# SCHOLARLY PUBLICATIONS

*A CURRENT AWARENESS BULLETIN*

*OF RESEARCH OUTPUT*

# *@DTU*

(115th Edition)

***JULY 2022***

## 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 Fifteenth** 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, 2020**. 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 literature review and perspectives in deepfakes: generation, detection, and applications, ***6.Deepak Dagar*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

2. A N2CNN-based Anomaly Detection Method for Cardiovascular data in Cyber-Physical System, Raju Pal, ***6.1.Ashish Kumar Tripathi***, Avinash Chandra Pandey, Mohammad Ayoub Khan, Varun G. Menon and Himanshu Mittal, CSE, DTU

3. A nonlinear SAIR epidemic model: Effect of awareness class, nonlinear incidences, saturated treatment and time delay, ***6.Kanica Goel*** and ***3.Nilam***, Applied Mathematics, DTU

4. A Novel Traffic Based Framework for Smartphone Security Analysis, Sumit Kumar, ***3.S. Indu*** and Gurjit Singh Walia, Electronics, DTU

5. A Performance Study on Deep Learning Covid-19 Prediction through Chest X-Ray Image with ResNet50 Model, ***3.Dhirendra Kumar***, ***8.Pulkit Sharma***, ***3.Anupama*** and ***8.1.Parth Sharma***, Mathematics and Electrical, DTU

6. Accessible Polycystic Ovarian Syndrome Diagnosis Using Machine Learning, ***8.Akanksha Tanwar***, ***8.Anima Jain*** and ***3.Anamika Chauhan***, IT, DTU

7. Advances in Applications of Non-Destructive Testing (NDT): A Review, ***8.Mridul Gupta*** and ***8.Muhsin Ahmad Khan***, Mechanical, DTU

8. Aerodynamic study of tall buildings under wind load, ***6.Kunwar Navodit Srivastava*** and ***3.Ritu Raj***, Civil, DTU

9. Agri-tech: experiential learning from the Agri-tech growth leaders, ***6.Sanjay Chaudhary*** and ***3.P. K. Suri***, DSM, DTU

10. An assessment of historical and future coastal dynamic response along the Odisha coast, ***7.Badal Mohanty*** and ***3.Raju Sarkar***, Civil, DTU
11. Applications of metal–organic framework-based bioelectrodes, ***8.Vidushi Aggarwal***, ***3.Shipra Solanki*** and ***3.Bansi D. Malhotra***, Biotechnology and Applied Chemistry, DTU

**12.** ASSESSMENT OF 3D MESH WATERMARKING TECHNIQUES, ***6.Neha Sharma*** and ***3.Jeebananda Panda***, Electronics, DTU

**13.** Assessment of Bio-Based Polyurethanes: Perspective on Applications and Bio-Degradation, ***3.Raminder Kaur***, ***6.Pooja Singh***, ***6.Surya Tanwar***, ***7.1.Gunjan Varshney*** and ***6.Sarla Yadav***, Chemistry, DTU

**14.** Automatic illness prediction system through speech, Husam Ali Abdulmohsin, Belal Al-Khateeb, Samer Sami Hasan and ***6.1.Rinky Dwivedi***, CSE, DTU

**15.** Brain tumor detection and classification by MRI images using deep learning techniques, ***7.Rashika Mangla*** and ***7.Chetna***, Applied Mathematics, DTU

**16.** Complaint and Severity Identification from Online Financial Content, Apoorva Singh, ***8.Rohan Bhatia*** and Sriparna Saha, DTU

**17.** Computation of energy dissipation across the type-A piano key weir by using gene expression programming technique, ***8.1.Deepak Singh*** and ***3.Munendra Kumar***, Civil, DTU

**18.** COVID-19 Lock-down in Delhi: Understanding Trends of Particulate Matter in Context of Land-Use Patterns, GIS Mapping, and Meteorological Traits, Amrit Kumar, Kiranmay Sarma, ***6.Abhinav Pandey***, ***3.Rajeev Kumar Mishra*** and Panuganti China Sattilingam Devara, Environmental, DTU

**19.** Cyclic behaviour of stratified soil under liquefied states, Arpit Jain, Satyendra Mittal and ***6.Sanjay Kumar Shukla***, Civil, DTU

**20.** Deep Learning Framework for Macro Marine Litter Classification and Quantification, ***3.Dr. Shailender Kumar***, ***8.Abhinav Gautam***, ***8.Khushbu Mehlawat*** and ***8.Kunal Rawat***, CSE, DTU

**21.** Design and implementation of OTA based fractional-order oscillator, ***6.Garima Varshney***, ***3.Neeta Pandey*** and ***3.Rajeshwari Pandey***, Electronics, DTU

**22.** Design of novel MIMO UOWC link using gamma–gamma fading channel for IoUTs, ***6.Dushyant Singh Chauhan***, ***3.Gurjit Kaur*** and ***3.Dinesh Kumar***, ECE, DTU

**23.** Determinants of Spiritual Tourism Consumption: A Hierarchical Approach, Suneel Kumar, ***7.Shivangi*** Gupta and Shekhar, USME, DTU

**24.** Determination of ciprofloxacin using the Ziziphus mauritania derived fluorescent carbon dots based optical sensor, ***6.Navneet Chaudhry***, Damini Verma, Amit K Yadav, ***3.Jai Gopal Sharma*** and Pratima Solanki, Biotechnology, DTU

**25.** Dro-Mal Detector: A Novel Method of Android Malware Detection, ***8.Aman Chandok***, ***8.Aditya Verma*** and ***3.Rahul Gupta***, IT, DTU

**26.** Eco-design requisites for solar desaltification still augmented evacuated annular tube collectors with parabolic concentrator: an optimum-environ-economic viability, ***8.Ashok Kumar Singh*** and ***3.1.Samsher***, Mechanical, DTU

**27.** Eco-friendly Green Synthesis of stable ZnO nanoparticles using citrus limon: X-Ray Diffraction Analysis and Optical Properties, Anamika Kem, Mohd Rehan Ansari, P. Prathap, ***3.M. Jayasimhadri*** and Koteswara Rao Peta, Applied Physics, DTU

**28.** Effect of Interference Between Twin Tall Buildings with Recessed Corners on Wind-Induced Pressure, ***6.Suresh Kumar Nagar***, ***3.Ritu Raj*** and ***3.Nirendra Dev***, Civil, DTU

**29.** Effect of preheating on the mechanical and high temperature tribological behaviour of reusing marble dust with aluminium composite, Sumit Chaudhary, Sonia Dangi, ***8.Vikas Yadav***, Ravinderjit Singh Walia, Narendra Mohan Suri and Mohit Tyagi, Mechanical, DTU

**30.** Effect of wind load on irregular shape tall buildings having different corner configuration, ***6.RAHUL KUMAR MEENA***, ***3.RITU RAJ*** and ***3.S ANBUKUMAR***, Civil, DTU

**31.** Encryption and Decryption of Marker Based 3- Dimensional Augmented Reality Image Using Modified Hill Cipher Technique for Secure Transfer, ***3.Abhilasha Sharma***, ***8.Anshul Singh*** and ***8.Appu Kumar***, Software Engineering, DTU

**32.** Entanglement in interacting quenched two-body coupled oscillator system, Sayantan Choudhury, Rakshit Mandish Gharat, Saptarshi Mandal, ***8.Nilesh Pandey***, Abhishek Roy and Partha Sarker, Applied Physics, DTU

**33.** Examining Systemic Risk using Google PageRank Algorithm: An Application to Indian Non-Bank Financial Companies (NBFCs) Crisis, ***3.Anurag Chaturvedi*** and ***3.Archana Singh***, USME AND DSM, DTU

**34.** Exploring Multimodal Sentiment Analysis through Cartesian Product approach using BERT Embeddings and ResNet-50 encodings and comparing performance with pre-existing models, ***3.Dr. Aruna Bhat***, ***8.Rahul Mahar***, ***8.Rahul Punia*** and ***8.Rahul Srivastava***, CSE, DTU

**35.** Fabrication and investigation on influence of metal oxide nanoparticles on thermal, flammability and UV characteristics of polyethylene glycol based phase change materials, ***6.Surya Tanwar*** and ***3.Raminder Kaur***, Applied Chemistry, DTU

**36.** Flower Pollination Algorithm for Scheduling Optimization of Flexible Manufacturing System, ***6.Gayathri Devi K***, ***3.RS Mishra*** and ***3.AK Madan***, Mechanical and Production Industrial and Automobiles, DTU

**37.** Food Review Analysis and Sentiment Prediction using Machine Learning Models, ***8.Dhruv Gupta***, ***8.Ausho Roup***, ***8.Diksha Gupta*** and ***3.Avinash Ratre***, ECE, DTU

**38.** Grid Tied Photovoltaic Based Electric Vehicle Charging Infrastructure, ***7.Mamta Gupta***, ***6.Aakash Kumar Seth*** and ***3.Mukhtiar Singh***, Electrical, DTU

**39.** Identication of storm eye from Satellite image data using fuzzy logic with machine learning, ***7.Monu Yadav*** and ***3.Laxmi Narayan Das***, Applied Mathematics, DTU

**40.** Impact of plasma process parameters on the growth of vertically aligned carbon nanotube array and its optimization as field emitters, ***6.Umang Sharma*** and ***3.Suresh C. Sharma***, Applied Physics, DTU

**41.** Impedance based damage assessment of concrete under the combined effect of impact and temperature using different piezo configurations, ***6.Indrajeet Singh***, ***3.Nirendra Dev*** and ***3.Shilpa Pal***, Civil, DTU

**42.** Implementation of Inductive Wireless Power Transmission system for Battery Charging applications, ***7.Abhishek Shakya*** and ***3.Mayank Kumar***, Electrical, DTU

**43.** Implementation of Z-Source Resonant Converter for Power Factor Correction & Wireless Charging Applications, ***7.Ayush Kumar Srivastava*** and ***3.Narendra Kumar***, Electrical, DTU

**44.** Importance of Initialization in K-Means Clustering, ***8.Anubhav Gupta***, ***8.Antriksh Tomer*** and ***3.Sonika Dahiya***, Software Engineering, DTU

**45.** Improved resistive switching of RGO and SnO2 based resistive memory device for nonvolatile memory application, ***7.Km. Komal***, Govind Gupta, ***3.Mukhtiyar Singh*** and ***3.Bharti Singh***, Applied Physics, DTU

**46.** Kernel picture fuzzy clustering with spatial neighborhood information for MRI image segmentation, ***3.Dhirendra Kumar***, ***8.Inder Khatri***, ***8.Aaryan Gupta*** and Rachana Gusain, Applied Mathematics, DTU

**47.** Memory Recurrent Elman Neural Network-Based Identification of Time-Delayed Nonlinear Dynamical System, ***3.Rajesh Kumar***, Electrical, DTU

**48.** Multilabel Facial Emotion Classification, ***7.Mansi Bhardwaj*** and ***7.Mansi Panchal***, Applied Mathematics, DTU

**49.** Multi-level Classification and Detection of Cardiac Arrhythmias with High-Resolution Superlet Transform and Deep Convolution Neural Network, Prashant Mani Tripathi, Ashish Kumar, ***3.Manjeet Kumar*** and Rama Komaragiri, ECE, DTU

**50.** Optimal Reactive Power Dispatch Using Giza Pyramids Construction Algorithm, ***6.Ganesh Kumar Jaiswal***, ***3.Uma Nangia*** and ***3.N. K. Jain***, Electrical, DTU

**51.** Optimization of process parameters for 3D printing process using Taguchi based grey approach, Ramesh Raju, M. Murali M.K. Varma and ***6.1.Pushp Kumar Baghel***, Mechanical, DTU

**52.** Parametric optimisation for detoxification of non-oil Jatropha residual material using Taguchi method, ***6.Shilpi Ahluwalia***, Amrit Pal Toor, ***3.Pushpendra Singh*** and ***3.Jai Gopal Sharma***, Biotechnology, DTU

**53.** Performance Analysis of CSS Over α–η–μ and α–κ–μ Fading Channel Using Clustering-Based Technique, Sandeep Kumar, Puspraj Singh Chauhan, Rahul Bansal, ***8.Manpreet Kaur*** and ***3.Rajesh Kumar Yadav***, CSE, DTU

**54.** Power Flow Management and DC bus Stabilization for SPV based Battery Energy Storage System, ***6.Anupam***, ***3.Mayank Kumar*** and ***3.Suman Bhowmick***, Electrical, DTU

**55.** Real-time Vehicle Detection, Classification & Tracing Using State-Of-The-Art Dataset, ***8.Aadhar Dutta***, ***8.Aman Jha***, ***8.Rajat Rao*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

**56.** Review of factors affecting facial recognition algorithms performance, ***3.Shweta Meena***, ***8.Kanishk Vats*** and ***8.Deepak Kumar***, Software Engineering, DTU

**57.** Sign Language to Text Conversion using Deep Learning, ***3.Dr. Aruna Bhat***, ***8.Vinay Yadav***, ***8.Vishesh Dargan*** and ***8.Yash***, CSE, DTU

**58.** Stimulating CSR learning collaboration by the mentor universities with digital tools and technologies – an empirical study during the COVID-19 pandemic, Namita Jain, Asha Thomas, ***3.Vikas Gupta***, Mario Ossorio and Daniele Porcheddu, DSM, DTU

**59.** Sustainable Reservoir Sediment Management in Indian River Basins, ***3.S.K. MAZUMDER***, Civil, DTU

**60.** Synthesis, structural characterizations, electrochemical properties and DFT calculations of highly fluorescent phosphorus(V) corroles, ***6.Ritika Kubba***, Omprakash Yadav, Suman Maji, Natalia Fridman and ***3.Anil Kumar***, Applied Chemistry, DTU

**61.** Two/Three-element textile Diversity MIMO Antenna for Biomedical and Wearable Applications, Anubhav Kumar, ***3.Asok De*** and R.K Jain, Electronics, DTU

**62.** Type-2 Diabetes Mellitus (T2DM): Spatial-temporal Patterns of Incidence, Mortality and Attributable Risk Factors from 1990 to 2019 among 21 World Regions, ***6.Mehak Nanda***, ***3.Rajesh Sharma***, Sumaira Mubarik, ***6.Aashima Aashima*** and Kai Zhang, USME, DTU

**63.** Understanding Percentage Impedance (%Z) of Transformer wiht Examples, DR. ***7.1.Rajesh Kumar Arora***, Electrical, DTU

**64.** What causes outward foreign direct investment (OFDI) from India into least developed countries (LDCs)?, ***6.Apoorva Jain*** and ***3.Sonal Thukral***, DSM, DTU

**65.** Wind Effects on Rectangular and Triaxial Symmetrical Tall Building Having Equal Area and Height, Astha Verma, ***6.Rahul Kumar Meena***, ***3.Hrishikesh Dubey***, ***3.Ritu Raj*** and S. Anbukumar, Civil, DTU

**66.** Wind interference effect on a hexagonal-shaped high building with orifices, ***8.Harmanbir Singh***, ***8.Harsh Goyal***, ***8.Mayank Singh*** and ***3.Ritu Raj***, Civil, 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*

International Journal of Multimedia Information Retrieval
https://doi.org/10.1007/s13735-022-00241-w

TRENDS AND SURVEYS

# A literature review and perspectives in deepfakes: generation, detection, and applications

**Deepak Dagar**[1] · **Dinesh Kumar Vishwakarma**[1]

Received: 2 February 2022 / Revised: 2 June 2022 / Accepted: 13 June 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

**Abstract**
In the last few years, with the advancement of deep learning methods, especially Generative Adversarial Networks (GANs) and Variational Auto-encoders (VAEs), fabricated content has become more realistic and believable to the naked eye. Deepfake is one such emerging technology that allows the creation of highly realistic, believable synthetic content. On the one hand, Deepfake has paved the way for highly advanced applications in various fields like advertising, creative arts, and film productions. On the other hand, it poses a threat to various Multimedia Information Retrieval Systems (MIPR) such as face recognition and speech recognition systems and has more significant societal implications in spreading misleading information. This paper aims to assist an individual in understanding the deepfake technology (along with its application), current state-of-the-art methods and gives an idea about the future pathway of this technology. In this paper, we have presented a comprehensive literature survey on the application of deepfakes, followed by discussions on state-of-the-art methods for deepfake generation and detection for three media: Image, Video, and Audio. Next, we have extensively discussed the architectural components and dataset used for various methods of deepfakes. Furthermore, we discuss the various limitations and open challenges of deepfakes to identify the research gaps in this field. Finally, discuss the conclusion and future directions to explore the potential of this technology in the coming years.

**Keywords** Deepfake · Face-swap · Face reenactment · Lip-sync · Body-puppetry · Speech synthesis · Deep learning

## 1 Introduction

Nowadays, there has been an uproar of information dissemination in the form of images, video, and audio on the web. An earlier attempt that can be traced in the manipulation field was in 1865, with an iconic photograph of then US president Abraham Lincoln where his face was swapped [1]. Deepfake is the current state-of-the-art in the image, video, and audio manipulation.

Deepfake word is composed of two words, "Deep" and "fake," which means the fake media [1] that has been created using a deep neural network, a branch of machine learning. Fake media created by this technology appear so realistic (Fig. 1) and believable that it is difficult to identify as counterfeit to the naked eye. This word became famous when, in 2017, a Reddit user with the name "deepfakes" created pornographic content with a swapped face of a celebrity and posed it online.[1] Since then, it has become one of the hot topics, and there has been a lot of research going on in recent times.

Few deepfake detection challenges are going on to set up the benchmark for deepfake detection. Facebook Inc., in collaboration with Microsoft, AWS (amazon web services), and partnership on AI committee, has created the deepfake detection challenge[2] (DDC) to encourage the researchers to detect fake and manipulated media. ASV (Automatic Speaker Verification) spoofing Challenge[3] is another challenge to detect AI-driven counterfeit voices.

Deepfake has become a buzzword for its versatile application to create different manipulations, especially in images and videos. It has opened the door for various exciting applications in multiple fields such as advertising, film production,

✉ Dinesh Kumar Vishwakarma
dvishwakarma@gmail.com

[1] Biometric Research Laboratory, Department of Information Technology, Delhi Technological University, Delhi 110042, India

[1] AI-Assisted Fake Porn Is Here(vice.com).

[2] Deepfake Detection Challenge | Kaggle.

[3] ASVspoof.

Published online: 23 July 2022

**Fig. 1** Example of a Deepfake [2]

creative arts and video games. On the other hand, it poses a threat to various societal problems like manipulating public opinion during elections, undermining journalism, eroding the trust in institutions, committing fraud, creating a false narrative & massive civil unrest and jeopardizing the nation's security. To an individual, it can discredit, blackmail, and harass people to an unimaginative height and may ruin their lives and careers.

The rise of deepfake has also become a potential threat to various multimedia information retrieval systems (MIPR) such as face recognition, speech recognition, biometric and gait analysis systems. Information retrieval systems look for the intended semantic knowledge from multimedia sources (e.g., databases). As the data for various identities are freely available on the internet nowadays, such information allows for creating an intended deepfake mask or audio (using readily available deepfake software), which can persuade the MIPR systems to a fake identity. Moreover, as the manufactured multimedia information freely circulates on the internet, it would dilute the empirical evidence, factual information and critical features extracted by the MIPR systems, which may further create false interpretation of the reality.

## 1.1 Application of deepfakes

The intention behind using the technology makes it beneficial or threatening to society (Fig. 2).

### 1.1.1 Beneficial use of technology

This technology offers many benefits if used with the right intention, and it may include the following fields:

**Education** Deepfake gives a plethora of opportunities for educators with the way they can impart education. For example, videos of historical personalities like Mahatma Gandhi or Nelson Mandela teaching about themselves and their work. In 2018, a circulated deepfake video of Barack Obama wherein he warns about the dangers of deepfakes was an apt example [3]. Teaching in the above manner makes study compelling for the students.

**Entertainment** Deepfake has also contributed to entertainment purposes in video dubbing in other languages, memes, GIFs, animating dead or cartoon characters, and special effects in the movies [4]. This industry will indeed leverage advanced features of this technology in the coming times.

**Expression** Deepfake technology allows people suffering from disabilities, such as ALS (Amyotrophic Lateral Sclerosis, the patient has difficulty speaking and communicating) to express themselves through their deepfake video. Deepfake also allows one to have an avatar experience through virtual engagement that might be impossible to have physically, e.g., in video games [3]. Deepfake can be used for social communication, such as speech, to reach an audience that understands a different dialect than the speaker's natural language [5].

**Innovation** Deepfake has been used a great deal to attract customers for a brand. For example, Reuters demonstrated the use of an AI-generated presenter-led sports news feed.[4] In the fashion retail industry, deepfake allows customers to turn into models (virtually) to try out new apparels.[5] A Japanese company named Data Grid is already using an AI-generated virtual model for advertising [5]. In the future, deepfakes will play a huge role in producing stimulating effects through advertising and branding.

[4] Synthesia Case studies: Reuters.

[5] Digital Doubles: The Deepfake Tech Nourishing New Wave Retail (forbes.com).

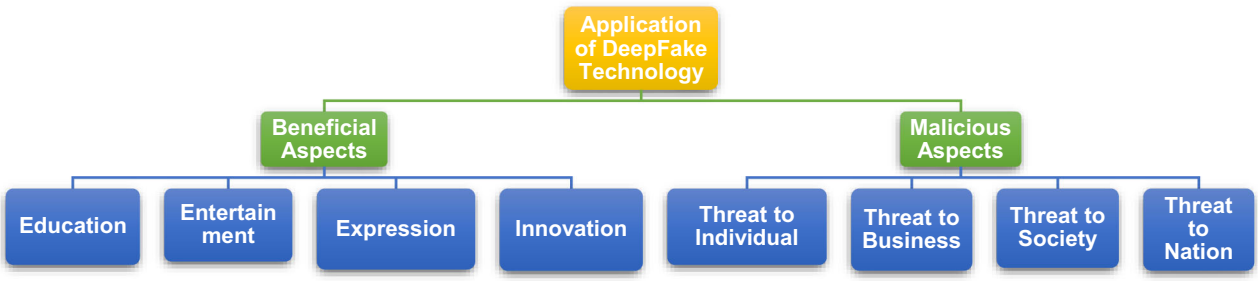



**Fig. 2** Application of Deepfake

### 1.1.2 Malicious use of technology

The real danger of this technology lies in the different ways it can be misused and the large-scale impact it can have, courtesy of being misused. Here are the some of the threats that it may create:

**Threat to individuals** Deepfake holds excellent potential for inflicting tangible harm, psychological stress, and physical pain. A fraudster may use deepfake to extract something of value for inflicting damage. To prevent the release of such deepfake, the victim provides money, personal banking details and business secrets [3]. Also, the most common form of exploitation is in the form of deepfake pornographic videos. One can victimize the individual to any violent or humiliating act to gratify their wants. At the workplace, deepfake videos can be used depicting someone indulging in anti-social actions or behavior like damaging property, hurling abuse or making racist remarks toward a co-worker. Also, deepfake audio or video can be used as evidence for harassment or sexual abuse, which can cause severe repercussions for an individual's career perspective and future aspirations [3].

**Threat to the business** Deepfake technology allows anyone to impersonate voices of different identities like the Business leader and CEO to incur fraud. A recent example of it, in March 2019, the CEO of a UK-based corporation was asked by his boss impersonated voice to supply $243,000 to a supplier [6]. Also, a deepfake can manipulate the market quickly if rumors back a fake media, and someone may lose or make a considerable profit [7]. Deepfake ads could be used to defame a company's product and lower its brand value, which could be disastrous for a brand [8]. Deepfake can cause significant harm to an organization's reputation and future perspective in the market and can give undue advantage to its competitor [3].

**Threat to society** Deepfake can have a substantial societal impact considering its realism and fast propagation through different social media networks. The rise of deepfake would give tough times to people comprehending what is real and fake, which undermines the very existence of journalism. The Trust of the people in the institutions will erode as this technology prevails. Even if later the video proved fake, it cannot reverse its effect entirely. Deepfake acts as a powerful tool to spread misinformation among the masses on a large scale, creating chaos and panic and division in society. In the worst case, it can lead to a civil war in the community [3].

**Threat to nations** Deepfake can affect national and international relations; it can sour bilateral ties whose impact may last generations. Deepfake gives the option to external entities to influence the democratic process of a nation [3], which could trigger substantial civil unrest, and protests can turn violent, jeopardizing the nation's security. The well-timed circulation of deepfake can sway elections in a democratic country. The rise of deep fakes could dilute discussions and debates over the policies; empirical evidence would get enmeshed with doctored data. The advancement of such technology dilutes information, putting the speaker's credibility at stake [3].

## 1.2 Scope and contribution of the surveys

This paper primarily focuses on the application and technical aspects of the deepfake for three media. Technical aspects include deepfake creation and detection methods and datasets proposed from 2019 onwards. Furthermore, we have considered the loosely defined definition of deepfake, where fake media is created by the deep neural networks (DNNs). This paper has comprehensively reviewed existing literature for visual and audio deepfake. The **contribution of this survey** is summarized below:

- Categorize the application of deepfake initially into two categories, which further expand to give a detailed description with different scenarios about the usefulness and misuse of this technology.

- Compare the various state-of-the-art surveys and identify the gaps this review literature aims to address.
- Segregation of detection and generation methods into their relevant categories. Also, detailed descriptions of various generation and detection methods in tabular forms.
- Discuss the various architecture components and datasets used in the generation and the detection methods.
- Outline the various open-source tools used for deepfake creation for multiple platforms and the type of manipulation they do.
- Identify the limitations of the existing generation and detection methods, which need to be addressed by the research community.
- Present the future research trend and pathway of this technology in the coming years.

### 1.3 Organization of the review

Various research papers have extensively explored the deepfakes, categorizing methods at multiple levels. Tolsana et al. [9] have divided the deepfake techniques into four categories, limited to image and video media without explaining the detection mechanism based on the feature representation. Mirsky et al. [4] have focused on face replacement and face reenactment in great depth. Masood et al. [10] have considered the audio deepfake, but the categorization of detection methods is coarse and not refined further. Earlier literature has assisted in refining the categorization of generation methods for visual and audio media which helps to categorize procedures covering the entire range of techniques. Tolsana et al. [9] help arrive at the categorization for generation methods. Juefei-Xu et al. [11] give insight into feature representation based on detection traces, and also Yu et al. [12] provide an idea for the video media feature representation. These two manuscripts helped arrive at a feature representation that helps refine feature representation based on traces used for detection.

This review paper is segregated into six parts. The first part (Sect. 1) discussed deepfake technology and its application. The subsequent section (Sect. 2) will discuss the deepfake techniques: deepfake generation and detection. Each category contains state-of-the-art methods that have been proposed from 2019 onwards. Subsequently, this paper has presented the popular datasets in the tabular form (Sect. 3). Proposed methods contain various architectural components for their models; this paper briefly introduces such components and tools for deepfake creation (Sect. 4). Next, this paper has highlighted the various limitations of generation and detection methods (Sect. 5). Finally, we end the discussion with the conclusion and future pathway this technology may take in the coming years (Sect. 6).

## 2 Deepfake techniques

The two major algorithm classes in this area are deepfake generation and detection.

### 2.1 Deepfake generation

Based on the generation process of different media, the generation mechanism has been divided into visual deepfake, including image and video media, and audio deepfake generation. Figure 3 presents the further categorization of deepfake generation methods for these two categories.

#### 2.1.1 Visual deepfake generation

Visual deepfake generation involves the generation mechanism for image and video media. This section will cover various categories under which manipulation can take place.

**Identity swap (IS)** The identity/face of the source/frame is transferred to the target image/frame in this type of manipulation. Face-swap is the other name for it. The face-swap [13] and fake app [14] software made it easier for anyone to generate face-swap. Their typical approach for IS is based on a pair of auto-encoder and decoder architecture. For this process, encoder–decoder pairs are required, where the encoder converts the images into their latent representation, while the decoder reconstructs the image back from the latent representation. During the training, encoder–decoder pair is required for each image for the model to learn its embedding, where encoder weights are shared. Once training is complete, the decoder is interchanged during the generation stage, and the decoder of the source image and encoder of shared weights are used to generate the target results (Fig. 4).

The researcher proposed various sophisticated algorithms over the years. The first well-known Identity swap is FaceSwap [15], used to develop the FaceForensics + + dataset [16]. The traditional method uses 3D morphable models, and facial textures are replaced with the estimated 3D model's geometry with the target image. Dale et al. [17] model has been one of the old approaches that uses the multi-linear model to track the facial performance in both videos and then use 3D geometry to warp the faces. Nowadays, IS architecture uses DNN, which usually uses two modules; one uses latent space for disentanglement of identity from other attributes and then the other module transfers and refines the identity from source to target. Faceshifter [18] manipulation comprises two stages. In the first stage, the method extracts and embeds the target attributes thoroughly and adaptively, and in the second stage, the network recovers the anomalies region in a self-supervised manner. Most of the recent IS methods are subject agnostic, where once the model gets trained, it can be applied to any new faces without re-training

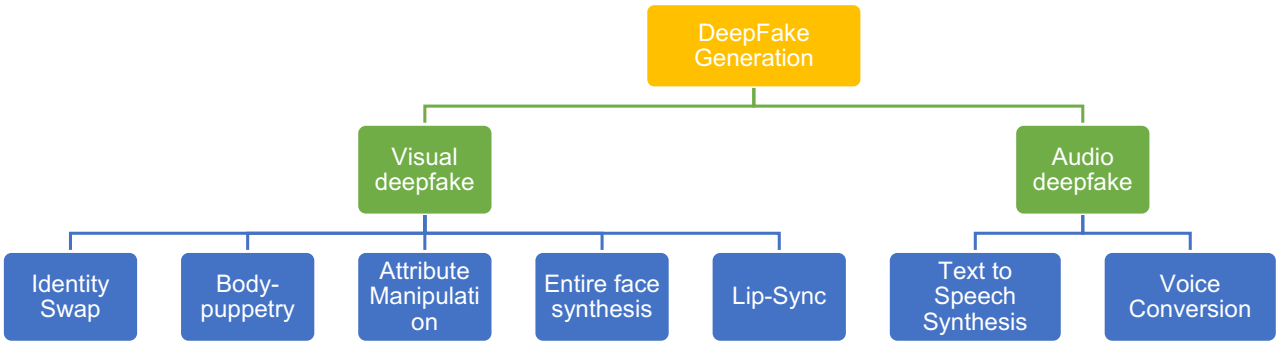



**Fig. 3** Classification of deepfake generation techniques

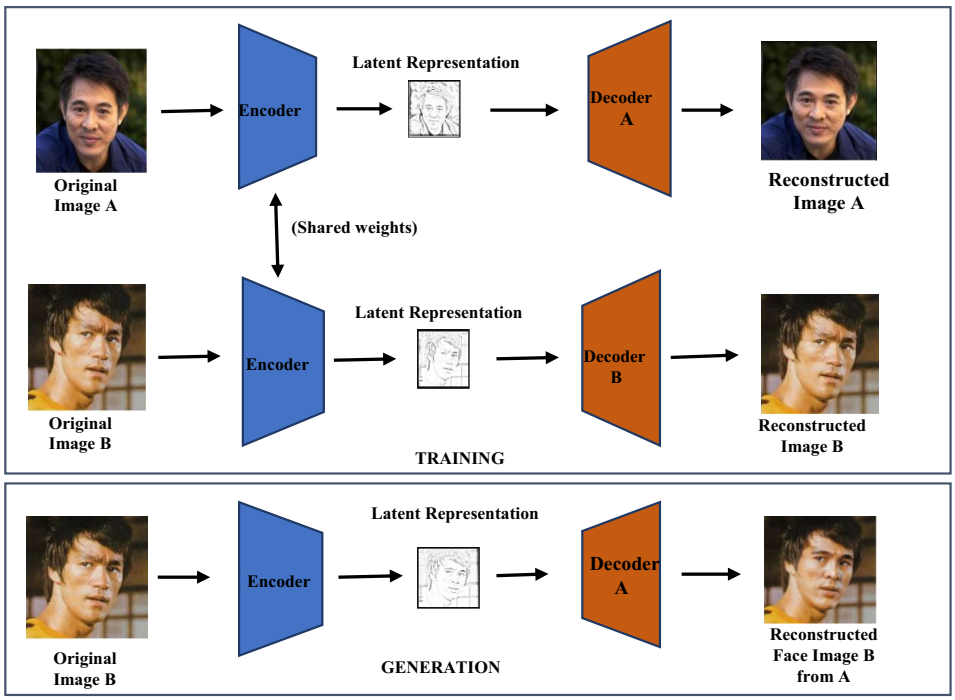



**Fig. 4** Identity Swap generation model using auto-encoder & decoder

on subject-specific targets. Table 1 describes recent popular identity swap methods.

**Body puppetry (BP, aka reenactment)** Body puppetry (aka reenactment) deepfake is where the source derives the content of the target; it can be facial gestures, eye and head movements, or different body poses. Face reenactment is its subset, where the facial attributes are derived. It is significantly used for post-production editing of movies or short videos [25]. Most of the time, the target content is derived either from some source media in the form of images/frames using landmark key points, 3D morphable models, skeletons or any other mapping method. Chan et al. [25] proposed a method to transfer dance moves from the source to the target using intermediate pose Skelton transfer and predicting the two consecutive frames to produce coherent results. However, the samples cannot generate realistic poses, especially at the body's joints. Thies et al. [26] proposed the famous Face2Face approach that allows for the real-time reenactment of facial expression, where source facial expressions are tracked using a dense photometric consistency measure, and then a transfer function exploits the deformation transfer in semantic space. Many techniques [27, 28, 29] require multiple input images of the source samples, while few methods require few instances [30, 31] or even a single [32] to generate results. Many methods have the limitation that they can synthesize one attribute at a time and apply it to only low-resolution images. FaceSwapNet [33] resolves this issue using two modules: landmark swapper and landmark-guided generator to generate the face expression enacted photo-realistic image. Table 2 lists the popular body puppetry with their analysis in various fields. Some other methods [34, 35, 36, 37, 38] have also been proposed over the years. Most of the reenactment has been done on dance poses and facial

**Table 1** Overview of popular identity swap methods

| Ref | Year | Model description | Subject agnostic | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | O/P media | | Limitation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | Img | Vid | |
| [19] | 2019 | Proposed a framework based on a novel recurrent neural network for both face swapping and facial reenactment for a pair of the subject without the training on the subject-specific data | ✓ | IJB-C dataset, FaceForensics + +, LFW, Figaro | SSIM = 0.51, Euler Score = ~ 3.21, Landmark score = ~ 24.5, Verification Score = 0.38 | SSIM = 0.54, Euler Score = ~ 2.49, Landmark score = ~ 22.2, Verification Score = 0.38 | GANs | 128 × 128 & 256 × 256 | • | • | The method fails to capture facial expressions & blurriness in image texture fully. Identity and texture quality degrade in case of significant angular differences |
| [18] | 2020 | Employed a two-stage GAN framework (called faceshifter) for a face-swapping algorithm. In the first stage, AEI-NET is used to extract & embed target attributes adaptively. In the second stage, (HEAR-Net) attributes are further refined by recovering face anomalies in a self-supervised manner | ✓ | FaceForensics ++, CelebA-HQ, FFHQ, VGGFace | ID retrieval = 97.38, pose error = 2.96, expression error = 2.06, Human evaluation(identity) = 52.9, Human evaluation(attr.) = 62.1, Human evaluation(realism) = 82.9 | | Encoder-Decoder | 256 × 256 | • | | Stripped artifacts of the generated samples |
| [20] | 2020 | Employed ID Injective Module (IIM) enables identity transfer from source to target image, extending the architecture from specific to arbitrary face manipulation | ✓ | VGGFace2, FaceForensics + + & CelebMask-HQ | Accuracy rate (ID retrieval) = 73.64%, Pose Estimation (L2 Distance) = 1.22 | Accuracy rate (ID retrieval) = 96.57%, Pose Estimation (L2 Distance) = 2.47 | GAN + Encoder-Decoder | 224 × 224 | • | | Blurriness at the edges |

**Table 1** (continued)

| Ref | Year | Model description | Subject agnostic | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | O/P media | | Limitation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | Img | Vid | |
| [21] | 2021 | Proposed one-shot learning methods for face swapping, where a hierarchical face encoder encodes face representation in new extended latent space to maintain facial details. A second module transfers the identity from the source to the target by nonlinear trajectory. Finally, the synthesis mechanism of styleGAN2 swaps faces | ✓ | CelebA, CelebA-HQ,FFHQ, FaceForensics + + | ID similarity = 0.5014, ID retrieval = 90.83, pose = 3.58, expression = 2.87, FID = 10.16 | ID similarity = 0.5014, ID retrieval = 90.83, pose = 2.64, expression = 2.96, FID = 10.16 | GANs | 256 × 256, 1024 × 1024 | • | | Fuzzy detailing at the edges and corners |
| [22] | 2021 | Proposed a GAN framework that allows specific attributes control in a high-fidelity face-swapping process. For this, they have used a pose expression (PE) block, which uses a feature-wise boundary map to correct the pose and expression of the manufactured face, along with a discriminator which keeps week supervision on them. Also, it uses a perceptual loss that leverages the discriminator to help preserve the facial attributes like skin color, illumination and occlusion | ✓ | FF-HQ, CelebA-HQ, VGGFace2 & FF + + | Identity retrieval(id) = 0.98, DIPD = 0.28, pose dist.(pose) = 2.43, expression dist.(exp) = 2.19, SSIM = 0.77, Avg. endpoint error(AEE) = 1.50, FEW = 0.22, FID = 30.31, SSIM = 0.78, MMS = 0.08 | Identity retrieval(id) = 0.98, DIPD = 0.28, pose dist.(pose) = 2.43, expression dist.(exp) = 2.19, SSIM = 0.77, Avg. endpoint error(AEE) = 1.50, FEW = 0.22, FID = 19.72, SSIM = 0.81, MMS = 0.90 | GANs | 256 × 256 | • | • | Visual inconsistencies and blurry artifacts at the boundaries |

**Table 1** (continued)

| Ref | Year | Model description | Subject agnostic | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | O/P media | | Limitation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | Img | Vid | |
| [23] | 2021 | Unified framework for high fidelity for face-swap and face reenactment using disentanglement of 3D poses shape and different expression factors and then recombining for various tasks | ✓ | FaceForensics ++, Ceelb-DF v2, voxCeleb2 | ID Similarity = ~ 0.62, Expression Error = ~ 1.71, Pose Error = ~ 2.62, SSIM = ~ 0.90, Human Deceive rate = ~ 0.35 | ID Similarity = ~ 0.80, Expression Error = ~ 1.67, Pose Error = ~ 1.97, SSIM = ~ 0.90, Human Deceive rate = ~ 0.53 | CNNs | 256 × 256 | • | | Blending inconsistencies in skin tones, wrinkles, 3D warping error, and contour errors in case of occlusion objects |
| [24] | 2021 | Proposed a unified framework for face swapping and face reenactment for video editing. The framework has a 3D dynamic training sample selection mechanism, and the optical flow loss constraint achieves temporal coherence through barycentric coordinate interpolation. In addition, a region-aware conditional normalization layer achieves more realistic results | ✓ | VoxCeleb2 | FID = 34.1, Identity error = 0.26, pose error = (1.12, 1.31, 0.65), expression error = 2.31 | | GANs | 64 × 64 | | • | Few generated samples are blurry |

**Table 2** Overview of popular body puppetry methods

| Ref. | Year | Model description | Subject Agnostic | Attribute transfer (reenactment) | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | O/P media | | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Img/frame | Video | |
| [25] | 2019 | Employed a motion transfer method that learns the pose mapping and transfers the appearance to the target, predicting two consecutive frames to produce coherent results for the synthesis | ✘ | Dance poses | YouTube Videos | Human Evaluation = 53.9%, SSIM = 0.816, LPIPS = 0.050 | Human Evaluation = 58.8%, SSIM = 0.838, LPIPS = 0.036 | GAN | 1920 × 1080 | | • | Model struggles to generate good illustrations for extreme poses. Sensitive to the variation in the scene lighting |
| [27] | 2019 | Network combining the traditional graphics pipeline with learnable rendering components through feature maps to produce a photo-realistic image re-synthesis content | ✘ | Facial Expression | Synthetic objects from artec3d website | Mean Square Error (MSE) = ~ 1.5 | Mean Square Error (MSE) = ~ 0.4 | Encoder-decoder | 512 × 512 | | • | Reliance on a geometry proxy. Requires to re-train the model for new samples |
| [39] | 2020 | Two main components, Dense Mapping Network (DMN) and Editing Behavior Simulated Training (EBST), are employed for diverse, interactive face manipulation | • | Facial Expression & Style | CelebA-HQ, CelebA | FID = 37.55%, A.C.Acc. = 68.1%, H. eval. = 28%, F.V.A = 76.41%, seg. = 92.31% | FID = 37.14%, A,V.Acc. = 89.5%, H.eval = 44%, F.V.A = 76.41%, seg. = 92.34% | GAN | 512 × 512 | • | | Fail to preserve the finer details of the regions |

**Table 2** (continued)

| Ref. | Year | Model description | Subject Agnostic | Attribute transfer (reenactment) | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | O/P media | | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Img/frame | Video | |
| [30] | 2019 | Proposed a few-shot learning-based pre-train model transforming facial landmark positions to generate a realistic-looking head | ○ | Facial Expression & head pose | VoxCeleb1, VoxCeleb2 | FID = 46.1, SSIM = 0.61, CSIM = 0.15, Human Evaluation = 0.62 | FID = 29.5, SSIM = 0.74, CSIM = 0.47, Human Evaluation = 0.33 | GAN | 256 × 256 | ○ | • | Sensitive to the source identity leakage and Lack of gaze adaption |
| [40] | 2019 | Employed a progressive pose attention transfer network that smoothly cascaded the input image features and the conditional pose using Pose-attentional Transfer blocks (PATBs) | • | Poses | Deepfashion and Market-1501 | SSIM = 0.311, IS = 3.209, M.SSIM = 0.811, mask-IS = 3.773, DS = 0.74, H. eval = 19.14% | SSIM = 0.773, IS = 3.323, M.SSIM = 0.811, mask-IS = 3.773, PCKh = 0.96, DS = 0.976, H.eval = 63.47% | GAN + Encoder-Decoder | 128 × 64, 256 × 256 | • | | Unwanted, distorted and blurry visual artifacts |
| [31] | 2019 | Employed a framework vid2vid utilizes the attention mechanism, which uses a few images of the target image to generate videos of unseen using input semantic videos or masks | ○ | Head poses, dance poses & street scene | CityScapes, ApolloScape, Camvid, FF & YouTube videos | FID = 144.24, Pose Error = 6.01, mIoU = 0.408, Pixel Accuracy = 0.831 | FID = 80.44, Pose Error = 6.01, mIoU = 0.408, Pixel Accuracy = 0.831 | Encoder + Decoder | | | • | Does not generalize to unseen domains, and the input is limited to semantic estimations |

**Table 2** (continued)

| Ref. | Year | Model description | Subject Agnostic | Attribute transfer (reenactment) | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | O/P media | | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Img/frame | Video | |
| [28] | 2019 | Employed a neural network that first tracks the 3D motion, transferring it to synthetic renderings to generate corresponding realistic human poses | ✘ | Dance poses | Videos from Youtube | L2 error 15.67433, SSIM = 0.65328 | | GAN + Encoder-Decoder | 512 × 512 | | • | Does not generalize for an arbitrary person. Visual artifacts occur mainly at the end-effectors like hands or feet |
| [41] | 2019 | Employed a generative network that uses two branches; one branch learns the generation of various poses, and the other improves the temporal coherence of unseen poses | ✘ | Dance poses | Web videos | P.to.P = 17.2, p.to.S metric = ~ 29(dist), MSE = 16.9, T.C.E = 0.0140 | P.to p = 4.2, p.to.seq = 0(dist), MSE = 15.4, T.C.error = 0.0092 | GAN | 256 × 256 | | • | Fail to generate the fine details and also does not change the background |
| [42] | 2019 | Leverages two generative models: the first synthesized novel poses, and then the latter fuses the background details, further refining the realism of the video frames | ✘ | Dance poses | YouTube videos | MSE = 642.9080, SSIM = 08,115, PSNR = 21.3286, Human Evaluation = 29.66% | MSE = 171.3259, SSIM = 0.9352, PSNR = 26.172, H.eval = 41.73% | GAN + Encoder-Decoder | 256 × 256, 512 × 512 | | • | Limited to the generation of static background frames |

**Table 2** (continued)

| Ref. | Year | Model description | Subject Agnostic | Attribute transfer (reenactment) | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | O/P media | | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Img/frame | Video | |
| [43] | 2020 | A two-stage pipeline in which the first extracts the facial expression and attributes(Action Units), and the latter transfers those attributes onto the image | • | Facial expression poses | VoxCeleb, CelebA | Accuracy = 59.88, F-score = 0.3759 | Accuracy = 62.86, F-score = 0.4185 | GAN | 128 × 128 | • | | Limited resolution and lack of diversification of the output image poses |
| [44] | 2019 | Network that first produces the intermediate generation of the correct and detailed poses, conditioned to create temporally coherent photo-realistic video frames | • | Poses | Fashion & Tai-Chit | P.loss = 0.5960, FID = 75.44, AKD = 3.77 | P.loss = 0.2811, FID = 13.09, AKD = 1.36 | GAN + Encoder-Decoder | 256 × 256 | | • | Limited to the single source video generation |
| [33] | 2019 | FaceSwapNet framework uses two modules, landmark swapper, and landmark-guided generator, to generate the face expression enacted photo-realistic image | • | Facial Expression | RaFD | SSIM = 0.659, FID = 13.26, Human Evaluation = 74.9% | SSIM = 0.711, FID = 11.67, Human Evaluation = 74.9% | GAN + Encoder-Decoder | 256 × 256 | • | | Visual inconsistencies of the generated samples |

**Table 2** (continued)

| Ref. | Year | Model description | Subject Agnostic | Attribute transfer (reenactment) | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | O/P media | | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Img/frame | Video | |
| [29] | 2021 | Framework uses a sequential generator and an ad-hoc discriminator to capture the complex non-rigid facial movement to synthesize a temporally consistent video | ✘ | Facial expression, head pose & eye gaze | Head2Head(Self-built), FaceForensics + + | Av. Exp. Dist.(AED) = 1.509, Av. Rot. Dist.(ARD) = 0.83, DAI = 8.86, FID = 1242, Max. Mean Discrepancy($MMD^2$) = 3915, Av. Eye land. Dist.(AELD) = 0.71 | | GANs | 256 × 256 | | • | Data-intensive model and also computationally expensive |
| [32] | 2021 | Composed of three networks, where the first separates the pose and other frame generation parts, the second converts the posture and target id to a masked frame, and the third network does the further refinement | • | Dance poses | Multi-Human Parsing(MHPv2), CIHP, DFDC | SSBS = 0.902, SSIS = 0.218, Dense Pose Binary similarity(DPBS) = 0.928, Dense Pose Index Similarity(DPIS) = 0.5, LPIPS = 0.375, SSIM = 0.116, FID = 83.95 | | CNNs | 512 × 320 | • | • | Does not produce coherent results, appearance is not preserved, and edges are pixelated |

expressions. There has been a considerable improvement in the quality of dance poses generated samples, but still, they are far from appearing realistic.

**Lip-syncing (LS)** This category of video manipulation involves synthesizing the mouth region of a target identity consistent with the arbitrary input audio. Lip movement and the corresponding expression are key elements to convey the information effectively. Usually, Influential leader's deepfakes are developed, as their audios, videos, and images are readily available, and their generated samples create more impact. Suwajanakorn et al. [45] created one of the first well-known lip-sync of ex-President Obama from the audio using a recurrent neural network to map raw audio features to different mouth textures. However, a new identity requires the model to be trained again. Earlier lip-sync methods [46, 45] construct 3D talking face models for a specific by animating 3D face meshes of the particular chosen subject, and such techniques are hard to scale to arbitrary identities. However, as the technology evolves, they disentangle audio-visual representation, allowing methods [47, 48] to use few subject samples to generate results. Real-time reenactment of audio has also become possible nowadays. Jamaluddin et al. [49] proposed a real-time cross-modal self-supervision model for synthesizing talking heads. They employ encoder–decoder multi-stream CNN, which uses a joint embedding of still images and audio to generate lip-synched video frames in real time. However, the model lacks the synthesis of real-time emotional facial expressions. There has been significantly less work for lip-sync manipulation, and also, methods have difficulties generalizing to any arbitrary identity. Table 3 gives an overview of popular lip-sync methods.

**Attribute manipulation (AM)** Attributes like expressions, hair, eyes, the color of skin, age, gender, mustache, etc. that are manipulated in an image fall into the manipulation category.

Generally, attribute manipulation methods employ either encoder–decoder (ED) or a combination of ED and GANs with a conditioned attribute. ED-based model decodes the latent representation of attributes in a latent representation. A relationship is established between latent representation and attribute independent editing, which allows the independent attribute manipulation without the identity information loss, leading to a distorted or over-smooth generation of the results.

StarGAN and STGAN are classic examples. Earlier domains used to do the image-to-image translation between two domains, which was time-consuming, but StarGAN's [50] approach uses multi-domain image-to-image translation using a single model. It allows the training of multiple datasets of different domains within the same network. However, the model can only produce a limited number of expressions despite such flexibility. To address this limitation, Albert et al. [51] novel GAN-based model named ganimation uses a weakly supervised attention mechanism that takes annotated facial action units (AU) as input and generates a broader range of expressions. Encoder-decoder and GAN architecture has bottleneck layers, which results in blurry and low-quality results and adding skip connection to overcome these, results in weakened attribute manipulation. Ming et al. [52] proposed STGAN using selective transfer manipulation that incorporates specific target units into the encoder-decoder model, changing target face attributes. While manipulating, Guim et al. [24] use latent space and conditional attribute representation, which helps regenerate the image by modifying the required attribute. Table 4 gives an overview of some popular method of attribute manipulation methods. Some other techniques [53, 54], which have been proposed recently, manipulate the style of an image using StyleGAN [55] latent space.

**Entire image synthesis (EIS)** The powerful GANs create entire non-existent images with high realism. ProGAN [67] and StyleGAN1 [55] have leveraged the power of GANs to create highly realistic synthetic high-resolution images. Terro et al. [67] proposed the ProGAN methodology, which allows for generating high-resolution images progressively by adding the number of layers gradually with the training. They started with a low-resolution image and then gradually increased the layers of the GANs and the image's resolution. The image generated by the model is of high quality but is far from appearing real at times. Another method, StyleGAN1 [55], interpolates the various features such as pose and human identity by disentangling the high-level attributes from the stochastic variation (like freckles, hair) of the generated image in an unsupervised setting. The method enables intuitive, scale-specific control of the synthesis. However, they found several typical artifacts of StyleGAN. To improve the model, they redesigned the architecture StyleGAN2 [68] with the normalization used in the generator. They also adapted the progressive GAN approach by regularizing the mapping of the generator from the latent code to images. Most methods have a hard time finding the trade-off between fidelity and the variety of generated samples. Table 5 gives an overview of popular image synthesis methods.

**Summary of the visual deepfake generation methods** Several deepfake generation methods have been proposed over the years. Realism has increased over the past few years, and now the naked have difficulty identifying. Despite being so realistic, in the case of multi-modal data like the video at specific frames, the visual inconsistencies appear, causing flickering and jitteriness. Lip-sync and Body-puppetry, especially in the case of poses, need some improvements to be more realistic at specific frames.

**Table 3** Overview of popular lip-sync methods

| Ref. | Year | Model description | Subject agnostic | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | Limitations |
|---|---|---|---|---|---|---|---|---|---|
| [56] | 2019 | Employed a framework called Disentangled Audio-Visual System (DAVS) that generates high-quality talking video frames by integrating the disentangled speech and Identity space representation using associative and adversarial training processes | • | LRW, MS-Celeb-1 M, VoxCeleb | PSNR = 25.4, SSIM = 0.859, Human Evaluation(realistic) = 51.5%,, Human Evaluation(Lip-Audio sync) = 72.3% | PSNR = 26.8, SSIM = 0.884, Human Evaluation(realistic) = 87.8%,, Human Evaluation(Lip-Audio sync) = 88.4% | GAN + Encoder-Decoder | 256 × 256 | Method lacks the synthesis of emotional facial expressions |
| [46] | 2019 | Employed a model based on a dynamic programming strategy for talking head video that allows you to edit, insert and delete the text in an existing transcript | ✘ | DF-TIMIT videos also taken from YouTube | RMSE = 0.021, Human Evaluation(real) = 57.1% | RMSE = 0.018, Human Evaluation(real) = 62.1% | RNN + GAN + Encoder-Decoder | 512 × 512 | The model requires huge time (at least 1 h of video) to produce the best quality results. In addition, occlusion of the lower face region would lead to synthesis artifacts |

**Table 3** (continued)

| Ref. | Year | Model description | Subject agnostic | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | Limitations |
|---|---|---|---|---|---|---|---|---|---|
| [57] | 2019 | Employed a method for generative talking videos by converting audio signals to facial landmarks and mapping them to facial appearances in video frames. Also employed an adjustable pixel-wise loss with an attention mechanism and regression-based discriminator for smooth facial movement and audio-visual consistency | • | LRW, GRID, VoxCeleb,TCD and real-world samples from YouTube | LMD(Landmarks Distance) = 1.82, PSNR = 28.78, SSIM = 0.72 | LMD(Landmarks Distance) = 1.29, PSNR = 0.83, SSIM = 32.15 | RNN + Encoder-Decoder | 128 × 128 | Limited to the conscious head movement of the identity |
| [58] | 2019 | Employed an end-to-end temporal generative model that uses three discriminators. It helps attain a detailed level frame, audio-visual sync, and realistic facial expression synthesis to generate talking heads with facial movements using an audio clip and a still image | • | GRID, TCDM TIMIT, CREMA-D, LRW | PSNR = 20.107 SSIM = 0.658, CPBD = 0.189, ACD = 2.61 × $10^{-4}$, WER = 58.2%, AV offset = 8, AV Confidence = 1.4 | PSNR = 27.1 SSIM = 0.818, CPBD = 0.308, ACD = 1.02 × $10^{-4}$, WER = 23.1%, AV offset = 1, AV Confidence = 7.4 | GAN + Encoder-Decoder | 96 × 128 | Limited output resolution and the methods work for only frontal poses |

**Table 3** (continued)

| Ref. | Year | Model description | Subject agnostic | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | Limitations |
|---|---|---|---|---|---|---|---|---|---|
| [59] | 2020 | Employed an audio-driven facial reenactment model that inherently maps the audio stream to represent the person talking style using 3D latent space. Then, a novel rendering network uses the representation of latent mapping to produce photo-realistic talking video frames | ✘ | News-speaker videos are taken from the internet | Human Evaluation(visual) = ~ 60%, Human Evaluation(Audio-visual sync) = ~ 60% | | RNN | 512 × 512 | The method fails when the audio stream contains multiple voices and also when the video contains occluded objects |
| [60] | 2019 | Employed subject-independent temporal GAN method that directly maps the audio features into visual features to produce the accurate lip-sync photo-realistic talking heads with facial movements | • | GRID, TCD TIMIT dataset | PSNR = 23.01, SSIM = 0.654, CPBD = 0.252, FDBM = 0.097, ACD = 2.29 × $10^{-4}$, WER = 37.2%, Human Evaluation = 20.22% | PSNR = 27.98, SSIM = 0.844, CPBD = 0.280, FDBM = 0.114, ACD = 1.02 × $10^{-4}$, WER = 25.4%, Human Evaluation = 79.77% | GRU + GAN + Encoder-Decoder | 96 × 128 | Facial expression is solely derived from the spoken word; it does not represent the speaker's mood |
| [49] | 2019 | Proposed a cross-modal self-supervision model of synthesizing talking head that employs encoder-decoder multi-stream CNN, which uses a joint embedding of still images and audio to generate lip-synched video frames in real time | • | LRS2, VoxCeleb2, VGG face | MSE = 705, Embedding distance = 0.434, Retrieval accuracy = 79.7% | MSE = 327, Embedding distance = 0.115, Retrieval accuracy = 83.9% | Encoder-Decoder | 109 × 109 | Model struggles in making the chins move with the mouth, and it also lacks the synthesis of emotional facial expression |

**Table 3** (continued)

| Ref. | Year | Model description | Subject agnostic | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | Limitations |
|---|---|---|---|---|---|---|---|---|---|
| [48] | 2021 | Proposed a model to synthesize low-guided high-quality talking head generation consisting of two cascaded modules. The first module contains a generator to produce animate facial attributes like the movement of the mouth, head, eyebrow etc. The second module transforms these animations into more expression details, and the flow-guided mechanism synthesizes the video | • | Videos collected from the YouTube | PSNR = 24.4174, SSIM = 0.84, CPBD = 0.153, Human Visual Evaluation, MSE = 0.0875, LMD = 0.1899, CCA = 0.786 | | Encoder-decoder | 720p ~ 1080p | Method cannot produce temporally coherent results, unnatural speaking style, and the head pose is not extreme enough |

**Table 3** (continued)

| Ref. | Year | Model description | Subject agnostic | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | Limitations |
|---|---|---|---|---|---|---|---|---|---|
| [47] | 2021 | Proposed a framework for animating talking heads faces from arbitrary audio. First, they convert the video footage into normalized that decouples pose, texture, lighting etc. Secondly, they leverage facial symmetry to approximate skin constancy, isolate spatiotemporal lightning, and then use an auto-regressive approach to stabilize temporal dynamics of the resultant frames | ○ | GRID, TCD-TIMIT, CREMA-D | SSIM = 0.91, LMD = 1.57, CPBD = 0.25, WER = 18%, Human Eval(Lip-sync) = ~ 2.46, Human Eval(Vis. Quality) = ~ 4.10 | SSIM = 0.94, LMD = 0.80, CPBD = 1.57, WER = 18%, Human Eval(Lip-sync) = ~ 2.46, Human Eval(Vis. Quality) = ~ 4.10 | Encoder-Decoder | 256 × 256 | Synthesized heads movement seems out of place at various frames, and also, the method is limited to short videos, and it is a bit slower. Hence cannot be deployed in real-world scenarios |

**Table 4** Overview of popular Attribute Manipulation methods

| Ref | Year | Model description | Subject agnostic | Attributes manipulated | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | O/P media | | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Img/frame | Video | |
| [61] | 2019 | Employed attribute classification constraints for manipulating desired attributes, reconstruction loss to preserve the non-manipulated details of an image, and adversarial loss for realistic visual learning of the attribute | ✘ | Bald, bangs, Black hair, Blond hair, Brown hair, Bushy eyebrows, Eyeglasses, Gender, Mouth open, Moustache, No beard, Pale skin, Age | CelebA & LFW | Facial Attribute Editing Accuracy = ~ 0.35 Attribute Preservation error = ~ 0.10 | Facial Attribute Editing Accuracy = ~ 0.95 Attribute Preservation error = ~ 0.01 | GAN | 64 × 64, 128 × 128, 384 × 384 | • | | Manipulation is limited to facial attributes only and cannot apply to other general attributes of a person |
| [62] | 2020 | InterfaceGAN allows semantic editing of facial attributes by interpreting the latent Space representation learned by the GAN model | ✘ | Gender, age, expression, eyeglasses, pose | CelebA-HQ | Classification Accu = 75.3%, Correlation matrix = -0.08 | Classification Accu = 90.3%, Correlation matrix = 1.0 | GAN & SVM | - | • | | Method is more inclined to the manipulation of synthetic images |

**Table 4** (continued)

| Ref | Year | Model description | Subject agnostic | Attributes manipulated | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | O/P media | | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Img/frame | Video | |
| [52] | 2019 | Employed STGAN having an encoder-decoder network that incorporates the specific transfer units as input for target attributes manipulation while improving the resolution of an image | ✘ | Bald. Bangs, Black hair, Blonde hair, Busy eyebrows, Eyeglasses, Male, Mouth slightly open, Mustache, No Beard, Pale skin and Young | CelebA | Human Evaluation = 47.6, Attribute generation accuracy = ~ 0.25 | Human Evaluation = 69.96, Attribute generation accuracy = ~ 0.85, PSNR/SSIM = 31.67/0.9948 | GAN + Encoder-Decoder | 384 × 384 | • | | Model performs poorly for multiple attribute manipulation |
| [63] | 2019 | Employed a fully end-to-end convolutional network SC-FEGAN using additional loss that uses free-form user input in the form of masks, color, and sketches as a guide to generate the photo-realistic images | ✘ | Any desired attribute | CelebA-HQ | PSNR = 27.9131, SSIM = 0.9543, L2 loss = 0.2, LPIPS = 0.0795 | PSNR = 31.1687, SSIM = 0.9671, L2 loss = 0.0937, LPIPS = 0.0552 | GAN | 512 × 512 | • | | Finer details of the skin texture are not preserved |

**Table 4** (continued)

| Ref | Year | Model description | Subject agnostic | Attributes manipulated | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | O/P media | | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Img/frame | Video | |
| [64] | 2020 | Employed a framework InterFaceGAN that allows interpreting disentangled face representation in linear subspaces that paves the way for manipulating gender, age, expression, and presence of eyeglasses via subspace projection | ✘ | Gender, age, expression, Eyeglasses, pose | CelebA-HQ(PGGAN) and FF-HQ(StyleGAN) | Re-scoring analysis = -024, Identity discrepancy = 0.01, Disentanglement analysis = -0.42 Class. Acc. = ~ 0.65 | Re-scoring analysis = 0.59, Identity discrepancy = 0.61, Disentanglement analysis = 1.00 Class. Acc. = ~ 0.10 | GAN | 224 × 224 | • | | - |
| [51] | 2019 | Employed a novel GAN based on a weakly supervised attention mechanism that takes annotated facial action units (AU) as input and generates a broader range of expressions | • | Facial Expressions | EmotioNet, RaFD, CelebA | Average Content Distance (ACD) = 0.4, User Preference = 0, Expression Distance(ED) = 4.8 | IS = 1.48, ACD = 0.0, User Preference = 87, Expression Distance(ED) = 0.4 | GAN | – | • | | Lack of adaption for various poses and gaze. The application of the model to the video sequence is still questionable |

**Table 4** (continued)

| Ref | Year | Model description | Subject agnostic | Attributes manipulated | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | O/P media | | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Img/frame | Video | |
| [65] | 2021 | Employed a two-stage framework for joint manipulation of pose and facial expression for high-resolution images. In the first stage, the boundary is predicted, and in the second stage, structure and texture are disentangled to generate the desired manipulation. A high-resolution database named MVF-HQ is proposed that contains 120,283 images at resolution 6000 × 4000 for 479 identities | • | Poses and expression | CelebA-HQ, RaFD, MulitPIE, MVF-HQ | FID = 38.65, Rank-1 Accuracies = 48.2, Rank-1 recognition rates = 60.4% | FID = 12.94, Rank-1 Accuracies = 60.4, Rank-1 recognition rates = 100% | Encoder-Decoder | 128 × 128, 512 × 512, 1024 × 1024 | • | | Method is sensitive to the variation in poses and illumination |

**Table 4** (continued)

| Ref | Year | Model description | Subject agnostic | Attributes manipulated | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | O/P media | | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Img/frame | Video | |
| [66] | 2021 | Proposed a domain-guided Noise-optimization based on an inversion approach for facial image attribute manipulation. First, they map the image into latent space, aligning the inverse code with semantic knowledge. Then, a noise optimization mechanism for noise addition to capture high-frequency details like edges and corners, finally, a mask fuses image and local style | ✘ | Any desired attribute | FFHQ, CelebA-HQ | MSE = 0.0054, RMSE = 0.0731. PSNR = 23.318, UQI = 0.9266, SSIM = 0.9980, MS-SSIM = 0.9998, VIF = 0.9012, Att. Gen Acc. = ~ 0.8 | MSE = 0.0043, RMSE = 0.0570. PSNR = 25.669, UQI = 0.9481, SSIM = 0.9978, MS-SSIM = 0.9998, VIF = 0.9535, Att. Gen Acc. = ~ 1.0 | GANs + Encoder | 1024 × 1024 | • | | Method takes a few minutes to optimize the semantics of the code, while the embedding algorithm is done in seconds. Also, the process is not so attractive for interactive editing of images |

**Table 5** Overview of popular Image Synthesis methods

| Ref | Year | Model description | Dataset used | Worst performance | Best performance | Architectural components | Output resolution | Limitation |
|---|---|---|---|---|---|---|---|---|
| [55] | 2019 | Implemented an enhanced unsupervised style transfer to disentangle high-level attributes from the stochastic variation. To further quantify the properties of disentanglement and interpolation, they have employed two automated methods, i.e., perceptual path length and linear separability | Real images collected from Flickr website(FFHQ dataset) | FID = 5.06, Perceptual path(full) = 283.5, Perceptual path(end) = 285.5 separability scores = 9.88 | FID = 4.40, Perceptual path(full) = 217.8, Perceptual path(end) = 195.9 separability scores = 3.79 | GAN | 1024 × 1024 | Limited resolution of the generated samples |
| [69] | 2019 | Proposed a self-attention mechanism into GANs framework that models long-range and multi-level dependencies and more refined details from several locations of an image. In addition, spectral normalization stabilizes GANs training and a two-timescale learning rate(TTUR) to speed up the discriminator training | ImageNet2012 | FID = 22.98, IS = 43.15 | FID = 18.28, IS = 52.32 | GAN | 128 × 128 | Undesirable artifacts are visible at various patches of the image |
| [70] | 2019 | Train the GANs on the large-scale ImageNet dataset to study the instabilities and then apply orthogonal regularization to the generator to make it amenable for truncation trick, allowing explicit finer control for the trade-off between variety and fidelity of the image synthesis | ImageNet, JFT-300 M | FID = ~ 39.7, IS = ~ 124.5 | FID = ~ 5.7, IS = ~ 298 | GAN | 128 × 128, 256, 256, 512 × 512 | Conditional class synthesis of image |
| [68] | 2020 | Redesigned the architecture of StyleGAN1 [55] with the normalization used in the generator, adapting the progressive GAN approach to increase the number of layers as the training progresses and regularizing the mapping of the generator from the latent code to images | FFHQ, LSUN CAR | FID = 2.84, Perceptual Path length (PPL) = 415.5, Precision = 0.678, Recall = 0.492 | FID = 2.32, Perceptual Path length (PPL) = 145.0, Precision = 0.689, Recall = 0.514 | GAN | 256 × 256, 512 × 384, 512 × 512, 1024 × 1024 | – |

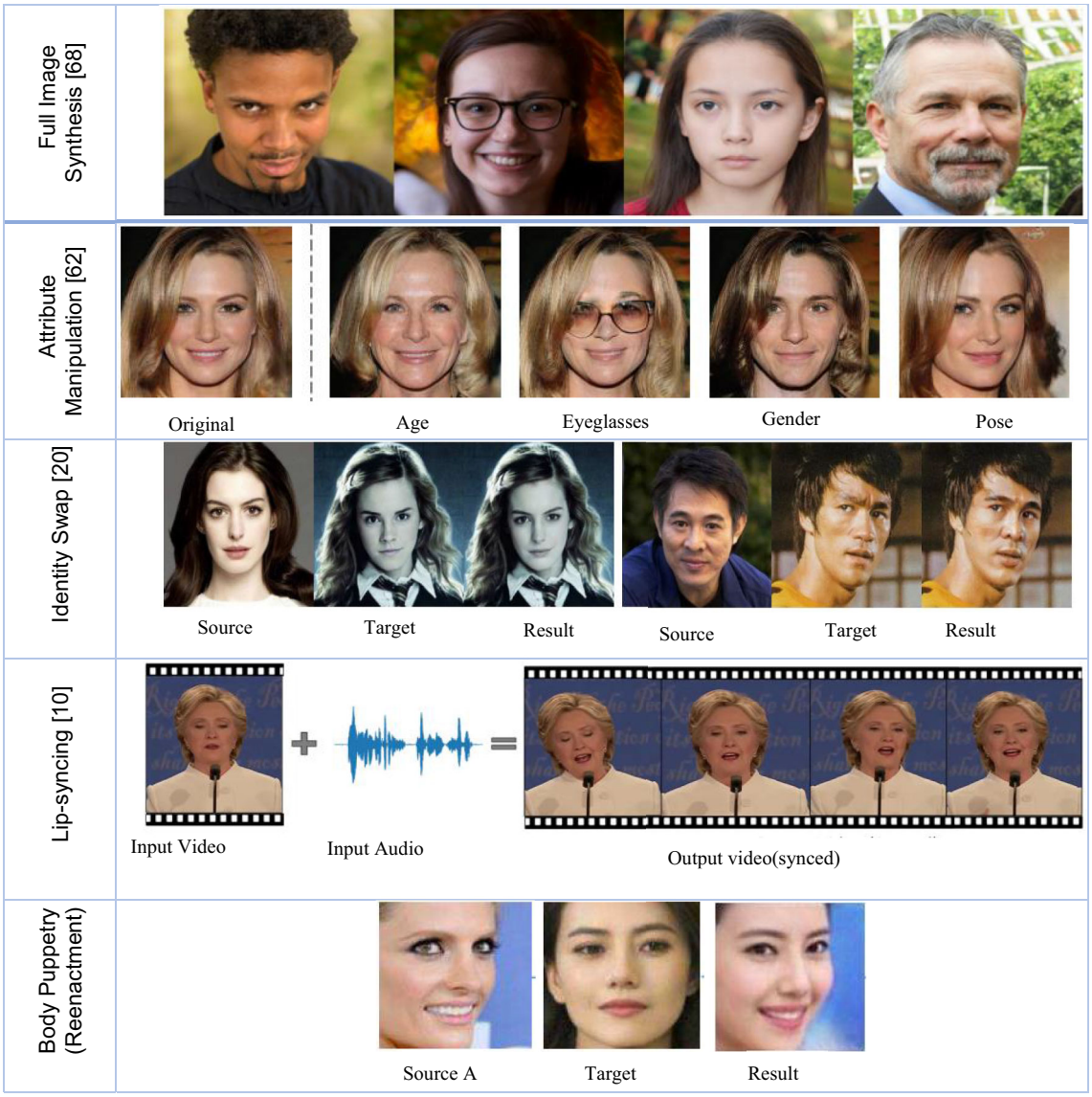



**Fig. 5** Types of visual deepfake manipulation

There has been a common trend that we can see that VAEs and GANs have been state-of-the-art generative models used more frequently. VAEs are often utilized for attribute manipulation, where the latent space is exploited to disentangle features and modify the attribute without affecting the other. RNNs like LSTM have been used for sequence data. Evaluating the fidelity of the generated results has always been an issue; most of the authors have adopted objective and subjective evaluation a great deal. For objective evaluation, FID, IS, SSIM, and PSNR have been used more frequently for evaluating the results. For subjective assessment, a human visual score has been adopted where the user is asked to rate the realistic nature of the visual results. Also, methods have a hard time finding a balanced trade-off between fidelity and variety of the generated samples. The CelebA dataset has been used most frequently, and some others even choose to take the dataset directly from the internet. Among the output resolution of the generated results, we can see that, on average, 256 $\times$ 256 is the most generated output of the models, suggesting that models struggle with high-resolution visual samples. Although few methods like PGGAN [67], and StyleGAN1 [55], can have high resolution and realistic content. The most common limitation that we encounter is that the generated results are not generalizable to other identities; they need to re-train their model for a new identity. Also, the generation mechanism has to go a long way before they provide real-time manipulation with realistic results. There have been few methods for Lip-sync and entire image synthesis among the five categories of manipulation (Fig. 5).

#### 2.1.2 Audio deepfake generation

The deepfake generation mechanism has also extended to audio manipulation. This manipulation form clones someone's voice or generates a speech from the text using audio samples. It just needs someone's audio samples to create the audio of the people, which they have never said. Speech synthesis is divided into two categories:

**Text to speech (TTS)** TTS synthesis refers to generating speech from the text in someone's voice. Such systems require a database of audio samples of the speaker to train the system; once the system is trained, it can generate the audio sample in the speaker's voice from the input text. TTS systems have become indispensable components of our daily lives. It is extensively used in human-technology interfaces like virtual voice assistants (e.g., Google Assistant, Apple's Siri, Microsoft's Cortana, & Amazon's Alexa), GPS navigation systems, speech-to-speech translation across multiple languages, and screen readers [71]. Such systems are beneficial for visually impaired people to access the technology without any aid. On the other side, the easy availability of TTS synthesis methods can be misused to generate someone's voice without their consent.

**Basic pipeline of TTS systems:** The input to the model is the text that goes through stages and finally generates an audio waveform (Figure 6) [72]. Different stages have different functions:

**Preprocessor:** This stage takes text as an input and breaks it into phonemes (linguistic features) based on pronunciation. From the phonemes, phoneme duration, pitch and energy (indicates the magnitude of Mel-spectrograms) are used to train the corresponding components of the systems to get the more natural output of voice [72].

**Encoder:** An encoder takes the linguistic features and converts them into n-dimensional latent feature embeddings. The speaker encoder also converts the speaker's voice into its embedding, which is concatenated with phonemes to get latent features representation of text and audio [72].

**Decoder:** At this pipeline stage, the decoder converts the latent feature into an acoustic feature, a Mel-spectrogram representation of audio. Mel-spectrogram is obtained from a short-time Fourier transform by applying a nonlinear transformation to the frequency axis [72].

**Vocoder:** It converts the Mel-spectrogram representation of audio into waveform. There are different ways to do it. It can be done using Griffin Lim methods or a neural network to map Mel-spectrogram to the waveform. Nowadays, neural network techniques outperform the Griffin Lim method [72].

TTS systems have been developed using two major approaches: concatenative and parametric approaches. In concatenative TTS, the process is based on collecting and fragmenting the good-quality speech recordings into the units. Then concatenating fragmented units into their corresponding text to generate the speech [74]. However, this approach produces the glitch text for the unseen part of the speech, and a new modification requires an entirely new dataset collection. In parametric TTS, the aim is to use the physical parametric model, usually, a function, to reenact the vocal tract of a human being and then adapt those parameters using recorded speech [74]. Such models allow modification of the voice of speech by tuning the parameters quickly. After the recent advances in deep learning techniques, TTS systems can easily modify the speech and produce a more natural speech. WaveNet [75] was one such kind developed by the researchers at DeepMind in 2016 using raw audio waveforms. The auto-regressive probabilistic model uses the previous audio sample to predict future audio, trained on thousands of audio samples of different subject identities to generate different voices. It produces more natural speech than the various state-of-the-art methods, and the model shows promising results when applied to other audio domains like music. However, WaveNet has a limitation of its prohibitive sequential speed that enables it to generate one audio sample at a time. Hence, it cannot be deployed in real-time settings requiring parallel data generation. Parallel WaveNet [76] was proposed to counter the prohibitive sequential speed of the Base WaveNet model. The model uses the probability distillation method to train the parallel feed-forward network using the sequential trained WaveNet model, generating high-fidelity audio samples at 20 $\times$ more speed than the usual WaveNet model.

Deepvoice1 [77] model, based on a deep neural network, is known for its production quality TTS systems. The framework with five building blocks requires few parameters for manual annotation, trains faster, and can work as standalone systems without the pre-existing model. DeepVoice2 [78] was another deep learning model developed in succession for multi-speaker voice systems to counter the limitation of single-speaker voice. The model uses the combination of Deep voice 1 & Tactron model, where each speaker's embedding is represented in low-dimensional space, and the model's parameters are shared among each other. Finally, deep voice3 [79] is a full sequence-to-sequence conventional acoustic model agnostic of the waveform synthesis methods that have been scaled to over two thousand recordings.

In 2017, Google introduced Tacotron [80], an end-to-end speech synthesis method from a given text. The model can be trained with random initialization using a < text, audio > pair and output spectrogram. However, Tacotron suffers from the flaw introduced due to the difference between the training and test processes, which affects the synthesis quality and length of the generated frames. Tacotron2 [84] uses an unsupervised GAN model to tackle the problem. The model introduced a training strategy called "random down" and

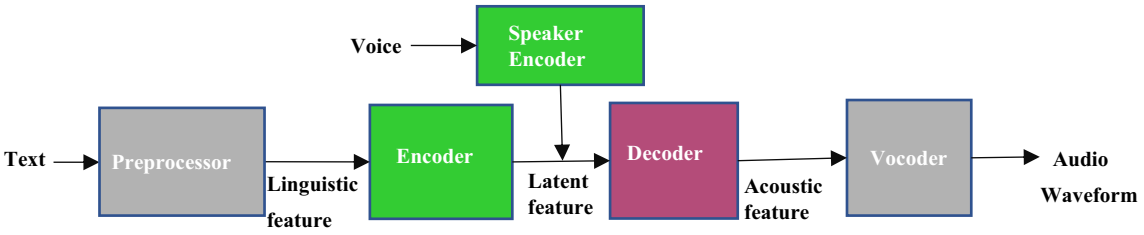



**Fig. 6** High-Level diagram of different components used in TTS systems [73]

added the window weights to take care of the minimum values of the attention weight. Table 6 gives an overview of current text-to-speech (TTS) methods.

**Voice conversion (VC)** Voice conversion is another category of speech synthesis, which deals with converting the source speaker's voice to another form as if uttered in the target speaker's voice while maintaining their linguistic content. This system can be applied as a component in personalized TTS systems for voice conversion [81]. Apart from that, it can be used for impersonating or hiding a person's speech, communication aid for speech-impaired people [82], speech enhancement/accent conversion in language learning and voice dubbing in movies [83]. Voice conversion has been one of the most active research areas in the past few years. Zhang et al. [84] perform non-parallel voice conversion methods by disentangling the linguistic information of the speech from the speaker's characteristics. Voice conversion methods generally require a large dataset for the source and target speaker to learn the mapping. However, few methods have been developed in the last few years that require less data and can even perform voice conversion for unseen voices. Wang et al. [85] propose a StarGAN framework that uses an incremental approach for voice conversion for the unseen dataset. Liu et al. [86] use a transformer network, which performs end-to-end voice conversion, based on three ideas: transformer, context preservation and model adaptation.

**Basic voice conversion pipeline:** The basic pipeline for VC, shown in Fig. 7, is also called as analysis-mapping-reconstruction pipeline [83]. Linguistic factors, supra-segmental speech factors (prosodic characteristics of speech), and segmental factors characterize the speaker's speech. During the training, mapping is done between the source and target speech features represented in the function form. At the inference time, source speech is given, decomposed into supra-segmental and segmental features by the speech analyzer module, and then features are extracted [83]. Then, in the mapping module, mapping is done between the source speech features and the corresponding target features learned during the training time. Finally, the reconstruction module synthesizes speech signals [83]. Table 7 briefly describes the Voice Conversion(VC) approaches.

**Summary of the audio deepfake generation methods** Various methods have been proposed for text-to-speech (TTS) and voice conversion. There are a few standard things about such approaches. CMU artic, LJspeech and VCTK database are the most preferred dataset among the research community for their models. Encoder–decoder architecture components are widely used, given the bottleneck it provides to alter the embedding of the speech to get the desired results. Like visual deepfake, audio deepfake has an issue with evaluating the generated samples; there has not been a consensus among the community about the metric to be used for the evaluation. That is why authors have preferred both subjective and objective assessments to prove their results. MCD, RMSE, and cosine similarity are the most frequently used metrics for objective results. For subjective evaluation, MOS is used, where a user rates the audio on various parameters like naturalness, similarity, emotional, etc. The most significant limitation that the methods suffer is that they fail to produce the natural and emotions contained voice. Also, the technique is not generalizable to various languages, and models also are data-intensive and need much data before generating results. As a result, methods struggle to generate results for low-quality and noisy data samples.

## 2.2 Deepfake detection

There is an arms race between manipulators and the detector; the detector uses some clue to find the detection, and the manipulator tries to diffuse it next time to make it detection-proof. Generalizing the deepfake technique, robustness against various post-processing operations and interpretability of the detection results are three critical factors for a detector to be deployed in the wild [11].

Visual deepfake and audio deepfake are the two different media for which different detection algorithms have been designed. Based on the clues/traces of feature representation, these two categories have been divided further, as shown in Fig. 8.

**Table 6** Overview of different Text-to-Speech methods proposed from 2019 onwards

| Methods | Year | Model description | Dataset | Architectural Components | Worst Performance (Evaluation Metrics) | Best Performance (Evaluation Metrics) | Limitations |
|---|---|---|---|---|---|---|---|
| [87] | 2019 | Extended the Tactron system to Japanese speech synthesis systems with a self-attention mechanism to capture long-term information on pitch accent | Japanese Speech Corpus | LSTM + Encoder-Decoder | MOS(naturalness) = 2 | MOS(naturalness) = 5 | The absence of word-level info in linguistic features and also model lacks real-time speech synthesis |
| [88] | 2019 | Proposed an adaptive model for TTS based on the meta-learning concept that uses WaveNet as a core model. The model first learns the speaker's embedding, then fine-tunes the architecture, and then predicts the embedding of the speaker | LibriSpeech & VCTK | Encoder + Deep neural networks | MOS = ~ 4.01, EER = 7.34, Cosine Similarity = 0.10 | MOS(naturalness) = ~ 4.12, EER = 1.85, Cosine Similarity = 0.75 | Performance degradation for a low-quality noisy audio sample. The model requires a lot of high-quality training data |
| [89] | 2019 | Introduced the unsupervised GAN model that optimized the prosody of the generated speech and incorporated a training strategy called "random down' to alleviate the cumulative error problem of Tactron2 | Text taken from the web | GANs | Graphical evaluation | | Method is limited to the input text length of 1000 |
| [90] | 2021 | Extending the architecture model of the tacotron by including the normalizing flow into the auto-regressive model. The sequential decoder model takes the input, produces conditioning features for a normalizing flow, and generates output waveforms containing various interdependent samples | Lj speech & a proprietary dataset | Vocoder with Attention mechanism | MCD = 6.87, MSD = 9.24, CER = 9.2, MOS = ~ 3.56 | MCD = 4.64, MSD = 11.44, CER = 9.4, MOS = ~ 4.23 | Model is computationally complex |

**Table 6** (continued)

| Methods | Year | Model description | Dataset | Architectural Components | Worst Performance (Evaluation Metrics) | Best Performance (Evaluation Metrics) | Limitations |
|---|---|---|---|---|---|---|---|
| [91] | 2021 | Proposed a multi-rate attention architecture to cope with the speech generation mechanism's low latency and slower synthesis speed. The encoder uses the parallelizable approach computing the compact multi-headed representation in a streamlined manner and then performing dynamic pooling to restrict the length to produce high-quality content | TTS dataset recorded in a voice production studios | CNNs with Attention Mechanism + •LSTM + Vocoder | MOS = ~ 4.08 | MOS = ~ 4.31 | - |
| [92] | 2021 | Proposed a novel training strategy to capture the association between input text and its corresponding prosody styles using Tacotron-based systems, improving speech's expressiveness. The supervised system training includes a fully differential perceptual and frame reconstruction loss that does not require any reference speech or manual style selection during execution | IEMOCAP, LJ-Speech database | LSTM with Attention Mechanism + Decoder + CNNs | MCD = 7.01, RMSE = 1.53, FD = 15.59, MOS = 3.63 | MCD = 6.37, RMSE = 0.94, FD = 13.96, MOS = 4.19 | Approach experimentation is limited to the Tacotron systems |
| [93] | 2021 | Proposed a framework for a sequence-to-sequence model for concatenative speech synthesis. The encoder has the same design as Tacotron2, while the decoder has been redesigned. The decoder contains an auto-regressive mechanism at both frame-level and phoneme levels that learn the mapping between phone sequences and acoustic using the transition probabilities | Chinese corpus | Encoder + Decoder-LSTM with Attention Mechanism | MSD = 3.70, MCD = 3.64, RMSE = 39.84, CORR = 0.89, UV = 7.31, Preference sore = 51.0 | MSD = 3.64, MCD = 3.46, RMSE = 37.76, CORR = 0.89, UV = 7, DRMSE, Preference sore = 55.0 | Method's experimentation is limited to the Chinese language corpus |

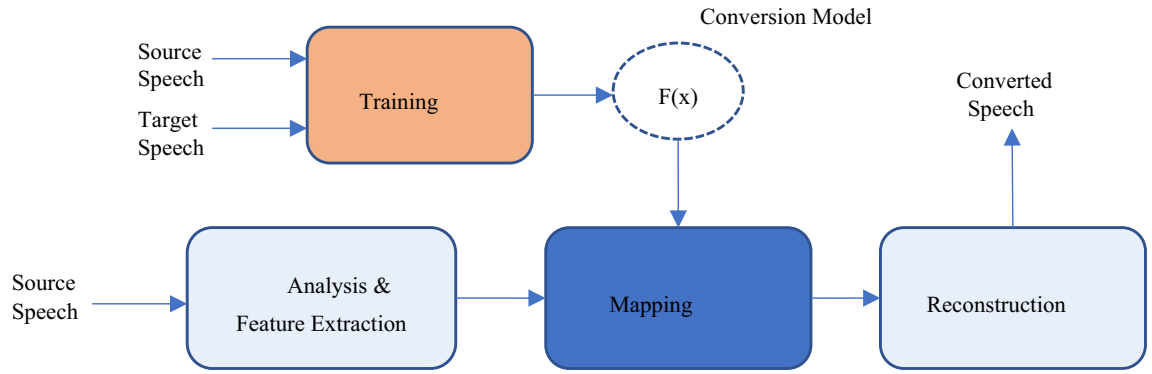



**Fig. 7** Typical flow diagram of Voice Conversion system [83]

### 2.2.1 Deepfake visual detection

Manipulated or manufactured samples have a certain peculiarity in spatial, temporal or frequency domain representation, which the different detection algorithm exploits. The subsequent sections will discuss the various detection techniques along with other domains.

**Spatial domain-based detection** The change in pixel distribution of manipulated samples can be reflected in the spatial domain properties. This section will discuss various types of spatial domain-based detection methods. Table 8 highlights the recent spatial domain-based detection approaches.

**Forensics based detection** Generation methods leave certain clues or traces that change the distribution of the samples, which is exploited by the detection methods by analyzing latent features and patterns. Li et al. [105] analyze the subtle distribution of the image statistics in the chrominance components of YCbCr and HSV color spaces, especially in the residual domain for the unseen DNG images detection. Chen et al. [106] use a multi-domain architecture where the features from the RGB domain and the noise vectors (which get added by the external mechanism) are fused to obtain richer robust features.

There are various variants of Forensics features are mentioned below:

- **GAN-Artifacts-based artifacts:** The imperfect design of the GANs leaves some traceable clues, which various researchers use for fake detection investigations. McCloskey et al. [107] utilized the prior knowledge about how the color is treated in GAN and camera models and used this knowledge as a cue to design the network. Methods perform exceptionally for the existing GAN model; however, the model's performance is unclear for new advanced GANs. Yu et al. [108] identify a unique GAN stable fingerprint that persists across different frequencies and patches of the generated image extracted by a neural network. However, the method fails for post-processing operations like compression, blur etc.
- **PRNU Noise-based detection:** Photo response non-uniformity (PRNU) is a noise-like pattern in the digital image caused by the camera's light sensor. Koopman et al. [109] proposed a method where Co-relation scores are computed between every eight groups of PRNU pattern frames, which serve for deepfake video detection. However, their evaluation is limited to the small dataset. PRNU based are generally low-cost based methods and have a high generalization capability.

**Visual-artifact based detection** The synthesized or manipulated faces would reveal inconsistencies in the appearance, especially the blending boundaries, landmark points or the shape of the manipulated facial attributes. Even sometimes, the content of the face also seems strange to the rest of the background. Li et al. [110] use the face warping artifact as a clue caused due to blending operation to match the configuration of the source face. Unfortunately, this affine warping operation leaves the artifact due to resolution inconsistency between the face and the surrounding area. The method is more robust than other existing methods, but there is still room for improvement.

Visual-Artifact based methods can achieve better generalization as they pay more attention directed to the specific artifacts. Li et al. [111] propose a generalized detector that uses face X-ray and blending boundaries for detection purposes. However, such a method fails for entirely synthetic images, indicating the blending operation's absence. Also, adversarial samples can be designed to bypass the detection mechanism by curbing such manipulation clues. Matern et al. [112] also exploited the visual artifacts like the difference in eye color, inconsistent illumination, missing teeth areas, etc.; once the algorithm extracts the artifacts, they are used for classification. These methods can localize the manipulation easily, as their detection is based on specific localized

**Table 7** Overview of various Voice Conversion(VC) methods proposed from 2019 onwards

| Ref | Year | Model description | Dataset | Architectural components | Worst Performance (Evaluation Metrics) | Best Performance (Evaluation Metrics) | Limitations |
|---|---|---|---|---|---|---|---|
| [94] | 2019 | Proposed a sequence-to-sequence learning-based method (ATTS2S-VC) that uses context preservation and attention mechanism for voice conversion tasks | CMU Arctic Database | Encoder-Decoder | Preference test(naturalness) = ~ 30% Preference test(speaker similarity) = ~ 25% | Preference test(naturalness) = ~ 60% Preference test(speaker similarity) = ~ 67% | Method is mainly limited to the supervised learning settings |
| [81] | 2019 | Proposed a Sequence-to-Sequence conversion model (SCENT) for acoustic feature modeling for voice conversion tasks. The model uses an attention mechanism that implicitly aligns the source and target acoustic feature sequences. At the conversion stages, simultaneous conversion of the acoustic features and duration happens, and finally, the WaveNet vocoder converts the output of the audio model into audio waveforms | Mandarin and CMU Arctic Database | Encoder-Decoder with Attention Mechanism + Vocoder | MCD = 3.802, RMS of $F_0$ = 41.748, Duration differences of converted and target utterances (DDUR) = 0.260, MOS = 3.66 | MCD = 3.212, RMS of $F_0$ = 9.899, Duration differences of converted and target utterances(DDUR) = 0.147, MOS = 4.01 | Model is computation expensive, requires a lot of training data, and does not generalize well to the unseen data |
| [95] | 2019 | Proposed non-parallel voice conversion methods using an auxiliary classifier VAE. First, the model uses a fully convolutional network to construct an encoder-decoder network to capture long-term time dependencies of acoustic feature sequences of speech. Also, the network uses information-rhetoric regularization during training to ensure that attribute class labels are preserved during the conversion process | Voice conversion Challenge(VCC) dataset 2018 | Variational Auto-encoder(VAE) | MCD = ~ 7.01, MOS = 1.75, ABX test for speaker similarity = 25% | MCD = ~ 6.29, MOS = 4.7, ABX test for speaker similarity = 75% | Method require a lot of training data |
| [96] | 2020 | Proposed a data-efficient approach to disentangle speech from the noise using Domain Adversarial Training (DAT) for speaker adaptation & speaker encoding methods | MULTI-SPK & CHiME-4 | Encoder-Decoder + RNNs | MOS(naturalness) = 3.30, MOS(similarity) = 2.96,MCD = 5.32, Cosine Similarity = 0.85 | MOS(naturalness) = 3.67, MOS(similarity) = 3.72,MCD = 3.66, Cosine Similarity = 0.96 | Sensitive to complicated acoustic conditional scenarios, e.g., room reverberations. However, the model also lacks in synthesizing utterances of a target speaker |

**Table 7** (continued)

| Ref | Year | Model description | Dataset | Architectural components | Worst Performance (Evaluation Metrics) | Best Performance (Evaluation Metrics) | Limitations |
|---|---|---|---|---|---|---|---|
| [97] | 2020 | Proposed a method named Deep conversion for voice conversion that requires a few samples for parallel data processing. Non-parallel data is used to train the general independent speaker model, while parallel data adapts the generalized model towards finer voice conversion between the source and the target speaker | CMU Arctic dataset | LSTM + Vocoder | MOS = 2.09, Preference test = ~ 20%, MCD = 9.32 | MOS = 4.63, Preference test = 75%, MCD = 4.11 | Model requires training the module separately, so it is computationally expensive |
| [98] | 2020 | Proposed a method named ConvS2S-VC for voice conversion based on an entirely conventional network that provides the flexibility of many-to-many, any-to-many real-time natural conversion of the fundamental frequency ($F_0$), speaker rate, rhythm, and speaker characteristics of the speech | CMU Arctic Database | Encoder-Decoder | MCD = ~ 7.08, Log $F_0$(LFC) = 0.771, LDR = 10.85, MOS(speaker simi.) = 2.9 MOS(sound quality) = 2.5 | MCD = ~ 5.78, Log $F_0$(LFC) = 0.877, LDR = 0.79, MOS(speaker simi.) = 3.5 MOS(sound quality) = 3.9 | Method is limited to English language conversion |
| [99] | 2020 | Proposed a method based on Phonetic Posteriorgrams (PPGs) to remove speaker-dependent information in VC models. During the adversarial training, the output of the encoder is fed to the speaker classifier to differentiate the corresponding speaker. Simultaneously, the encoder is optimized to fool the classifier, enabling the encoder to become more speaker-Independent | VCTK Corpus, LibriSpeech corpus and VoxCeleb2 dataset | ED with Attention mechanism + RNN vocoder | MCD = ~ 9.31, VSS(Voice Similarity Score) = ~ 4.80, MOS = ~ 3.77 | MCD = ~ 8.37, VSS(Voice Similarity Score) = ~ 5.02, MOS = ~ 3.86 | – |
| [85] | 2020 | Proposed a StarGAN framework for voice conversion for the target and source speaker unseen in the training dataset. The incremental training approach represents each speaker ID with an embedding vector for voice conversion | VCTK, CMU-Arctic & THCHS30 | GANs | MOS(reconstruction) = 1.32, MOS(conversion) = 2.89 | MOS(reconstruction) = 3.00, MOS(conversion) = 4.00 | Model is limited to the processing of two different timbers for voice conversion |

**Table 7** (continued)

| Ref | Year | Model description | Dataset | Architectural components | Worst Performance (Evaluation Metrics) | Best Performance (Evaluation Metrics) | Limitations |
|---|---|---|---|---|---|---|---|
| [86] | 2020 | Proposed an end-to-end voice conversion system based on three concepts: transformer, context preservation and model adaption. The transformer connects all positions, enabling the network to learn long-term dependencies. The context preservation mechanism accelerates the training, while the adaptation mechanism empowers the training to conduct smoothly with limited training data | Mandarin dataset | Transformer Networks + WaveNet vocoder | MOS(naturalness) = ~ 3.84, MOS(similarity) = ~ 4.11 | MOS(naturalness) = ~ 4.29, MOS(similarity) = ~ 4.31 | Method suffers from low inference and sometimes produces critical errors |
| [100] | 2020 | Extend the network of Cycle-GAN to be used for multi-speaker by using conditional cycle-GAN. The proposed network uses a single generator and discriminator conditioned on a pair of source and target identity vectors to steer the network to synthesize the speech for the target speaker | VCC2018 corpus | GANs | MCD = ~ 7.85, MSD = ~ 1.83, MOS(Sound quality) = ~ 3.00, MOS(Similarity test) = 28% | MCD = ~ 7.25, MSD = ~ 1.89, MOS(Sound quality) = 3..92, MOS(Similarity test) = 69% | Method does not work well for the unseen data |
| [101] | 2021 | Proposed a novel voice-conversion-based network using the transfer learning mechanism from TTS-based systems. The encoder takes the speech as an input instead of text, while the decoder has similar functionality to TTS systems. However, it is conditioned on speaker embedding to be trained for non-parallel data for any-to-any voice conversion. While training text or speech can be an input to the TTS and VC network, the VC network uses speech as an input during run-time | train-clean-360, VCC2018, AISHELL2 | Encoder-decoder with Attention mechanism | MCD = 10.38, RMSE = 0.66, AB preference test = ~ 25%, XAB preference test = ~ 25%, BWS test = 4.44, MOS = 2.39 | MCD = 3.31, RMSE = 0.27, AB preference test = ~ 95%, XAB preference test = ~ 90%, BWS test = 43.05, MOS = 4.64 | Method's performance is unknown for the compressed and noisy voice samples |

**Table 7** (continued)

| Ref | Year | Model description | Dataset | Architectural components | Worst Performance (Evaluation Metrics) | Best Performance (Evaluation Metrics) | Limitations |
|---|---|---|---|---|---|---|---|
| [102] | 2021 | Proposed a model to improve the data efficiency of many-to-many StarGAN-based models for voice conversion. To increase the relative number of speakers with limited training data, use a speaker-encoder to extract the speaker embedding from the target speech and then use StyleGAN2 weight adaptive instance normalization (W-AdaIN) on the convolutional layers | VCTK dataset | GANs + Encoder | Accuracy = 97%, EER = 0.66, AB test(naturalness) = 13.3, XAB(similarity) = 36.1 | Accuracy = 97%, EER = 0.66, AB test(naturalness) = 13.3, XAB(similarity) = 36.1 | Method cannot generate the voices for unseen samples without requiring a lot of such data to train the network |
| [103] | 2021 | Uses a method to convert the voice of songs from one singer to the other. The model takes phonetic posterior grams (PPGs) as input and uses an encoder to encode the PPGs and an additional encoder to compress Mel-spectrograms to obtain acoustic and musical information. Furthermore, it uses a singer confusion module and a mel-regressive representation module to improve the melody and timbre performance | Chinese Corpus | Encoder-decoder + vocoder | MOS(similarity) = ~ 3.42, MOS(naturalness) = ~ 3.64, Normalized cross-relation(NCC) = 0.889 | MOS(similarity) = ~ 3.57, MOS(naturalness) = ~ 3.75, Normalized cross-relation(NCC) = 0.902 | Model is data-intensive and uses a lot of data to generate good-quality voices |
| [104] | 2021 | Extending the framework of the Voice Transformer Network (VTN) for voice conversion to simultaneously learn the mapping among different speakers by extracting the various common latent features. In addition, the identity mapping loss is proposed to ensure the input feature sequences remain the same between the source and the target speaker indices, which eventually improves the model's performance at run-time | CMU Arctic Database | Transformer Net | MCD = 6.77, LFC = 0.717, LDR = 2.34, MOS = ~ 3.5, speaker similarity score = ~ 3.5 | MCD = 5.91, LFC = 0.835, LDR = 5.41, MOS = ~ 3.5, speaker similarity score = ~ 3.5 | Model is computationally expensive and requires different modules to train independently |

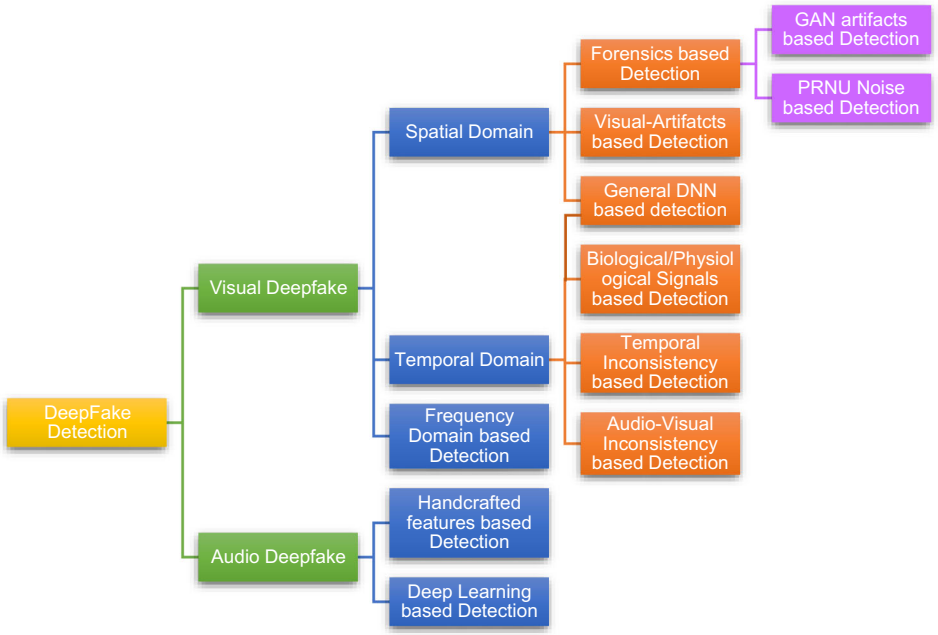



**Fig. 8** Classification of deepfake detection methods based on feature representation

artifacts. Some other techniques ( [113, 114, 115, 116]) also have been proposed, which look for visual clues for deepfake detection.

**Temporal domain-based detection** *Temporal information* is the sequential info that is relatable, coherent and changes with time. Manipulated artifacts or traces could appear as flickering/jittering. The subsequent sections will discuss various methods that investigated the temporal domain to find such clues. Table 9 gives a brief overview of temporal domain-based detection approaches.

**Audio-visual inconsistency based detection** For lip-sync methods, inconsistency between the mouth region (visual) and audio is one distinguishing factor for deepfake detection. Agarwal et al. [117] use a CNN to detect inconsistent mouth features, i.e., the shape of the mouth (visemes) is not aligned with spoken words (phenomes) in a manipulated video. They focused on the visemes related to the words M, B and P, in which the mouth almost gets completely closed. However, their method is specific to the videos of Barack Obama. Mittal et al. [118] simultaneously exploited visual & audio modalities and perceived the affective cues to detect any alteration. This exploration between these two modalities uses the Siamese network, where triplet loss is used to measure the similarity. Sometimes, this method cannot detect any manipulation if there is a similarity between these perceived affective cues. Moreover, this method is limited to a single person in a video. Chugh et al. [119] proposed an approach to calculate disharmony score between visual and audio modality and termed it Modality Dissonance Score(MDS). This dissimilarity score is calculated chunk-wise per video segment. The contrastive loss is employed to calculate such inter-frame modality similarity. However, these methods rarely look into visual consistencies, which could also be faked. It remains unclear whether such methods can be deployed in real-world scenarios where one may encounter any manipulation and undergo different post-processing operations.

**Temporal inconsistency based methods (TI)** In manipulating video frames, correlation structures between the frames are sometimes destroyed, reflecting in various forms like video flickering or shifting of the facial content [12]. Usually, sequence models like RNNs and their variants get employed to find such inconsistencies between the frames. Hosier et al. [120] use video speed manipulation as a temporal feature for detection, and the encoding used for each frame gives an idea about the number of deleted and added frames. Methods that use frame-level artifacts and temporal features for detection perform better than those focusing on either of the two. Guera et al. [1] propose a temporal-aware CNN-LSTM framework, which exploits the frame-level features and temporal inconsistency between frames. Temporal inconsistency introduced by the auto-encoder (used for face swapping) focuses on the face-swapping process unaware of inconsistency introduced by the process, which results in anomaly serving as crucial evidence for detection.

An optical flow mechanism has been used to estimate the per-pixel motion behavior of the adjacent frames. Amerini et al. [121] used a pre-trained model (trained on RGB images) with an optical flow mechanism to capture the dissimilarity

**Table 8** Overview of Spatial domain-based deepfake detection methods proposed from 2019 onwards

| Methods | Year | Key features | Dataset media | | Method types | Local. | Input resolution | Architectural components | Database | Worst performance | Best performance | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Img | Vid. | | | | | | | | |
| [105] | 2020 | Disparities in the chrominance components of YCbCr and HSV color spaces, especially in the residual domain | • | | Forensics based method | ✘ | 128 × 128 | SVM | Real images (CelebA, LFW, and FFHQ) but fake images are generated by different GAN models | Acc. = 51.85% FNR = 41.53% FPR = 96.29% (on mismatched image sources) | Acc. = 99.99% FNR = 0.0% FPR = 0.0% (on mismatched image sources) | Performs poorly for real-world scenarios |
| [106] | 2021 | RGB domain and noise feature vectors | | • | Forensics based method | ✓ | 640 × 640 | CNNs | FaceForensics + + | AUC = 65.30% | AUC = 99.99% | Performs poorly for the unseen type of face manipulation methods |
| [108] | 2019 | GANs fingerprint persisted across different frequency | • | | Forensics(GAN artifact) | ✘ | 128 × 128 | CNN | CelebA, LSUN | Accuracy = 97.04% FD Ratio = 6.27 | Accuracy = 99.93% FD Ratio = 454.76 | Performs poorly on post-processing operations |
| [163] | 2019 | Suppressed saturated and under-exposed pixel frequency to amplify GAN traces | • | | Forensics(GAN artifact) | ✘ | – | SVM | Media Forensics Challenge 2018 datasets, CelebA HQ | AUC = 0.61 | AUC = 0.92 | Does not consider the entire spectrum of GAN-generated images |
| [164] | 2020 | Convolutional traces of the image generation process | • | | Forensics(GAN artifact) | ✘ | 216 × 216, 256 × 256, 1024 × 1024 | KNN, SVM, LDA | Real Images(CelebA), Fake images(AttGAN, GDWCT, StarGAN, StyleGAN, StyleGAN2) | Acc.(KNN) = 71.17% Acc.(SVM) = 76.50% Acc.(LDA) = 76.00% | Acc.(KNN) = 99.61% Acc.(SVM) = 99.81% Acc.(LDA) = 99.61% | Not robust enough to the deepfakes generated using standard image editing(e.g., compression, photometric) operations |

**Table 8** (continued)

| Methods | Year | Key features | Dataset media | | Method types | Local. | Input resolution | Architectural components | Database | Worst performance | Best performance | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Img | Vid. | | | | | | | | |
| [165] | 2020 | GAN traces | • | • | Forensics(GAN artifact) | ✘ | 224 × 224 | CNNs(ResNet) | FaceForensics + + & other image are generated by different GANs | mAP = 81.3 | mAP = 93.0 | Not generalize enough to real-world images |
| [166] | 2021 | PRNU features are extracted as spectral and spatial features | | • | Forensics(PRNU) | ✘ | 10 × 10 | SVM | FaceForensics + +, Deepfake-TIMIT, DeepfakeDetection | Accuracy = 60% | Accuracy = 90% | Does not outperform the state-of-the-art deep learning-based methods |
| [110] | 2019 | Warping artifacts of the synthesizing methods | | • | Visual artifact | ✓ | 224 × 224 | CNNs | UADFV & DeepfakeTIMIT | AUC = 93.2% | AUC = 99.9% | Not robust enough for multiple video compression |
| [112] | 2019 | Visual inconsistencies of eye color, illumination inconsistencies, facial textures & geometry | • | | Visual artifact | ✘ | – | kNN, LogReg & MLP classifier | CelebA, FaceForensics | AUC(Gen. faces) = 0.704 AUC(Deepfake) = 0.402 AUC(Face2face) = 0.654 | AUC(Gen. faces) = 0.852 AUC(Deepfake) = 0.851 AUC(Face2face) = 0.866 | Limited to the specific dataset of face images with open eyes and clear teeth |
| [111] | 2020 | Blending boundaries of face manipulation | • | • | Visual artifact | ✓ | 64 × 64 | CNNs(HRNet) | FaceForensics + +, Celeb-DF, Deepfake Detection Challenge dataset | AUC = 70.56% AP = 68.99% EER = 32.62% Acc. = 85.69% | AUC = 99.31% AP = 93.34% EER = 8.37% Acc. = 97.73% | Unable to detect the entire synthetic image |
| [167] | 2020 | Discrepancies of the tightly bounded face and its context | • | • | Visual artifact | ✘ | 299 × 299 | CNNs | FaceForensics + +, DFDC, Celeb-DF-v2 | AUC = 66% Accuracy = 75.19% | AUC = 99.7% Accuracy = 75.49% | Performs badly for the low contrast and Blurry features samples |

**Table 8** (continued)

| Methods | Year | Key features | Dataset media | | Method types | Local. | Input resolution | Architectural components | Database | Worst performance | Best performance | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Img | Vid. | | | | | | | | |
| [168] | 2021 | Subtle texture differences of the image saliency | | • | Visual artifact | ✘ | 299 × 299 | CNNs | FaceForensics + + | Accuracy = 0.9926 | Accuracy = 0.999 | Computationally expensive and performs poorly for the data outside the training data scope |
| [169] | 2021 | Angular symmetrical differences in faces | • | • | Visual artifact | ✘ | 64 × 64 | CNNs | DF-TIMIT, FaceForensics + +, DFD, DFDC & Celeb-DF | AUC = 0.552 | AUC = 1.00 | May not work for the entire synthetic frames with excellent angular symmetry |
| [170] | 2021 | Landmark detection over an extended range | | • | Visual artifact | ✘ | 384 × 384 | Transformer-encoder | Celeb-DF, DFDC, FaceForensics | Accuracy = 74.35%, AUC = 92.34% | Accuracy = 99.2%, AUC = 94.96% | Not generalizable to the unseen type of face manipulations |
| [171] | 2021 | Multi-scale texture differences | | • | Visual artifact | ✘ | 224 × 224 | Encoder-decoder | FF ++, Celeb-DF, Deeperforensics, DFDC, Celeb-DF | AUC = 0.9579, Acc. = 55.46% | AUC = 0.9999, Acc. = 99.86% | Not generalizable to the unseen data |
| [172] | 2021 | Textures dynamics along spatial and temporal dimensions | | • | Visual artifact | ✘ | - | SVM | FaceForensics + + | Accuracy = 72%, AUC = 0.80% | Accuracy = 89.43%, AUC = 0.94% | Unsatisfactory results on the highly compressed videos |
| [173] | 2021 | Multi-scale Texture difference using pixel intensity and pixel gradient information | | • | Visual artifact | ✘ | 224 × 224 | CNNs | Ff ++, DFDC, DeeperForensics1.0, Celeb-DF | Accuracy = 53.43%(cross-database) | Accuracy = 99.99%(cross-database) | Requires to re-train for unknown face manipulation |

**Table 9** Overview of Temporal domain-based deepfake detection methods proposed from 2019 onwards

| Methods | Year | Key features | Dataset media | | Method types | Local. | Input resolution | Architectural components | Database | Worst performance | Best performance | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Img | Vid. | | | | | | | | |
| [123] | 2019 | Inconsistent head poses are determined using facial landmark points | | • | Physiological Signal | ✘ | 64 × 64 | SVM | UADFV, DARPA MediFor GAN | AUC = 0.843 | AUC = 0.89 | Performance degradation for Blurry images |
| [174] | 2019 | Unique trait of facial expression and head pose of a speaker | | • | Physiological Signal | ✘ | – | SVM | Videos of world leaders taken from Youtube | AUC = 0.92 | AUC = 1.00 | Performance is compromised when the person is looking off the camera |
| [125] | 2020 | Heartbeat rhythms of an image identity | | • | Biological signal | ✘ | – | CNNs + LSTM | FaceForensics + + & DFDC-preview | Accuracy = 0.975 | Accuracy = 0.997 | Not generalizable to unseen dataset |
| [126] | 2020 | Spatial coherence and temporal consistency of biological signals | | • | Biological Signal | ✘ | 128 × 128 | CNNs + SVM | Ff, Ff ++, Celeb-DF, UADFV & Deep Fakes dataset | Accuracy = 67.61% | Accuracy = 99.39% | Not robust enough for complex and diverse scenarios |
| [175] | 2020 | Discriminative lip feature indicates the talking style of an individual | | • | Physiological Signal | ✘ | – | CNNs | GRID & MOBIO | FRR = 2.7 HTER = 6.0 | FRR = 0.4 HTER = 0.6 | Approach is based on the visual speaker authentication, which cannot be considered a general detection mechanism |

**Table 9** (continued)

| Methods | Year | Key features | Dataset media | | Method types | Local. | Input resolution | Architectural components | Database | Worst performance | Best performance | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Img | Vid. | | | | | | | | |
| [176] | 2021 | Semantic inconsistencies in the emotions of human speakers are predicted using the subject face and voice | | • | Biological Signal | ✘ | – | LSTM, RF, XGBoost, LR, kNN | SEMAINE, DFDC | AUC = 0.921, Accuracy = 80.1% | AUC = 1.00, Accuracy = 99.5% | Method is biased to the Caucasian adults with British accents dialects; hence it may not generalize to other cultures |
| [177] | 2021 | Inconsistencies in the synthetic eyes and gazes | | • | Physiological Signal | ✘ | – | CNNs | Ff ++, Celeb-DF, Deep Fakes & DeeperForensics | Accuracy = 80.0% | Accuracy = 99.27% | LSTM model's performance decreased with the smaller dataset |
| [178] | 2021 | Inconsistencies of the corneal specular highlights between the two eyes | • | | Physiological Signal | ✘ | 1024 × 1024 | – | FFHQ,StyleGAN2 Images | AUC = 0.94 | | Method compared the pixel differences without considering the geometry and scene and assumes that images obey a particular portrait setting |
| [179] | 2021 | Lip-sync dynamics, where ear movement decoupled from the mouth and jaw movement | | • | Physiological Signal | ✘ | 256 × 256 | Logistic Regression Model | Videos are downloaded from the Youtube | AUC = 0.77 | AUC = 0.97 | Method may not work for long hair, large head movement or occluded ears |
| [121] | 2019 | Inter-frame dissimilarity | | • | Temporal inconsistency | ✘ | 300 × 300 | CNNs | FaceForensics + + | Accuracy = 75.41% | Accuracy = 81.61% | Method has reported very few results |

**Table 9** (continued)

| Methods | Year | Key features | Dataset media | | Method types | Local. | Input resolution | Architectural components | Database | Worst performance | Best performance | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Img | Vid. | | | | | | | | |
| [180] | 2019 | Temporal inconsistencies exhibit low-level artifacts of the face region | | • | Temporal inconsistency | ✘ | 224 × 2224 | CNN + GRU | FaceForensics + + | Accuracy = 94.35% AUC = 95.93% | Accuracy = 96.9% AUC = 99.59% | Approach is limited to FaceForensics + + dataset |
| [181] | 2019 | Temporal discrepancies across image streams along with the landmark face alignment features | | • | Temporal inconsistency | ✘ | 224 × 224 | CNNs + GRU | FaceForensics + + | Accuracy = 94.35% AUC = 0.9593 | Accuracy = 96.9% AUC = 0.9964 | Results are reported for the compression version of FaceForensics + + frames only |
| [120] | 2020 | Video speed manipulation feature | | • | Temporal inconsistency | ✘ | – | SVM | Deepfake detection challenge 2019 | Accuracy = 83.1% | Accuracy = 99.1% | Method may not detect the manipulation done within the existing frames of the video |
| [182] | 2020 | Structural inconsistencies of the frames and inter-frame temporal discrepancies | | • | Temporal inconsistency | ✘ | 256 × 256 | CNNs + LSTM | FaceForensics + + & videos from the Youtube | Accuracy = 93.33% | Accuracy = 95.25% | Method doesn't exploit frame-level inconsistencies. Also, performance degradation for compressed videos |

**Table 9** (continued)

| Methods | Year | Key features | Dataset media | | Method types | Local. | Input resolution | Architectural components | Database | Worst performance | Best performance | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Img | Vid. | | | | | | | | |
| [122] | 2021 | Inter-frame temporal structural dissimilarities | | • | Temporal inconsistency | ✘ | 224 × 224 | CNNs | FaceForensics + + | Accuracy = 71.65% | Accuracy = 98.41% | Approach is limited to the FaceForensics + + dataset |
| [183] | 2021 | Temporal features guided by the attention mechanism | | • | Temporal inconsistency | ✘ | 224 × 224 | CNNs with Attention mechanism | FF + + and Celeb-DF | AUC = 0.8278, Accuracy = 87.57% | AUC = 0.998, Accuracy = 99.29% | Scope of improvement for the model's generalization capability and the model is not robust for a different compression level |
| [184] | 2021 | Temporal inconsistencies of the frames using Prototype network | | • | Temporal inconsistency | ✓ | – | CNNs | FaceForensics ++, DFD, DeeperForensics, Celeb-DF | AUC = 68.20%, EER = 37.08%( unseen) | AUC = 92.44%, EER = 16.21% (on unseen dataset) | - |

**Table 9** (continued)

| Methods | Year | Key features | Dataset media | | Method types | Local. | Input resolution | Architectural components | Database | Worst performance | Best performance | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Img | Vid. | | | | | | | | |
| [118] | 2020 | Disparity between audio (speech), video (face) modality, and perceived affective cues | | • | Audio-Visual Inconsis-tency | ✘ | – | CNNs | DF-TIMIT, DFDC | AUC = 84.4% | AUC = 96.3% | Approach is limited to a single person per video. Model misclassifies when a Deepfake video does not contain a mismatch between modalities |
| [117] | 2020 | Mouth incon-sistencies between the shape of the mouth (visemes) with spoken words (phenomes) | | • | Audio-Visual Inconsis-tency | ✘ | 128 × 128 | CNNs | Videos created using A2V, T2V-S, T2V-L methods & videos also taken from Youtube | Accuracy = 93.9% | Accuracy = 97% | Method is limited to the MBP phonemes and the videos of Barack Obama |
| [119] | 2020 | Disharmony between audio and video modalities | | • | Audio-Visual Inconsis-tency | ✓ | 224 × 224 | CNNs | DFDC, DF-TIMIT | AUC = 91.5% | AUC = 97.9% | Does not look into visual inconsistency in the samples, which could also be faked |

between frames. However, the methods have reported very few results. Caldelli et al. [122] also proposed optical flow-based CNNs that exploit the motion dissimilarities in the temporal nature of the video sequences using optical flow fields. Again the approach is limited to the specific dataset. These techniques tend to perform better as they are independent of the specific type of manipulation. However, the overall performance improves when these methods are used in conjunction with spatial methods.

**Physiological/biological signals based methods (PBS)** In deepfake videos, inconsistencies are exhibited either at the physiological level (like inconsistencies in eye blinking patterns, head poses, blending boundaries) or biological level (inconsistencies in a heartbeat), exploited for deepfake video detection. Methods can generate deepfakes with high realism, but they cannot replicate every reasonable behavior, leading to inconsistencies. Such inconsistencies at the physical level may or may not be visible to the eyes; hence, some landmark detector that captures the coordinates of the desired location can be used further for classification. Yang et al. [123] use a landmark detector for 3D head poses to calculate their estimated positions, exploited by the SVM classifier. Their performance degrades to blurry images. Li et al. [124] proposed a method based on eye blinking patterns, which is not well preserved in the synthesized videos. Synthesized videos usually have less frequency of eye blinking, which leads to their detection. The technique can exploit abnormalities in the normal functioning of an organ for deepfake video detection. These signals are preserved neither spatially nor temporally, and different architectures exploit them for detection. Qi et al. [125] use heartbeat rhythms as a clue for detecting deepfake videos. Visual photoplethysmography (PPG) monitors the heartbeat rhythms and captures the abnormalities of the deepfake videos. However, the model does not generalize well to the unseen dataset. Ciftci et al. [126] proposed FakeCatcher methods that use biological signals such as heart rate to exploit the authenticity of a video. They have extracted signals on a pairwise basis, transformed them to a different domain (like frequency, time, etc.), and used this transformation for classification.

Although the methods based on these features perform well on the various datasets, such signals get seriously affected by the video's quality and hence a limited application for detection mechanism based on such signals [12]. [127, 128] some other methods that look for biological methods.

**Spatial and/or temporal domain-based detection** Few detection methods could leverage both the spatial and the temporal domain or either of the two. However, features fetched along both domains capture the broader range of manipulation traces which would eventually help in a better detection mechanism. The following section will discuss such detection methods.

**General DNN-based detection** Instead of focusing on the specific artifacts, some researchers let the network decide which latent features to analyze and learn the mapping accordingly. Deep neural networks drive such methods. Khalid et al. [129] use variational auto-encoder (VAE) to train with the real images, classify them, and treat others as anomalies. The generalization of the unseen dataset has been a more significant issue for such methods. They learn the specific type of manipulation on the data they are trained upon and hence tend to overfit and perform poorly on the other kind of manipulation in the wild. However, few authors have developed a generalizable detector. Xuan et al. [130] also proposed a generalized GAN images forensics detector that preprocesses the images with Gaussian blur and noise operation to enhance the high-frequency pixel noise to allow the CNN model to learn intrinsic discriminative features.

Nowadays, pre-trained models are used heavily to use the learned weights of similar problems to reduce the time complexity of the network. ResNet, XceptionNet, Densenet, AlexNet, and Inception models are recent state-of-the-art models generally used as pre-trained models. Zhou et al. [131] used GoogleNet Inception V3 pre-trained model in one of the branches of the architecture to detect the tampered artifacts evidence and noise inconsistency. Jeon et al. [132] proposed a framework for neural talking head detection using a pre-trained AlexNet model to extract features from a highly unbalanced dataset and then classify them further using a Siamese network-based classifier.

DNN models could also be used for sequence-based problems where data in audio or video frames pass through the model to analyze and learn the intrinsic pattern. Recurrent Neural Networks and their variants LSTM and GRU are generally used for the same purpose. Wu et al. [133] exploited temporal, spatial, and steganalysis features for deepfake video detection. The deep neural network extracts spatial features for the tampering artifacts like irregular shapes, color, etc. Steganalysis features are extracted by putting constraints on the convolution filter for underlying abnormal statistics of the pixels. The temporal inconsistencies are extracted using RNNs. This method beats the current state-of-the-art methods on the FF + + dataset.

DNN-based methods are very good at learning and extracting the intrinsic characteristics of several domains. Such methods tend to overfit the specific datasets they are trained upon, but they lack generalizability to other datasets. Also, the existing methods fail to prove their effectiveness against the adversarial noise attacks [11]. Also, the models do not have the interpretation of why their method proved something fake courtesy of the black-box nature of the model.

Some other methods [134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 130] have extracted discriminative features using DNN either implicitly or explicitly to perform deepfake detection.

Table 10 highlights the various DNNs approaches.

**Frequency domain-based detection** The frequency domain represents the change of pixel distributions along the different axis. Real images have a frequency distribution; such difference could be revealed in the frequency domain when the generative model does the manipulation. Frank et al. [160] explored the frequency spectrum of the GAN-generated images. They found that images exhibit severe artifacts consistent across different resolution images, caused mainly by the upsampling process of the GANs. Furthermore, the model is robust against image perturbation like blurring, cropping, compression and noise addition. Durall et al. [161] observe the behavior of the real and fake images in the classical frequency domain and use such behavior to be detected by the classifier. However, the model has low accuracy on the low-resolution images. Masi et al. [162] used a two-stream network to exploit frequency domain information in one of the streams using the Laplacian of Gaussian (LoG) operator. The LoG acts as a band-pass filter to suppress the image content and amplify the artifacts. Nevertheless, the method struggles to detect real-world data samples. Table 11 gives an overview of frequency domain-based detection methods.

**Summary of the visual deepfake detection methods** There has been a similar trend for various visual detection methods. Most methods use the deep learning-based classifier, while traditional machine classifiers like SVM, KNN, and RBF have also been used. AUC and accuracy have been the most preferred evaluation metric for accessing the model's performance. Also, researchers have preferred the public dataset to evaluate their model rather than designing their dataset. Most of the methods are doing the detection only; there is no localization of the affected area, especially in the case of video media. Three key issues continue to challenge the research community: generalizability of the detection mechanism, robustness against post-processing operations and adversarial noise attacks, and interpretability of the detection mechanism. However, few researchers have solved them to a certain extent; there is a long way before these systems can be implemented for real-world scenarios [11].

### 2.2.2 Deepfake audio detection

Audio deepfake detection is very different from video or image detection. Voice signals in audio are one-dimensional; artifacts would be hard to find in audio but provide sufficient evidence for investigation [199]. Voice is recorded where much noise is present, and then in such circumstances, it is easy for an attacker to fool the detector by adding the real-world noises [199]. Hence, there is a need for a generalized robust detector that detects every kind of manipulation. In addition, with the increase in authentication systems, people prefer to use automated speaker verification as a biometric system where a voice sample is used to authenticate the identity. However, such systems are vulnerable, as attackers easily get all biometric traits required for spoofing such biometric systems. Apart from using voice biometric spoofing, audio deepfake can also be misused in phishing scams, spreading misinformation, fake evidence, blackmail, and bullying [71]. Such vulnerabilities can only be restricted by developing such detector that detects their use and inform about their misuse. The detection mechanism is divided into handcrafted-based and deep learning-based detection methods based on the feature representation.

**Handicraft feature-based methods (HC)** Audio features are extracted from the audio signals by the hand-engineered algorithm; after processing them, they are fed into the classifier to learn the classification. Various audio features represent the relevant aspects of the audio signals extracted using other techniques. MFCCs (Mel-frequency cepstral coefficients), Ceptral coefficients, spectrogram, Constant Q Cepstral Coefficients (CQCCs), Linear Predictive Cepstral Coefficients (LPCCs), Spectral Sub-band Centroid Coefficients, and Complex Cepstral Coefficients (CCCs) are a few of the well-known acoustic features used in the different situation [200]. E.g., MFCC represents the perceptual aspect of the speech spectrum, and it is the most commonly used audio feature [200]. The most widely used conventional classifiers are GMM, GMM-UBM, and SVM. Saranya et al. [201] proposed a model that uses reverberation and channel information of a non-voiced segment of audio signal features (MFCC, CQCC, and MFS) and makes a classification using a GMM classifier. Many Replay attack detection methods investigate the frequency spectrum, as replay recording has noise added (in practical scenarios) reflected in the audio signal's frequency spectrum. Witkowski et al. [202] proposed a model that exploits the high-frequency signal of the replayed recording for spoof detection. AlBadawy et al. [203] did a bi-spectral analysis to find the high order spectral correlation introduced by the speech synthesis mechanism.

**Deep learning-based methods (DL)** Deep learning algorithms' performance is on the rise, and they perform much better than the conventional methods because of the feature abstraction capability, which a deeper model can learn [204]. The GMM classifier only captures first and second-order statistics of the feature vectors, while a deep neural network can also capture linear and nonlinear aspects of the features, giving robustness to various detection methods [204]. Wang

**Table 10** Overview of General DNNs based deepfake detection methods proposed from 2019 onwards

| Methods | Year | Key features | Dataset media | | Local. | Input resolution | Architectural components | Database | Worst performance | Best performance | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Img | Vid. | | | | | | | |
| [185] | 2019 | Latent space distribution features | • | • | ✘ | 256 × 256, 128 × 128 | Auto-encoder-decoder | FaceForensics, CelebA-HQ (real images) & diff. GANs samples | Acc. (synthetic images) = 82.05% Acc. (Inpainting) = 70.62% Acc. (face2/swap) = 72.57% | Acc. (synthetic images) = 100% Acc. (Inpainting) = 99.77% Acc. (face2/swap) = 94.47% | Computationally expensive due to large feature space |
| [186] | 2019 | Cross-layer common discriminative features | • | | ✘ | 64 × 64 | CNNs-DenseNet | CelebA & fake images generated by diff. GANs methods | Precision = 0.909 Recall = 0.865 (for general fake images) | Precision = 0.934 Recall = 0.936 (for general fake images) | Perform poorly for different distribution datasets |
| [187] | 2019 | Multi-level abstraction features of the manipulated region | • | | ✘ | 64 × 64 | CNNs + XGBoost + AdaBoost | SwapMe, FaceSwap | Accuracy = ~ 82.5% AUC = ~ 0.89 | AUC = 0.934 Accuracy = ~ 95.1% | Does not detect the entire synthetic image. It also extracts info from RGB channels only, which may always not be present |
| [132] | 2019 | Latent discriminative features of pre-trained Siamese network | | • | ✘ | – | CNN-AlexNet | VoxCeleb2 | Accuracy = 98.44%, F1 Score = 0.82 | Accuracy = 98.84%, F1 Score = 0.98 | Methods have shown results to the simpler dataset that are not diverse in nature |
| [188] | 2019 | Discriminative features of various spoofs using capsule network | | • | ✘ | 128 × 128 | Capsule- CNNs | REPLAY-ATTACK dataset, FaceForensics | Acc. (Swap) = 94.47% Acc. (reenact.) = 81.00% Acc. (CG) = 97% | Acc. (Swap) = 99.23% Acc. (reenact.) = 99.37% Acc. (CG) = 100% | Not robust enough against adversarial attacks |

**Table 10** (continued)

| Methods | Year | Key features | Dataset media | | Local. | Input resolution | Architectural components | Database | Worst performance | Best performance | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Img | Vid. | | | | | | | |
| [129] | 2020 | Anomaly scores determined using latent feature characteristics | | • | ✘ | 100 × 100 | Variational Auto-encoder | FaceForensics + + | F1score = 0.707, Acc = 70.70% | F1score = 0.982, Acc = 98.2% | Dependency on the RMSE function to compute the reconstruction score of an image, which may not be a better way |
| [189] | 2020 | Visual and temporal dynamics of frames | | • | ✘ | 224 × 224 | CNNs + RNNs(GRU) | DFDC | Accuracy = 91.88% | | Performance needs improvement, and the model dismisses any analysis of audio that could also be faked |
| [190] | 2020 | Frame-level features determined by the attention mechanism | | • | ✘ | 224 × 224 | Attention Map network | Celeb-DF & Youtube videos | Accuracy = 92% AUC = 94% | | Method did not consider the audio modality, which could also be faked |
| [162] | 2020 | Color and frequency domain features | | • | ✘ | 224 × 224 | CNNs + LSTM | Ff ++, Celeb-DF, DFDC preview | Accuracy = 86.34% AUC = 73.41% | Accuracy = 96.43% AUC = 99.12% | Not significant performance for practical web-application samples |
| [133] | 2020 | Spatial, statistical and temporal features | | • | ✘ | – | CNNs + LSTM | FaceForensics + + & GAN-based deepfakes | Accuracy = 83.78% | Accuracy = 98.57% | Method requires training the modules separately; hence, it is computationally expensive |
| [191] | 2020 | Spatial and temporal semantic features of frames | | • | ✘ | 299 × 299 | CNNs + LSTM | UADFV, Df-TIMIT, Ff + +, Celeb-DF, DFDC & Youtube videos | Accuracy = 55.91% | Accuracy = 100% | – |

**Table 10** (continued)

| Methods | Year | Key features | Dataset media | | Local. | Input resolution | Architectural components | Database | Worst performance | Best performance | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Img | Vid. | | | | | | | |
| [192] | 2021 | Multi-scale artifacts of manipulated regions | | • | ✓ | – | Encoder-decoder | FaceForensics ++, Celeb-DF and UADFV | AUC = 78.68% | AUC = 97.60% | Method doesn't show any robustness against different levels of compression |
| [193] | 2021 | Frame-level and temporal inconsistencies of videos | | • | ✘ | 256 × 256, 224 × 224 | CNNs | Celeb-DF, FaceForensics + + | AUC = 0.87, Accuracy = 80.05% | AUC = 0.98, Accuracy = 94.64% | Scope of improvement for the robustness of the model against a different level of compression |
| [183] | 2021 | Spatiotemporal features use an attention mechanism | | • | ✘ | 256 × 256 | CNNs(Inception) With Attention mechanism | FF + + & Celeb-DF | AUC = 0.8129 & Accuracy = 85.29% | AUC = 0.9979 & Accuracy = 99.29% | Method does not explain the interpretability of the detection results as to why the samples are declared as fake |
| [194] | 2021 | Semantic inconsistencies and noise features | | • | ✘ | 128 × 128 | CNNs | FaceForensics + + | Accuracy = 85.71% | Accuracy = 90.36% | Method is not generalizable to the unseen type of manipulation |
| [195] | 2021 | Multi-domain transfer learning-based features | | • | ✘ | 128 × 128 | CNNs | FaceForensics ++, DFDC and Celeb-Df | Accuracy = 73.12%(Low-quality videos) | Accuracy = 86.97%(Low-quality videos) | Method is not generalizable to the talking head types of deepfake |

**Table 10** (continued)

| Methods | Year | Key features | Dataset media | | Local. | Input resolution | Architectural components | Database | Worst performance | Best performance | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Img | Vid. | | | | | | | |
| [196] | 2021 | Multi-headed guided attention mechanism to fetch local discriminative texture features | | • | ✘ | 380 × 380 | CNNs with Attention Mechanism | DFDC, DF, FaceForensics + +, and Celeb-DF | AUC = 87.26%, Accuracy = 86.95% | AUC = 99.29%, Accuracy = 97.60% | Method is sensitive to the high compression rate, which curbs the valuable information of the spatial domain |
| [197] | 2021 | Temporal characterization over the precise geometric landmark inconsistencies | | • | ✘ | – | RNNs(GRU) | UADFV, FaceForensics + +, Celeb-DF & DeeperForensics 1.0 | AUC = 55.4% | AUC = 99.9% | Scope of improvement for the model on generalizability parameter. Lack of interpretation for the temporal features results captured by the model |
| [198] | 2021 | Intra-frame-level spatial artifacts and the temporal inconsistencies between the frames | | • | ✘ | 240 × 240 | LSTM | FaceForensics ++, DFD & DFW | F1 score = 87.58% | F1 score = 97.57% | Method does not consider the various level of compression; they performed results only on high-quality videos |

**Table 11** Overview of frequency-based deepfake detection methods proposed from 2019 onwards

| Methods | Year | Key features | Dataset media | | Local. | Input resolution | Architectural components | Database | Worst performance | Best performance | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Img | Vid. | | | | | | | |
| [161] | 2020 | Exhibition of different behavior at higher frequency domain by fake and real samples | • | • | ✘ | 1024 × 1024, 128 × 128 | SVM, Logistic Reg. & K-Means classifier | CelebA & FaceForensics + +, FF-HQ, 100 K faces project & samples from the thispersondoesnotexist.com | Acc.(SVM) = 58% Acc.(L.R) = 58% Acc.(K-means) = 37% | Acc.(SVM) = 100% Acc.(L.R) = 100% Acc.(K-means) = 94% | Low accuracy on the low-resolution samples |
| [160] | 2020 | GAN artifacts exhibited in the frequency domain | • | | ✘ | 128 × 128 | CNN, kNN & other classifiers | CelebA, LSUN | Accuracy = 69.15% | Accuracy = 99.91% | - |

et al. [199] propose a DeepSonar approach that monitors the layer-wise neurons, providing a deep, subtle understanding of AI-synthesized and genuine voice that enables robust detection. Wijethunga et al. [205] use a combination of CNN and RNN models to extract the dynamic acoustic features of the voice samples for detection. However, as various new types of audio detection methods are emerging, the generalizability of the algorithm is still a challenging problem. Chen et al. [206] proposed generalized DNN detection methods; the model uses marginal cosine loss function (LMCL) and frequency masking augmentation to learn robust discriminative features. As discussed earlier, ASV systems are vulnerable to spoofing attacks. There are four kinds of spoofing attacks: speech synthesis (SS), voice conversion (VC), impersonation, and replay attacks. Shim et al. [207] proposed a replay spoofing detection mechanism using the multi-tasking learning approach of DNNs. It is based on the hypothesis that spoofed signal contains replay noise and genuine signal, and internal noise. Yang et al. [208] propose a synthetic speech (SS) detection method by exploiting the high-frequency band of the long-term transform-based features.

**Summary of the audio deepfake detection methods** Various methods have been proposed for fake audio detection in the past few years. ASV challenge has been devoted to the researcher to develop new approaches to detect counterfeit voices. This challenge also proposed the dataset, which let anyone use and beat up the metrics of the state-of-the-art methods. Table 12 lists the various techniques that have been proposed over the year. The table describes the methods along various features used, type of method, architectural components used, dataset used for evaluating the model, the best performance of the model, and the limitation of the model. There are various insights the listing of methods provides; first, most of the techniques are DNN-based, which shows how such a network learns the latent, intrinsic discriminative features that eventually lead to better results. Second, the ASV challenge provides the dataset and the evaluation metric, EER (and t-DCF also, at times), the typical benchmarks to beat the state-of-the-art methods. Few limitations that visual deepfake detection face is also applied to audio deepfake. Procedures are not generalizable to the unseen type of manipulation, not robust against the various kind of attacks, performance degradation on the low-quality and noisy samples, and lack of interpretability of the detection results. Also, the model is computationally expensive and would need a lot of data for training before it is put for testing. The method needs to tackle these limitations for their deployment in real-world scenarios.

## 3 Deepfake datasets

With the emergence of a highly sophisticated deepfake algorithm, it becomes apparent to have a dataset that is a good representative of the distribution and enables the methods to leverage their generalizability power. Furthermore, as more and more datasets are evolving, it becomes easy for the detection methods to realize their full potential of detection and arrive at an accurate picture of the performance of their methods. In this section, we will discuss various datasets that have been proposed over the years.

### 3.1 Image and video deepfake datasets

This category is most sought after as its two visual classes, face swap and face reenactment, carry a huge potential for misuse. Many detection methods have been proposed to standardize the detection benchmark; each brings different manipulations with quantity and heterogeneity. Li et al. [225] have divided the dataset into the most common generations based on release time and synthesis algorithms. UADFV, FaceForensics ++, and Deepfake-TIMIT (DF-TIMIT) datasets belong to the first generations. Google-DFD datasets (later merged to FaceForensics) & Celeb-DF come under the second generation. Finally, Dolhansky et al. [226] proposed a third-generation under which the most recent dataset, DeeperForensics1.0, and DFDC datasets come. Table 13 gives an overview of the various deepfake image and video datasets. Figure 9 presents visual samples for popular visual deepfake datasets.

### 3.2 Audio deepfake datasets

Fake audio datasets have also unfolded in the past few years with the advent of audio spoof detection methods on biometric systems. In addition, the ASVspoof challenge is conducted, which results in a dataset every third year. Apart from that, various fake synthetic speech TTS datasets have surfaced with increased speech synthesis methods. Table 14 describes such datasets.

## 4 Architectural components and tools for deepfake

Various deepfake generation and detection frameworks use different components or variations of the below components for their architecture (Fig. 10). Table 15 gives a brief overview of such detection methods.

Today, many open-source tools allow you to create deepfake easily, even without knowing the underlying technology. Table 16 gives a brief description of the popular open-source tools available.

**Table 12** Overview of different audio deepfake detection methods from 2019 onwards

| Ref | Year | Model and features description | HC | DNN | Architecture components | Dataset | Worst performance | Best performance | Limitation |
|---|---|---|---|---|---|---|---|---|---|
| *Hand-crafted features based methods* | | | | | | | | | |
| [203] | 2019 | Exploits the high order correlation using bi-spectral analysis to distinguish the synthetic speech from the original human voice | • | | LR Classifier | Podcasts Recordings | AUC = 0.88 | AUC = 0.99 | Performance drops with the addition of noise in the audio samples |
| [209] | 2019 | Employed an approach that uses high-order spectral analysis, Gaussianity Statistics & linearity test statistics to capture the artifacts of the speech synthesis generative models | • | | – | Baidu cloned Audio dataset | Accuracy = 75% | Accuracy = 100% | Performance remains unclear against the various level of compression and for noisy samples |
| [210] | 2019 | Uses Cochlear model to extract the discriminative features of AM and FM speech signal components | • | | Cochlea model + GMMs cls | ASV Spoof 2017- 2 replay corpus | EER = 11.30% | EER = 7.32% | Model is computationally complex |
| [211] | 2021 | Supervised learning-based model where features are constituted by modeling the speech as an auto-regressive process | • | | RF, SVM, RBF SVM | ASVspoof2019 | Accuracy = 0.735 | Accuracy = 0.741 | Performance degradation due to compression |

**Table 12** (continued)

| Ref | Year | Model and features description | HC | DNN | Architecture components | Dataset | Worst performance | Best performance | Limitation |
|---|---|---|---|---|---|---|---|---|---|
| *DNN features based methods* | | | | | | | | | |
| [212] | 2019 | Uses an attentive Filtering deep neural Network that captures the discriminative feature in the time and frequency domain | | • | CNN-ResNet | ASVspoof2017 | EER = 8.98% | EER = 8.54% | Method needs a performance improvement |
| [213] | 2019 | Uses the DenseNet-LSTM framework that uses hybrid features of the speech segment to make a detection | | • | DenseNet-LSTM | ASVspoof2017 | EER = 9.56% | EER = 8.84% | Dependency on the background sound of the different record and playback devices; hence model does not perform well in the presence of good-quality hardware |
| [214] | 2019 | Proposed an LC-RNN framework that extracts the speech signals features as utterances level embedding used by the back-end classifier | | • | CNN-GRU + Classifier (SVM, LDA, PLDA) | ASVspoof2015ASVspoof2017 ASVspoof2019 | t-DCF(LA) = 0.1873 EER(LA) = 7.12% t-DCF(PA) = 0.0946 EER(PA) = 3.49% | t-DCF(LA) = 0.1523 EER(LA) = 6.28%, t-DCF(PA) = 0.061, EER(PA) = 2.23% | Method's performance remains unclear in noisy conditions and against various compression levels |
| [215] | 2019 | Uses signal-to-noise masks and gated neural network extracts the robust features at utterance level embedding's used by the backend recognizer to spoof audio detection | | • | DNNs + GRU | ASVspoof15, ASVspoof17, ASVspoof19 | EER = 3.96%((unseen condition) | EER = 2.44%((unseen condition) | Method lacks the interpretation of its processed and detection results |
| [199] | 2020 | Monitor the layer-wise neural behavior of DNNs to capture the subtle differences between real and AI-synthesized voices | | • | Deep Neural Networks | FoR, Sprocket-VC & MC-TTS | Acc. = 0.981, AUC = 0.982, F1 Score = 0.982 AP = 0.976, EER = 0.021 | | Not robust enough against the adversarial attacks and performance degradation for real-world noise samples |

**Table 12** (continued)

| Ref | Year | Model and features description | HC | DNN | Architecture components | Dataset | Worst performance | Best performance | Limitation |
|---|---|---|---|---|---|---|---|---|---|
| [205] | 2020 | Uses DNN to extract discriminative dynamic acoustic features of voice for classification | | • | CNNs + RNNs | FoR Dataset & AMI corpus | Accuracy = 88% | Accuracy = 89% | Limited to the speaker's utterances one word at a time. Performance degradation for noisy samples |
| [206] | 2020 | Uses a Marginal Cosine Loss Function & frequency mask augmentation to learn the robust features through DNNs | | • | CNNs | ASVspoof 2019 | EER = 4.04% t-DCF = 0.109 | EER = 1.26% t-DCF = 0.052 | Method not robust enough for noisy samples |
| [216] | 2020 | Uses an attention-based framework that extracts LBF features from the selected segments of the high-frequency domain | | • | DenseNet BiLSTM Net | BTAS2016, ASVspoof2017 | EER = 6.43% | EER = 0.53% | Method requires background sound of different records and playback devices and hence not diverse |
| [208] | 2020 | Exploited the high-frequency band info of the long-term transform-based features for synthetic speech detection | | • | DNNs | ASVspoof2015 ASVspoof2019 | EER = 0.345% | EER = 0.090% | Method does not perform well for real-world scenarios |
| [217] | 2020 | Employed a model that exploits the characteristics distribution of the genuine speech and uses it as a feature for classification | | • | CNNs + Transformer | ASVspoof2019 | EER = 4.07%, t-DCF = 0.102 | | Approach does not detect the replay attack detection |

**Table 12** (continued)

| Ref | Year | Model and features description | HC | DNN | Architecture components | Dataset | Worst performance | Best performance | Limitation |
|---|---|---|---|---|---|---|---|---|---|
| [218] | 2020 | Uses an approach where dense connectivity strengthens the propagation of features and ensures the maximum flow of information through dense layers | | • | CNNs | ASVspoof19 | t-DCF = 0.1853, EER = 8.99%(LA) | t-DCF = 0.0469, EER = 1.98%(LA) | Method does not solve the issue of generalization and robustness against different levels of processing |
| [219] | 2020 | Uses a parallel network fetches the RDF features in one branch using CQT, BLSTM, DCT network, while the other network extracts CQCCE acoustic features, fused for classification | | • | DNNs + LSTM | ASVspoof2017-V2 | EER = 15.08% | EER = 8.89% | Method's robustness against the different kinds of processing remains unclear, and also, it is not generalizable to the unseen samples |
| [220] | 2020 | Uses a capsule network modifying the dynamic routing algorithm to focus on the local artifacts to yield better results on the unseen fake audio attacks | | • | CNNs | ASVspoof2019 | t-DCF = 0.0982 EER = 3.19%(logical access) | | Method lacks the interpretability of detection results as to what features the capsule network uses for detection |
| [221] | 2021 | Uses a method that extracts the temporal waveform features. It uses the since convolution and squeeze excitation module, which implements the module's recalibration, and captures the interdependencies between the audio channels, enhancing the good bands for detection | | • | CNNs | ASVSpoof19 | EER = 16.39% | EER = 7.23% | Scope of improvement for the generalizability of the algorithm |

**Table 12** (continued)

| Ref | Year | Model and features description | HC | DNN | Architecture components | Dataset | Worst performance | Best performance | Limitation |
|---|---|---|---|---|---|---|---|---|---|
| [222] | 2021 | Explored the method that extracts the intrinsic deep learning features using CNNs from a raw wave, then CQCC features extracted from the same raw wave | | • | LSTM + CNNs | BTAS2016 & ASVspoof2017 | EER = 7.73% | EER = 0.79% | Performance is limited to the dataset on which it is trained |
| [223] | 2021 | Uses a capsule network that uses different input features to learn various artifacts of spoofed voices that assume to bear the similarity with the spatial information | | • | CNNs | ASVspoof2019 | t-DCF = 0.1561 EER = 6.32% | t-DCF = 0.1198 EER = 4.93% | Not generalizable to the unknown spoofing attacks |
| [224] | 2021 | Modified the ResNet block architecture to capture multi-scale features. It splits the features map within a block into different channel groups and designs a residual-like connection across such groups to enable capturing multi-scale features | | • | CNNs | ASVspoof2019 | t-DCF = 0.0743 EER = 2.502%(LA) | t-DCF = 0.0452 EER = 1.892%(LA) | Scope of improvement for better generalizability for ASV anti-spoofing tasks |

**Table 13** Image and Video deepfake datasets

| Dataset | Year | Media | Manipulation technique | No. of original samples | No. of manipulated samples | Resolution | Format |
|---|---|---|---|---|---|---|---|
| Swap Me & FaceSwap dataset [131] | 2017 | Image | FaceSwap | 2300 | 1005 | | JPEG |
| UADFV [123] | 2018 | Image | FakeApp [14] | 49 | 49 | 294 × 500 | – |
| Deepfake-TIMIT [227] | 2018 | Video | FaceSwap-GAN [13] | – | 620 | 64 × 64(LQ), 128 × 128(HQ) | JPG |
| FaceForensics [228] | 2018 | Video | Face2face [26] approach | 500,000 frames edited from 1004 videos | | Videos at least 480p and at most 1080p | Frame rate @ 30fps, either raw with 'DIB' codec or compressed with H.264 |
| CelebA-HQ [67] | 2018 | Image | ProGAN | | 3000 | 1024 × 1024 | JPEG |
| Fake Face in the wild(FFW) [229] | 2018 | Image | CGI, FakeApp, Face Swap | 53,000 images from the 150 videos | | At least 480p | |
| FaceForensics + + [16] | 2019 | Image | Face2Face, FaceSwap, DeepFakes, Neural Textures | More than 1.8 million images from 4000 fake videos | | 480p, 720p, 1080p | - H.264, CRF = 0,23,40 |
| Deepfake Detection Chal-lenge(DFDC) Dataset [226] | 2019 | Video | FaceSwap | 23,654 | 104,500 | 1080 × 1920 | H.264 |
| DFDC Preview [230] | 2019 | Video | Faceswap | 1131 | 4113 | 1080 × 1920 | H.264 |
| Google DFD [231] | 2019 | Video | Faceswap | 363 | 3068 | 1080 × 1920 | - |
| Celeb-DF [225] | 2020 | Video | DeepFake algorithm | 590 | 5639 | 256 × 256 | MPEG4.0 |
| DeeperForensics 1.0 [232] | 2020 | Video | FaceSwap | 50,000 | 10,000 | 1920 × 1080 | - |
| WildDeepfake [233] | 2020 | Face sequences | Faceswap | 3805 | 3509 | - | - |
| Vox-Deepfake [234] | 2020 | Video | Faceswap | 1,045,786 | 1,125,429 | 224 × 224 | - |
| Deepfake MNIST + [235] | 2021 | Animated Video | Siarohin's framework | 10,000 | 10,000 | 256 × 256 | Raw & H.264 in c23 & c4 |

## 5 Current limitations and open challenges

Since the advent of this technology, there has been significant progress; but still, there are open challenges that need to be addressed to make the best use of the technology for beneficial purposes.

### 5.1 Generation methods

The quality of deepfake samples has evolved with the sophistication of the generative methods; however, several concerns of these methods would remain the challenge to be foreseen. Various limitations and challenges which may need to look to move forward are as follows:

- **Generalizability:** Deepfake generative models are data-driven; to generate the specific identity samples, we need similar identity training data to train the network [4]. It constrains the model to develop the particular sample type, limiting their generalizability. Moreover, it is easier to get enough training data for driving samples than specific examples [4]. In addition, to model each separate identity,

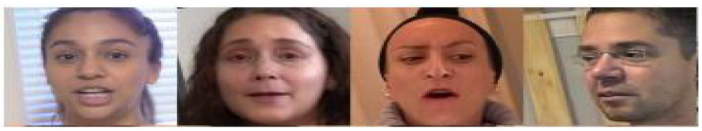
| Dataset | Data Samples |
|---|---|
| Deepfake-TIMIT [227] | |
| Faceforensics [228] | |
| Celeb-DF [225] | |
| Faceforensics++ [16] | |
| Deepfake Detection Challenge(DFDC) Dataset [226] |  |
| DeeperForensics 1.0 [229] | |

**Fig. 9** Data samples of popular visual deepfake dataset

we need to train the model again, which would be computationally expensive, making it difficult for the system to deploy anywhere. Therefore, extensive efforts have been put into finding ways to model the methods that need less training data and work on multiple unseen identities.

- **Driver's content dependency:** For face reenactment, content is derived from the driving samples, usually with a frontal pose, constraining the flexibility of pose variations. Also, with the face-swap techniques, the frontal face is derived from the lookalike identity, and finding a good match or similar personality is not possible every time [4]. This limits the flexibility of the generation methods that end with a static performance. Therefore, designing the target identity with expected expression and personality would remain a challenge to be solved.
- **Paired training:** The trained, supervised model generates the desired output based on similar input data. It requires data pairing, which becomes laborious when it is done for every specific identity. The process becomes more complicated when the model is designed for multiple identities for multiple tasks. However, many existing methods resolve this issue by using encoding in the ED network or CycleGAN.
- **Spatial inconsistency and temporal coherence:** Visual-spatial inconsistencies appear in various forms like inconsistent facial features due to interpolation (facial reenactment or lip-sync methods) or blending boundaries (face-swap), which is challenging for few techniques. It also happens due to irregular illumination or abrupt change in lighting conditions of the environment. Another concern is temporal coherence in the form of flickering or jittery among frames of the videos. It usually happens due to considering the individual frames without considering the relationship among them [4]. To overcome these issues, some methods provide this context either to the GANs, employing temporal coherence loss, or using RNNs to consider the temporal relation between the frames or combining both of them [4].
- **Identity leakage:** During the face reenactment or body-puppetry task, the partial driver's identity is transferred to the target identity resulting in inconsistencies in the facial features. It happens when the training is done on a single identity or the network is trained on multiple identities; the data pairing is done with the same identity [4].
- Heavy reliance on the massive dataset for audio deepfake generation: To produce the high-quality synthetic speech

**Table 14** Audio deepfake datasets

| Dataset | Year | Techniques/tools used for generation | No. of Manipulated Samples | No. of Original Samples | Language | Format |
|---|---|---|---|---|---|---|
| CMU Arctic Speech Database [236] | 2004 | Recorded | 1200 phonetically utterances | | English | Wave |
| LibriSpeech [237] | 2015 | LibriVox's API & Sequitur G2P toolkit | 1000 h of Audio book recording | | English speech | FLAC(Free Lossless Audio Codec) |
| ASVSpoof2015 [238] | 2016 | Audio spoofed by Voice conversion(VC) & Speech Synthesis(SS) algorithms | 16,651(Genuine) | 246,500(Spoofed) | – | RIFF/Wave |
| LJ Speech Dataset [239] | 2017 | Recorded | 13,100 short clips of a single speaker | | – | Wave |
| ASVSpoof2017 [240] | 2018 | Recorded | 3565(Bona fide) | 14,465(Replay) | English | – |
| VoxCeleb2 [241] | 2018 | Two-stream CNNs | 1,128,246 number of utterances from 6112 speakers | | – | – |
| CSTR VCTK Corpus [242] | 2019 | Recorded | 110 English Accent | | English | – |
| Fake-or-Real(FoR) Dataset [243] | 2019 | Deep voice 3, WaveNet, Amazon AWS Polly, Baidu TTS, Microsoft Azure TTS, Google Traditional & Cloud TTS, | 111,000 real utterances | 87,000 synthetic utterances | English | MP3/WAV |
| VoxCeleb1 [244] | 2019 | Two-stream CNNs | 153,516 number of utterances from 1251 speakers | | – | – |
| M-AILabs Dataset [245] | 2019 | Recorded | Around 1000 h recording | | German, English, Spanish, Italian, Ukrainian, Russian, French, Polish | Wave |
| ASVSpoof2019 [246] | 2020 | Tacotron2 & WaveNet | 122,157 Samples | | – | Mp3 |

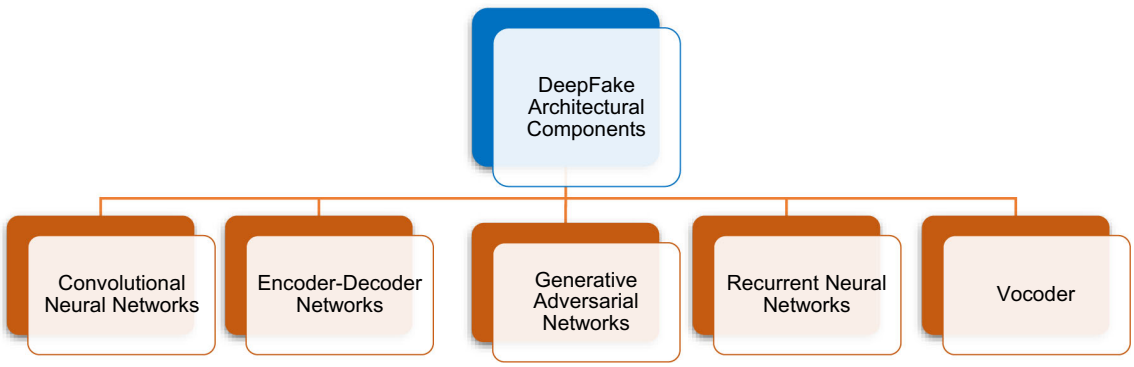



**Fig. 10** Deepfakes Architectural components

**Table 15** Different Architecture components of deepfake methods

| Components | Descriptions |
|---|---|
| Convolutional Neural Networks(CNNs) | CNNs extract the visual features from the images or video frames in an abstract feature map using convolutional kernels. Pooling layers are used for dimensionality reduction, while upsampling layers increase it. Non-activation functions like ReLu introduce the nonlinearity that allows learning of nonlinear discriminative features. It is used frequently to analyze the distribution of pixels and extract visual data features using a feature map, which can be used further to manipulate the distribution or find abnormalities for detection |
| Encoder-Decoder(ED) | Encoder-decoder network is a particular class of neural network where the encoder network learns the mapping of compressed latent features to its distribution. In contrast, a decoder network learns the mapping of compact features to its original input. If the encoder-decoder network is symmetrical and the network regenerates back the original image, then the network is called auto-encoder [4]. Another variant of ED is Variational Auto-encoder (VAE), where the network learns the posterior distribution of input. However, it is better at disentangling the various latent features, which ease the process of modification and interpolation [4]. ED provides the latent space vectors, which different techniques exploit and use to change the attribute in a specific direction. It has also been exploited to identify the distribution pattern of real and fake samples |
| Generative Adversarial Networks(GANs) | Google researcher Ian J. Goodfellow proposed GANs in 2014 [247]. GANs are composed of two neural networks: Generator () and Discriminator (Đ). Generator generates the fake samples from the random noise distribution while the Discriminator Đ identifies the artificially generated samples. Both are trained in an adversarial manner with the aim that generator produces more realistic samples while discriminator Đ gets better at identifying them. Visual deepfake has become so realistic because of the arms race between generator and discriminator that challenge each other to improve the distribution. It goes on and on until we have such real samples that even discriminators have difficulty distinguishing them |
| Recurrent Neural Networks(RNNs) | RNNs are also a type of neural network that can learn the long-term dependencies of the data; hence, they are good at handling variable length and sequential data. Therefore, it is used to process temporal data like videos and audio. Nowadays, a more advanced form of RNNs is used: Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU). They are very good at learning the dependencies and retaining the order of features to be used for detection or manipulation problems |
| Vocoder | A voice encoder captures the characteristics of an audio signal that can convert one form of acoustic feature to various other formats. It is used as a component in the varied architecture of speech synthesis, which converts acoustic features to the waveform. Different deepfake techniques are used to extract the acoustic features and give them for the final conversion of acoustic features to the waveform |

or audio waveforms, the current model employs deep neural networks that require a lot of training data to learn the speaker's characteristics. Acquiring such extensive data for each speaker is a difficult task, and in addition, it requires retraining the model again for a new speaker, which is a highly computational task. Moreover, few TTS systems require manual annotation for the dataset, which is an uphill task. A few methods use the pre-existing model or employ few-shot learning methods to mitigate this issue, which is a good step in this direction, but there is still a lot to be done.

- **Lack of realism for synthetic voice:** The voice quality is getting better courtesy of the advanced model, but still, there is enough scope for improvement. Artificial voice lacks pauses, accent, tone variation, emotions, and the natural pace of the intended speaker. Mainly, voice-cloning methods cannot produce a clean synthetic voice and contain the background noise, which can be taken care of by using a clean voice dataset.
- **The realism of real-time deepfakes:** Various methods [26, 256] have generated real-time deepfakes, but they are far from realistic. The realism of the content would be beneficial for imparting education in no time. However, the evolution of these techniques can invite various phishing attacks, which is again a challenge that needs focus.

## 5.2 Detection methods

Since the inception of deep learning frameworks, the performance of the classification algorithms has gone to the next level; deepfake detection methods are too leveraging their capability for advancing the purpose. Apart from the deep learning-based detection methods, hand-engineered feature-based methods also perform well. Despite all these advances, some serious concern needs to be taken care of to stop this

**Table 16** Popular open-source tools available for deepfake generation

| Tools | Brief Description | Platform | Manipulation Types | Media | Scale of dataset used |
|---|---|---|---|---|---|
| FakeApp [14] | Reddit users developed FakeApp software, and its algorithm has progressed ever since | Desktop Application | FaceSwap | Videos | A single sample of source and target video is required |
| DeepFaceLab [248] | Pioneer software is famous for manipulating lips and replacing faces and heads. The authors claim to have an accuracy of up to 95% | Open-source implementation | FaceSwap | Videos | A single sample of source and target video is required |
| Deepfakes web [249] | Online web software has a powerful GPU on the cloud that creates video/image in around 4 h. High-quality deepfakes have a visible watermark to state that it is not real | Web-based Application | FaceSwap | Videos/Images | A single sample of source and target video is required |
| FaceApp [250] | The smartphone app creates deepfake images using many AI filters and effects, and it garnered attention due to trans-change features | Mobile application | Facial Attribute manipulation | Images | A single sample is required |
| Faceswap [13] | The open-source tool with lots of functionality allows (even) saving your model's training. It runs on all platforms and requires a high-capability processor GPU to run it | Open-Source Implementation | FaceSwap | Videos | A single sample of source and target video is required |
| Zao [251] | Smartphones app is mainly used for swapping faces with a celebrity in seconds. Therefore, it works best on Chinese faces | Mobile App | FaceSwap | Videos/Images | A single sample is required |
| MachineTube [252] | Web-based software allows creating deepfake video/image free. However, it is a bit slow; requires high-end computing GPUs | Web-based App | FaceSwap | Videos/Images | A single sample is required |
| Doublicat [253] | The smartphone app allows the creation of images, memes, and GIFs quickly. In addition, it makes use of RefaceAI to smoothen the outcomes | Mobile App | FaceSwap | Videos/Images | A single sample is required |
| Resemble [254] | A text-to-speech (TTS) deepfake software allows cloning, voice modulation, and intonation with an emotion-embedded generated voice | Web-based App | Text-to-speech Synthesis | Audio | Around 5 min clip of audio is required |
| Wav2Lip [255] | A lip-sync software allows you to sync the mouth region according to the arbitrary audio | Open-source Implementation | Lip-Sync | Videos | A single sample of video and audio files is required |

technology's misuse. Following are some concerns that may need to be addressed:

- **Generalization:** Existing deepfake detection methods perform well on the seen dataset, but their performance degrades on the unseen dataset. This concern affects the generality of the algorithms, which makes them unfit for their deployment in real-world scenarios [11]. Moreover, its absence gives the upper hand to the anti-social elements to misuse the technology at their whims and fancies. That's why generalization is one of the most crucial indicators of the performance of the methods [4].
- **Lack of interpretability of methods:** Most detection methods use a neural network with an inherent problem of lack of explainability for their results due to the black-box nature [12]. For real-world scenarios, human-understandable justification is required for any forensic methods. E.g., suppose any deepfake detection tool is deployed in the courtroom to detect the evidence. In that case, reason or explanation may be required for various aspects of it being a deepfake. Therefore, detection methods need to focus on the explainability of the detection results, which undoubtedly remain a concern to look at in the future [11].
- **Scarcity of a large quality dataset:** Dataset plays a pivotal role in the detection methods, as it allows learning the different sets of features required to identify various artifacts [10]. Other proposed datasets suffer from problems due to which detection methods suffer. Some of the noticeable issues:
- Small in size and not representative of various kinds of manipulation.
- Inconsistent and blurriness of facial features.
- Flickering or jittery of video frames.
- Uneven illumination in images.
- Lack of noisy audio dataset (to represent real work scenarios).
- Lack of occluding objects in images.
- Low-quality images/video frames.
- **Performance degradation in real-world scenarios:** Existing methods tend to perform well in the controlled environment, where we have a dataset that hardly represents real-world scenarios. For real-world data, which contains various noises and manipulation, their performance degrades as the detection methods are designed to identify specific types of artifacts. In addition, real-world data are multi-facet data where subjects are manipulated in various ways, and samples detected as not manipulated do not mean that they have not been changed in any form. Therefore, the binary classification may not be an ideal way to go about it; a multi-class category could be more suitable that tells the varied degree of manipulation [10].
- **Lack of manipulation localization methods:** Various methods can detect the manipulation but cannot localize it. Localization methods, in a way, guarantee that the detection methods have learned the correct features required for detection. In addition, localization may explain the technique used to produce the deepfake, which may be helpful for forensics investigation.

## 6 Conclusion and future work

There is an arms race between attacker and defender; this race challenges each other to get the best out of them and develop more sophisticated and optimized approaches. However, not all generated deepfakes are malicious. As we saw earlier, deepfakes can be used for creative purposes like education, innovation, entertainment, etc.; it raises the alarm when used for malicious which can afflict pain, exploit and harm individuals, and become a threat to society and nation. This paper comprehensively discussed the application of deepfakes, various types of deepfakes, their generation and detection methods, datasets for multiple media, limitations and open challenges of deepfakes generation and detection methods.

Future holds the promise for this technology to be leveraged fully. Fabricated content would become more and more realistic, which may easily convince naked eyes. Detection tools would be deployed on various platforms to prove the authenticity of the media. Based on papers within this survey, the research community has remarked on the future challenges and work as follows:

- Existing methods primarily focus on the frontal face, face swapping or face reenactment. Future deepfake methods could potentially manipulate the upper or entire body.
- Real-time deepfake generation is expected to be seen more frequently in the future. This deepfake generation technology matures that would require fewer data and generates more believable deepfakes in less time.
- We may expect that deepfake will prevail in other domains in the future. Deep learning approaches generate fingerprints of an individual more convincingly, and deepfakes would be used to create news articles and tweets more frequently. Deepfake can be applied to financial accounts more often to evade detection [4].
- The current trend of deepfake has been focused mainly on generating visual content; in the future, we could see more and more high-quality fabricated audios that resemble human voices in terms of accent, tone variation, emotions, etc.
- We could expect to see human robots speaking in human-realistic voices in the future. These models could be

identity agnostic and trained using a few samples of human voice data. More and more TTS systems would become part of daily chores, and voice impersonation would be easy to obtain.

As the deepfake generation technique continues to evolve in the future, it tends to challenge the detection techniques with the same level of sophistication, which would lead to, some highly optimized, more robust, and generalized approaches. The following may be the future work that would majorly improve the limitations of the existing detection methods:

- Multi-tasking learning strategies may be implemented in future models [12]. Multi-tasking allows doing more than one task at a time, like detection and localization. Each task complements the other, as localization allows the detection procedure to learn the correct artifacts, eventually leading to even a better localization.
- The current problem for the detection methods is to generalize well on the unseen data due to the different distribution of the dataset. A Triplet training strategy could reduce the distance between the identical distributions and widen the gaps between the samples of the other distributions, which subsequently helps in the classification tasks [12].
- Existing detection methods focus on finding the drawbacks of deepfake generation methods and use them as classification artifacts. Such information may not be available in future deepfakes, especially in an adversarial environment where attackers intentionally try not to reveal the artifacts. Futuristic models need to boost approaches' robustness, generalizability, and scalability.
- As this technology grows, social media platforms will be exploited by this technology. There will be a need to integrate deepfake detection methods on such platforms to deal with the widespread malicious effect of this technology. The legal requirement could be framed to enforce this requirement effectively [257].
- The future trend would be employing a unified detection framework that may work on finding multiple forgeries for video media. The framework could use parallel branches to detect the visual content, while the other unit could detect the fake audio. Such a framework could effectively deal with the synthetic content in any form [10].
- Existing datasets are not robust enough to consider various deepfake generation techniques. We have seen fewer data samples based on lip-sync approaches or entire image fabrication. The success of detection approaches is directly proportional to the robustness and variedness of the datasets [10]. Future work could include the proposal of a more varied and robust dataset.
- Deepfake detection methods could prove the evidence in police investigations and court cases. In the future, digital media forensics will be collaborating with the investigation to examine and analyze the prosecutor's proofs [257].

## Declarations

**Conflict of interest** There is no conflict of interest

## References


1. Güera D, Delp EJ (2018) Deepfake video detection using recurrent neural networks. In: 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS), Auckland
2. Strickland E (2019) Facebook AI launches its deepfake detection challenge. In: IEEE, December 2019. https://spectrum.ieee.org/facebook-ai-launches-its-deepfake-detection-challenge
3. Chesney R, Citron DK (2018) Deep fakes: a looming challenge for privacy, democracy, and national security, 68
4. Mirsky Y, Lee W (2021) The creation and detection of deepfakes: a survey. ACM Comput Surv 54(1):1–41
5. Jaiman A (2020) Positive uses of deepfakes, towards data science, 15 Aug 2020. https://towardsdatascience.com/positive-use-cases-of-deepfakes-49f510056387. Accessed 11 April 2021
6. Damiani J (2019) A voice deepfake was used to scam a CEO Out Of $243,000, Forbes, 3 September 2019. https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/?sh=70583a482241. Accessed 10 July 2021
7. Jaiman A (2020) Deepfakes harms and threat modeling, 19 Aug 2020. https://towardsdatascience.com/deepfakes-harms-and-threat-modeling-c09cbe0b7883. Accessed 14 April 2021
8. . Rizzotto L (2019) Deepfake ads, 4 Dec 2019. https://medium.com/futurepi/why-deepfakes-will-change-advertising-forever-2949ec3f87ee. Accessed 18 April 2021
9. Tolosana R, Vera-Rodriguez R, Fierrez J, Morales A, Ortega-Garcia J (2020) Deepfakes and beyond: a survey of face manipulation and fake detection. Inf Fusion 64:131–148
10. Masood M, Nawaz M, Malik KM, Javed A, Irtaza A (2021) Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward, http://arxiv.org/abs/2103.00484v1
11. Juefei-Xu F, Wang R, Huang Y, Guo Q, Ma L, Liu Y (2021) Countering malicious deepfakes: survey, battleground, and horizon. In: http://arxiv.org/abs/2103.00218v1
12. Yu P, Xia Z, Fei J, Lu Y (2021) A survey on deepfake video detection. IET Biometrics 10(6):607–624
13. Faceswap, https://faceswap.dev/. Accessed 6 April 2021
14. FakeApp, https://www.malavida.com/en/soft/fakeapp/. Accessed 6 April 2021
15. deepfakes/Faceswap, github, 2016. https://github.com/deepfakes/faceswap
16. Rössler A, Cozzolino D, Verdoliva L, Riess C, Thies J, Niessner M (2019) FaceForensics++: learning to detect manipulated facial images. In: IEEE/CVF International Conference on Computer Vision (ICCV), Seoul
17. Dale K, Sunkavalli K, Johnson MK, Vlasic D, Matusik W, Pfister H (2011) Video face replacement. ACM Trans Gr 30(6):1–10
18. Li L, Bao J, Yang H, Chen D, Wen F (2020) Advancing high fidelity identity swapping for forgery detection. In:

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle
19. Nirkin Y, Keller Y, Hassner T (2019) FSGAN: subject agnostic face swapping and reenactment. In: IEEE/CVF International Conference on Computer Vision (ICCV), Seoul
20. Chen R, Chen X, Ni B, Ge Y (2020) SimSwap: an efficient framework for high fidelity face swapping. In: Proceedings of the 28th ACM International Conference on Multimedia, Seattle
21. Zhu Y, Li Q, Wang J, Xu C, Sun Z (2021) One shot face swapping on megapixels. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville
22. Zhang L, Yang H, Qiu T, Li L (2021) AP-GAN: improving attribute preservation in video face swapping. IEEE Trans Circuits Syst Video Technol (Early Access) 32(4):2226–2237
23. Peng B, Fan H, Wang W, Dong J, Lyu S (2021) A unified framework for high fidelity face swap and expression reenactment. IEEE Trans Circuits Syst Video Technol (Early Access) 32(6):3673–3684
24. Cao M, Huang H, Wang H, Wang X, Shen L, Wang S, Bao L, Li Z, Luo J (2021) UniFaceGAN: a unified framework for temporally consistent facial video editing. IEEE Trans Image Process 30:6107–6116
25. Chan C, Ginosar S, Zhou T, Efros A (2019) Everybody dance now. In: IEEE/CVF International Conference on Computer Vision (ICCV), Seoul
26. Thies J, Zollhöfer M, Stamminger M, Theobalt C, Nießner M (2016) Face2Face: real-time face capture and reenactment of RGB videos. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas
27. Thies J, Zollhöfer M, Nießner M (2019) Deferred neural rendering: image synthesis using neural textures. ACM Trans Gr 38(4):66
28. Liu L, Xu W, Zollhöfer M, Kim H, Bernard F, Habermann M, Wang W, Theobalt C (2019) Neural rendering and reenactment of human actor videos. ACM Trans Gr 38(5):1–14
29. Christos Doukas M, Koujan MR, Sharmanska V, Roussos A, Zafeiriou S (2021) Head2Head++: deep facial attributes retargeting. IEEE Trans Biometrics Behav Identit Sci 3(1):31–43
30. Zakharov E, Shysheya A, Burkov E, Lempitsky V (2019) Few-shot adversarial learning of realistic neural talking head models. In: IEEE/CVF International Conference on Computer Vision (ICCV), Seoul
31. Wang T-C, Liu M-Y, Tao A, Liu G, Kautz J, Catanzaro B (2019) Few-shot video-to-video synthesis. In: Advances in Neural Information Processing Systems (NeurIPS), Vancouver
32. Gafni O, Ashual O, Wolf L (2021) Single-shot freestyle dance reenactment. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville
33. Zhang J, Zeng X, Pan Y, Liu Y, Ding Y, Fan C (2019) FaceSwapNet: landmark guided many-to-many face reenactment. http://arxiv.org/abs/1905.11805v1
34. Zhang Y, Zhang S, He Y, Li C, Loy CC, Liu Z (2019) One-shot face reenactment. http://arxiv.org/abs/1908.03251v1
35. Gu K, Zhou Y, Huang T (2020) FLNet: landmark driven fetching and learning network for faithful talking facial animation synthesis. In: Proceedings of the AAAI Conference on Artificial Intelligence, Hilton New York Midtown
36. Lee J, Ramanan D, Girdhar R (2020) MetaPix: few-shot video retargeting. In: International conference on learning representations
37. Sanchez E, Valstar M (2020) A recurrent cycle consistency loss for progressive face-to-face synthesis. In: IEEE international conference on automatic face and gesture recognition, Buenos Aires
38. Tripathy S, Kannala J, Rahtu E (2021) FACEGAN: facial attribute controllable rEenactment GAN. In: IEEE winter conference on applications of computer vision (WACV), Waikoloa
39. Lee C-H, Liu Z, Wu L, Luo P (2020) MaskGAN: towards diverse and interactive facial image manipulation. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR), Seattle
40. Zhu Z, Huang T, Shi B, Yu M, Wang B, Bai X (2019) Progressive pose attention transfer for person image generation. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR), Long Beach
41. Aberman K, Shi M, Liao J, Lischinski D, Cohen-Or D, Chen B (2019) Deep video-based performance cloning. In: European association for computer graphics, Genova
42. Zhou Y, Wang Z, Fang C, Bui T, Berg TL (2019) Dance dance generation: motion transfer for internet videos. In: IEEE/CVF international conference on computer vision workshop (ICCVW), Seoul
43. Tripathy S, Kannala J, Rahtu E (2020) ICface: interpretable and controllable face reenactment using GANs. In: IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass
44. Zablotskaia P, Siarohin A, Zhao B, Sigal L (2019) DwNet: dense warp-based network for pose-guided human video generation. In: British Machine Vision Conference (BMVC), Cardiff
45. Suwajanakorn S, Seitz SM, Kemelmacher-Shlizerman I (2017) Synthesizing Obama: learning lip sync from audio. ACM Trans Gr 36(4):1–14
46. Fried O, Tewari A, Zollhöfer M, Finkelstein A, Shechtman E, Goldman DB, Genova K, Jin Z, Theobalt C, Agrawala M (2019) Text-based editing of talking-head video. ACM Trans Gr 38(4):1–14
47. Lahiri A, Kwatra V, Frueh C, Lewis J, Bregler C (2021) LipSync3D: data-efficient learning of personalized 3D talking faces from video using pose and lighting normalization. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville
48. Zhang Z, Li L, Ding Y, Fan C (2021) Flow-guided one-shot talking face generation with a high-resolution audio-visual dataset. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR), Nashville
49. Jamaludin A, Chung JS, Zisserman A (2019) You said that?: Synthesising talking faces from audio. Int J Comput Vis 127:1767–1779
50. Choi Y, Choi M, Kim M, Ha J-W, Kim S, Choo J (2018) StarGAN: unified generative adversarial networks for multi-domain image-to-image translation. In: IEEE/CVF conference on computer vision and pattern recognition, Salt Lake City
51. Pumarola A, Agudo A, Martinez AM, Sanfeliu A, Moreno-Noguer F (2019) GANimation: one-shot anatomically consistent facial animation. Int J Comput Vis 128:698–713
52. Liu M, Ding Y, Xia M, Liu X, Ding E, Zuo W, Wen S (2019) STGAN: a unified selective transfer network for arbitrary image attribute editing. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach
53. Liang H, Hou X, Shen L (2021) SSFlow: style-guided neural spline flows for face image manipulation. In: Proceedings of the 29th ACM international conference on multimedia, New York
54. Wang R, Chen J, Yu G, Sun L, Yu C, Gao C, Sang N (2021) Attribute-specific Control Units in StyleGAN for Fine-grained image manipulation. In: Proceedings of the 29th ACM international conference on multimedia, New York
55. Karras T, Laine S, Aila T (2019) A style-based generator architecture for generative adversarial networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach
56. Zhou H, Liu Y, Liu Z, Luo P, Wang X (2019) Talking face generation by adversarially disentangled audio-visual representation. In: AAAI Conference on Artificial Intelligence (AAAI-19), Honolulu

57. Chen L, Maddox RK, Duan Z, Xu C (2019) Hierarchical cross-modal talking face generation with dynamic pixel-wise loss. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach
58. Vougioukas K, Petridis S, Pantic M (2019) Realistic speech-driven facial animation with GANs. Int J Comput Vis 128:1398–1413
59. Thies J, Elgharib M, Tewari A, Theobalt C, Nießner M (2020) Neural voice puppetry: audio-driven facial reenactment. In: European conference on computer vision (ECCV), Glasgow
60. Vougioukas K, Petridis S, Pantic M (2019) End-to-end speech-driven realistic facial animation with temporal GANs In: Computer Vision and Pattern Recognition (CVPR), Long Beach
61. He Z, Zuo W, Kan M, Shan S, Chen X (2019) AttGAN: facial attribute editing by only changing what you want. IEEE Trans Image Process 28(11):5464–5478
62. Shen Y, Gu J, Tang X, Zhou B (2020) Interpreting the latent space of GANs for semantic face editing. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle
63. Jo Y, Park J (2019) SC-FEGAN: face editing generative adversarial network with user's sketch and color. In: IEEE/CVF international conference on computer vision (ICCV), Seoul
64. Shen Y, Yang C, Tang X, Zhou B (2020) InterFaceGAN: interpreting the disentangled face representation learned by GANs. IEEE Trans Pattern Anal Mach Intell (Early Access), p 1
65. Fu C, Hu Y, Wu X, Wang G, Zhang Q, He R (2021) High-fidelity face manipulation with extreme poses and expressions. IEEE Trans Inf Forensics Secur 16:2218–2231
66. Yang N, Zheng Z, Zhou M, Guo X, Qi L, Wang T (2021) A domain-guided noise-optimization-based inversion method for facial image manipulation. IEEE Trans Image Process 30:6198–6211
67. Karras T, Aila T, Laine S, Lehtinen J (2018) Progressive Growing of GANs for improved quality, stability, and variation. In: International conference on learning representations (ICLR), Vancouver
68. Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, Aila T (2020) Analyzing and improving the image quality of StyleGAN. In: IEEE/CVF Conference on computer vision and pattern recognition (CVPR), Seattle
69. Zhang H, Goodfellow I, Metaxas D, Odena A (2019) Self-attention generative adversarial networks. http://arxiv.org/abs/1805.08318v2
70. Brock A, Donahue J, Simonyan K (2019) Self-attention generative adversarial networks. In: International Conference on Learning Representations (ICLR), New Orleans
71. Martin K, Marketing V (2021) What is voice cloning?, ID R&D, https://www.idrnd.ai/what-is-voice-cloning/. Accessed 24 July 2021
72. Maheshwari H (2021) Basic text to speech, explained," towards data Science, https://towardsdatascience.com/text-to-speech-explained-from-basic-498119aa38b5. Accessed 11 July 2021
73. Maheshwari H (2021) Text to speech system for multi-speaker setting, towards data science, https://towardsdatascience.com/text-to-speech-system-for-multi-speaker-setting-35e83f84e669. Accessed 12 July 2021
74. Singh J (2018) WaveNet: google Assistant's voice synthesizer, towardsdatascience, 7 November 2018. https://towardsdatascience.com/wavenet-google-assistants-voice-synthesizer-a168e9af13b1. Accessed 10 July 2021
75. Oord AVD, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior A, Kavukcuoglu K (2016) WaveNet: a generative model for raw audio. In: Proceedings of the 9th ISCA Speech Synthesis Workshop, Sunnyvale
76. Oord A, Li Y, Babuschkin I, Simonyan K, Vinyals O, Kavukcuoglu K, Driessche G, Lockhart E, Cobo L, Stimberg F, Casagrande N, Grewe D, Noury S, Dieleman S, Elsen E, Kalchbrenner N, Zen H, Graves A, King H, Walters T, Belov D, Hassabis D (2018) Parallel WaveNet: fast high-fidelity speech synthesis. In: Proceedings of the 35th international conference on machine learning, Stockholm
77. Arık SO, Chrzanowski M, Coates A, Diamos G, Gibiansky A, Kang Y, Li X, Miller J, Ng A, Raiman J, Sengupta S, Shoeybi M (2017) Deep voice: real-time neural text-to-speech. In: International conference on machine learning, Sydney
78. Arık SÖ, Diamos G, Gibiansky A, Miller J, Peng K, Ping W, Raiman J, Zhou Y (2017) Deep voice 2: multi-speaker neural text-to-speech. In: Advances in neural information processing systems, Long Beach
79. Ping W, Peng K, Gibiansky A, Arık SO, Kannan A, Narang S, Raiman J, Miller J (2018) Deep voice 3: scaling text-to-speech with convolutional sequence learning. In: International conference on learning representations (ICLR), Vancouver
80. Wang Y, Skerry-Ryan R, Stanton D, Wu Y, Weiss RJ, Jaitly N, Yang Z, Xiao Y, Chen Z, Bengio S, Le Q, Agiomyrgiannakis Y, Clark R, Saurous RA (2017) Tacotron: towards end-to-end Speech Synthesis. http://arxiv.org/abs/ 1703.10135v2
81. Zhang J-X, Ling Z-H, Liu L-J, Jiang Y, Dai L-R (2019) Sequence-to-sequence acoustic modeling for voice conversion. IEEE/ACM Trans Audio Speech Lang Process 27(3):631–644
82. Veaux C, Yamagishi J, King S (2013) Towards personalized synthesized voices for individuals with vocal disabilities: voice banking and reconstruction. In: Speech and language processing for assistive technologies (SLPAT), Grenoble
83. Sisman B, Yamagishi J, King S, Li H (2021) An overview of voice conversion and its challenges: from statistical modeling to deep learning. IEEE/ACM Trans Audio Speech Lang Process 29:132–157
84. Zhang J-X, Ling Z-H, Dai L-R (2019) Non-parallel sequence-to-sequence voice conversion with disentangled linguistic and speaker representations. IEEE/ACM Trans Audio Speech Lang Process 28:540–552
85. Wang R, Ding Y, Li L, Fan C (2020) One-shot voice conversion using Star-GAN. In: ICASSP 2020 - 2020 IEEE international conference on acoustics, speech and signal processing (ICASSP), Barcelona
86. Liu R, Chen X, Wen X (2020) Voice conversion with transformer network. In: ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP), Barcelona
87. Yasuda Y, Wang X, Takaki S, Yamagishi J (2019) Investigation of enhanced tacotron text-to-speech synthesis systems with self-attention for pitch accent language. In IEEE international conference on acoustics, speech and signal processing (ICASSP), Brighton
88. Chen Y, Assael Y, Shillingford B, Budden D, Reed S, Zen H, Wang Q, Cobo LC, Trask A, Laurie B, Gulcehre C, Oord AVD, Vinyals O, Freitas ND (2019) Sample efficient adaptive text-to-speech. In: International Conference on Learning Representations (ICLR), New Orleans
89. Liu R, Yang J, Liu M (2019) A new end-to-end long-time speech synthesis system based on Tacotron2. In: International conference proceeding series (ICPS), Beijing
90. Weiss RJ, Skerry-Ryan R, Battenberg E, Mariooryad S, Kingma DP (2021) Wave-Tacotron: spectrogram-free end-to-end text-to-speech synthesis. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto
91. He Q, Xiu Z, Koehler T, Wu J (2021) Multi-rate attention architecture for fast streamable text-to-speech spectrum modeling. In: 2021 IEEE international conference on acoustics, speech and signal processing (ICASSP), Toronto

92. Liu R, Sisman B, Gao G, Li H (2021) Expressive TTS training with frame and style reconstruction loss. IEEE/ACM Trans Audio Speech Lang Process 29:1806–1818
93. Zhou X, Ling Z-H, Dai L-R (2021) UnitNet: a sequence-to-sequence acoustic model for concatenative speech synthesis. IEEE/ACM Trans Audio Speech Lang Process 29:2643–2655
94. Tanaka K, Kameoka H, Kaneko T, Hojo N (2019) ATTS2S-VC: sequence-to-sequence voice conversion with attention and context preservation mechanisms. In: ICASSP 2019 - 2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), Brighton
95. Kameoka H, Kaneko T, Tanaka K, Hojo N (2019) ACVAE-VC: non-parallel voice conversion with auxiliary classifier variational autoencoder. IEEE/ACM Trans Audio Speech Lang Process 27(9):1432–1443
96. Cong J, Yang S, Xie L, Yu G, Wan G (2020) Data efficient voice cloning from noisy samples with domain adversarial training. In: Interspeech 2020, Shanghai
97. Zhang M, Sisman B, Zhao L, Li H (2020) DeepConversion: voice conversion with limited parallel training data. Speech Commun 122:31–43
98. Kameoka H, Tanaka K, Kwaśny D, Kaneko T, Hojo N (2020) ConvS2S-VC: fully convolutional sequence-to-sequence voice conversion. IEEE/ACM Trans Audio Speech Lang Process 28:1849–1863
99. Ding S, Zhao G, Gutierrez-Osuna R (2020) Improving the speaker identity of non-parallel many-to-many voice conversion with adversarial speaker recognition. In: INTERSPEECH, Shanghai
100. Lee S, Ko B, Lee K, Yoo I-C, Yook D (2020) Many-to-many voice conversion using conditional cycle-consistent adversarial networks. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona
101. Zhang M, Zhou Y, Zhao L, Li H (2021) Transfer learning from speech synthesis to voice conversion with non-parallel training data. IEEE/ACM Trans Audio Speech Lang Process 29:1290–1302
102. Chen M, Shi Y, Hain T (2021) Towards low-resource stargan voice conversion using weight adaptive instance normalization. In: IEEE international conference on acoustics, speech and signal processing (ICASSP), Toronto
103. Li Z, Tang B, Yin X, Wan Y, Xu L, Shen C, Ma Z (2021) PPG-based singing voice conversion with adversarial representation learning. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto
104. Kameoka H, Huang W-C, Tanaka K, Kaneko T, Hojo N, Toda T (2021) Many-to-many voice transformer network. IEEE/ACM Trans Audio Speech Lang Process 29:656–670
105. Li H, Li B, Tana S, Huang J (2020) Identification of deep network generated images using disparities in color components. Signal Process 174:107616
106. Chen P, Liu J, Liang T, Yu C, Zou S, Dai J, Han J (2021) DLFM-Net: end-to-end detection and localization of face manipulation using multi-domain features. In: IEEE international conference on multimedia and expo (ICME), Shenzhen
107. McCloskey S, Albright M (2018) Detecting GAN-generated imagery using color cues. http://arxiv.org/abs/ 1812.08247v1
108. Yu N, Davis L, Fritz M (2019) Attributing fake images to GANs: learning and analyzing GAN fingerprints. In: IEEE/CVF international conference on computer vision (ICCV), Seoul
109. Koopman M, Rodriguez AM, Geradts Z (2018) Detection of deepfake video manipulation. In: Irish machine vision and image processing conference (IMVIP), Belfast
110. Li Y, Lyu S (2019) Exposing DeepFake videos by detecting face warping artifacts. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR) workshops, Long Beach
111. Li L, Bao J, Zhang T, Yang H, Chen D, Wen F, Guo B (2020) Face X-ray for more general face forgery detection. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR), Seattle
112. Matern F, Riess C, Stamminger M (2019) Exploiting visual artifacts to expose deepfakes and face manipulations. In: IEEE winter applications of computer vision workshops (WACVW), Waikoloa
113. Zhao Y, Ge W, Li W, Wang R, Zhao L, Ming J (2019) Capturing the persistence of facial expression features for deepfake video detection. In: International Conference on Information and Communications Security, Beijing
114. Li X, Yu K, Ji S, Wang Y, Wu C, Xue H (2020) Fighting against deepfake: Patch&Pair convolutional neural networks (PPCNN). In: Companion Proceedings of the Web Conference 2020, New York
115. Lee S, Tariq S, Shin Y, Woo SS (2021) Detecting handcrafted facial image manipulations and GAN-generated facial images using Shallow-FakeFaceNet. Appl Soft Comput 105:107256
116. Shang Z, Xie H, Zha Z, Yu L, Li Y, Zhang Y (2021) PRRNet: Pixel-Region relation network for face forgery detection. Pattern Recognit 116:107950
117. Agarwal S, Farid H, Fried O, Agrawala M (2020) Detecting deepfake videos from phoneme-viseme mismatches. In: IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW), Seattle
118. Mittal T, Bhattacharya U, Chandra R, Bera A, Manocha D (2020) Emotions don't lie: an audio-visual deepfake detection method using affective cues. In: ACM international conference on multimedia, New York
119. Chugh K, Gupta P, Dhall A, Subramanian R (2020) Not made for each other- audio-visual dissonance-based deepfake detection and localization. In: ACM international conference on multimedia, New York
120. Hosier BC, Stamm MC (2020) Detecting video speed manipulation. In: IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW), Seattle
121. Amerini I, Galteri L, Caldelli R, Bimbo AD (2019) Deepfake video detection through optical flow based CNN. In: IEEE/CVF international conference on computer vision workshop (ICCVW), Seoul.
122. Caldelli R, Galteri L, Amerini I, Bimbo AD (2021) Optical Flow based CNN for detection of unlearnt deepfake manipulations. Pattern Recognit Lett 146:31–37
123. Yang X, Li Y, Lyu S (2019) Exposing deep fakes using inconsistent head poses. In: IEEE international conference on acoustics, speech and signal processing (ICASSP), Brighton
124. Li Y, Chang M-C, Lyu S (2018) In Ictu Oculi: exposing AI created fake videos by detecting eye blinking. In: IEEE International Workshop on Information Forensics and Security (WIFS), Hong Kong
125. Qi H, Guo Q, Juefei-Xu F, Xie2 X, Ma L, Feng W, Liu Y, Zhao J (2020) DeepRhythm: exposing DeepFakes with attentional visual heartbeat rhythms. In: ACM international conference on multimedia, New York
126. Ciftci UA, Demir I, Yin L (2020) FakeCatcher: detection of synthetic portrait videos using biological signals, IEEE Trans Pattern Anal Mach Intell (Early Access)
127. Hernandez-Ortega J, Tolosana R, Fierrez J, Morales A (2020) DeepFakesON-Phys: deepfakes detection based on heart rate estimation. http://arxiv.org/abs/2010.00400v3
128. Yasrab R, Jiang W, Riaz A (2021) Fighting deepfakes using body language analysis. Forecast MDPI Open Access J 3(2):1–19
129. Khalid H, Woo SS (2020) OC-FakeDect: classifying deepfakes using one-class variational Autoencoder. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle

130. Xuan X, Peng B, Wang W, Dong J (2019) On the generalization of GAN image forensics. In: Chinese conference on biometric recognition, Zhuzhou
131. Zhou P, Han X, Morariu VI, Davis LS (2017) Two-stream neural networks for tampered face detection. In: IEEE conference on computer vision and pattern recognition workshops (CVPRW), Honululu
132. Jeon H, Bang Y, Woo SS (2019) FakeTalkerDetect: effective and practical realistic neural talking head detection with a highly unbalanced dataset. In: IEEE/CVF international conference on computer vision workshop (ICCVW), Seoul
133. Wu X, Xie Z, Gao Y, Xiao Y (2020) SSTNet: detecting manipulated faces through spatial, steganalysis and temporal features. In: IEEE International conference on acoustics, speech and signal processing (ICASSP), Barcelona
134. Tariq S, Lee S, Kim H, Shin Y, Woo SS (2019) GAN is a friend or foe? A framework to detect various fake face images. In: Proceedings of the 34th ACM/SIGAPP symposium on applied computing, Cyprus
135. Sohrawardi SJ, Chintha A, Thai B, Seng S, Hickerson A, Ptucha R, Wright MK (2019) Poster: towards robust open-world detection of deepfakes. In: ACM SIGSAC conference on computer and communications security, London
136. Fernando T, Fookes C, Denman S, Sridharan S (2019) Exploiting human social cognition for the detection of fake and fradulent faces via memory networks. http://arxiv.org/abs/1911.07844v1
137. Sun X, Wu B, Chen W (2020) Identifying invariant texture violation for robust deepfake detection. http://arxiv.org/abs/2012.10580v1
138. Ding X, Raziei Z, Larson EC, Olinick EV, Krueger P, Hahsler M (2020) Swapped face detection using deep learning and subjective assessment. EURASIP J Inf Secur, vol. 6
139. Kumar A, Bhavsar, A, Verma R (2020) Detecting deepfakes with metric learning. In: International Workshop on Biometrics and Forensics (IWBF), Porto
140. .Rana MS, Sung AH (2020) DeepfakeStack: a deep ensemble-based learning technique for deepfake detection. In: IEEE international conference on cyber security and cloud computing, New York
141. Zhou X, Wang Y, Wu P (2020) Detecting deepfake videos via frame serialization learning. In: IEEE 3rd International Conference of Safe Production and Informatization (IICSPI), Chongqing City
142. Nguyen XH, Tran TS, Le VT, Nguyen KD, Truong D-T (2021) Learning Spatio-temporal features to detect manipulated facial videos created by the Deepfake techniques. Forensic Sci Int Digital Investig 36:301108
143. Xu Z, Liu J, Lu W, Xu B, Zhao X, Li B, Huang J (2021) Detecting facial manipulated videos based on set convolutional neural networks. J Vis Commun Image Represent 77:103119
144. Chen Z, Yang H (2021) Attentive semantic exploring for manipulated face detection. In: IEEE International Conference on Acoustics, Speech and Signal Processing, Toronto
145. Zhang J, Ni J, Xie H (2021) DeepFake videos detection using self-supervised decoupling network. In: IEEE International Conference on Multimedia and Expo (ICME), Shenzhen
146. Gu Z, Chen Y, Yao T, Ding S, Li J, Huang F, Ma L (2021) Spatiotemporal inconsistency learning for deepfake video detection. In: Proceedings of the 29th ACM international conference on multimedia, New York
147. Tu Y, Liu Y, Li X (2021) Deepfake video detection by using convolutional gated recurrent unit. In: International conference on machine learning and computing, Shenzhen
148. Zhuang Y-X, Hsu C-C (2019) Detecting generated image based on a coupled network with two-step pairwise learning. In: IEEE international conference on image processing (ICIP), Taipei
149. Lima OD, Franklin S, Basu S, Karwoski B, George A (2020) Deepfake detection using spatiotemporal convolutional networks. http://arxiv.org/abs/2006.14749v1
150. Lang Y, Li X, Chen Y, Mao X, He Y, Wang S, Xue H, Lu Q (2020) Sharp multiple instance learning for deepfake video detection. In: Proceedings of the 28th ACM international conference on multimedia, Seattle WA
151. Chen B, Ju X, Xiao B, Ding W, Zheng Y, Albuquerque VHCD (2021) Locally GAN-generated face detection based on an improved Xception. Inf Sci 572:16–28
152. Chen H-S, Rouhsedaghat M, Ghani H, Hu S, You S, Kuo C-CJ (2021) DefakeHop: a light-weight high-performance deepfake detector. In: IEEE International Conference on Multimedia and Expo (ICME), Shenzhen
153. Das S, Seferbekov S, Datta A, Islam MS, Amin MR (2021) Towards solving the deepfake problem : an analysis on improving deepfake detection using dynamic face augmentation. In: IEEE/CVF international conference on computer vision workshops (ICCVW), Montreal
154. Nguyen HH, Fang F, Yamagishi J, Echizen I (2019) Multi-task learning for detecting and segmenting manipulated facial images and videos. In: IEEE 10th international conference on biometrics theory, applications and systems (BTAS), Tampa
155. Du M, Pentyala SK, Li Y, Hu X (2020) Towards generalizable deepfake detection with locality-aware autoencoder. In: ACM international conference on information & knowledge management, Virtual Event Ireland
156. He P, Li H, Wang H (2019) Detection of fake images via the ensemble of deep representations from multi color spaces. In: IEEE International conference on image processing (ICIP), Taipei
157. Guo Z, Yang G, Chen J, Sun X (2021) Fake face detection via adaptive manipulation traces extraction network. Comput Vis Image Underst 204:103170
158. Wang R, Juefei-Xu F, Ma L, Xie X, Huang Y, Wang J, Liu Y (2020) FakeSpotter: a simple yet robust baseline for spotting AI-synthesized fake faces. In: International joint conference on artificial intelligence (IJCAI), Yokohama
159. Khan SA, Dai H (2021) Video transformer for deepfake detection with incremental learning. In: Proceedings of the 29th ACM international conference on multimedia, New York
160. Frank J, Eisenhofer T, Schonherr L, Fischer A, Kolossa D, Holz T (2020) Leveraging frequency analysis for deep fake image recognition. Proc of Mach Learn 119:3247–3258
161. Durall R, Keuper M, Pfreundt F-J, Keuper J (2020) Unmasking deepfakes with simple feature. http://arxiv.org/abs/1911.00686v3
162. Masi I, Killekar A, Mascarenha RM, Gurudatt SP, AbdAlmageed W (2020) Two-branch recurrent network for isolating deepfakes in videos. In: European conference on computer vision, Glasgow
163. McCloskey S, Albright M (2019) Detecting GAN-generated imagery using saturation cues. In: IEEE International conference on image processing (ICIP), Taipei
164. Guarnera L, Giudice O, Battiato S (2020) DeepFake detection by analyzing convolutional traces. In: IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW), Seattle
165. Wang S-Y, Wang O, Zhang R, Owens A, Efros AA (2020) CNN-generated images are surprisingly easy to spot... for now. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR), Seattle
166. Lugstein F, Baier S, Bachinger G, Uhl A (2021) PRNU-based deepfake detection. In: Proceedings of the 2021 ACM workshop on information hiding and multimedia security
167. Nirkin Y, Wolf L, Keller Y, Hassner T (2020) DeepFake detection based on discrepancies between faces and their context. http://arxiv.org/abs/2008.12262v1

168. Yang J, Xiao S, Li A, Lan G, Wang H (2021) Detecting fake images by identifying potential texture difference. Futur Gener Comput Syst 125:127–135
169. Li G, Cao Y, Zhao X (2021) Exploiting facial symmetry to expose deepfakes. In: IEEE international conference on image processing (ICIP), Anchorage
170. Luo Z, Kamata S-I, Sun Z (2021) Transformer and node-compressed dnn based dual-path system for manipulated face detection. In: IEEE international conference on image processing (ICIP), Anchorage
171. Yang J, Xiao S, Li A, Lu W, Gao X, Li Y (2021) MSTA-net: forgery detection by generating manipulation trace based on multi-scale self-texture attention. IEEE Trans Circuits Syst Video Technol ( Early Access ), pp. 1–1
172. Bonomi M, Pasquini C, Boato G (2021) Dynamic texture analysis for detecting fake faces in video sequences. J Vis Commun Image Represent 79:103239
173. Yang J, Li A, Xiao S, Lu W, Gao X (2021) MTD-Net: learning to detect deepfakes images by multi-scale texture difference. IEEE Trans Inf Forensics Secur 16:4234–4245
174. Gu Y, He M, Nagano K, Li H (2019) Protecting world leaders against deep fakes. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR) Workshops, Long Beach
175. Yang C-Z, Ma J, Wang S-L, Liew AW-C (2020) Preventing deepfake attacks on speaker authentication by dynamic lip movement analysis. IEEE Trans Inf Forensics Secur 16:1841–1854
176. Hosler B, Salvi D, Murray A, Antonacci F, Bestagini P, Tubaro S, Stamm MC (2021) Do deepfakes feel emotions? A semantic approach to detecting deepfakes via emotional inconsistencies. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville
177. Demir İ, Ciftci UA (2021) Where do deep fakes look? Synthetic face detection via gaze. In ACM symposium on eye tracking research and applications, Germany
178. Hu S, Li Y, Lyu S (2021) Exposing GAN-generated faces using inconsistent corneal specular highlights. In: IEEE international conference on acoustics, speech and signal processing (ICASSP), Toronto
179. Agarwal S, Farid H (2021) Detecting deep-fake videos from aural and oral dynamics. In: IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW), Nashville
180. Sabir E, Cheng J, Jaiswal A, AbdAlmageed W, Masi I, Natarajan P (2019) Recurrent-convolution approach to deepfake detection – state-of-art results on FaceForensics++. http://arxiv.org/abs/1905.00582v1
181. Sabir E, Cheng J, Jaiswal A, AbdAlmageed W, Masi I, Natarajan P (2019) Recurrent convolutional strategies for face manipulation detection in videos. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR), Long Beach
182. Amerini I, Caldelli R (2020) Exploiting prediction error inconsistencies through LSTM-based classifiers to detect deepfake videos. In: ACM workshop on information hiding and multimedia security, New York
183. Lu C, Liu B, Zhou W, Chu Q, Yu N (2021) Deepfake video detection using 3D-attentional inception convolutional neural network. In: IEEE international conference on image processing (ICIP), Anchorage
184. Trinh L, Tsang M, Rambhatla S, Liu Y (2021) Interpretable and trustworthy deepfake detection via dynamic prototypes. In: IEEE winter conference on applications of computer vision (WACV), Hawaii
185. Cozzolino D, Thies J, Rossler A, Riess C, Nießner M, Verdoliva L (2019) ForensicTransfer: weakly-supervised domain adaptation for forgery detection. http://arxiv.org/abs/1812.s02510v2
186. Hsu C-C, Zhuang Y-X, Lee C-Y (2019) Deep fake image detection based on pairwise learning. Appl Sci 10(1):370
187. Dang LM, Hassan SI, Im S, Moon H (2019) Face image manipulation detection based on a convolutional neural network. Expert Syst Appl 129:156–168
188. Nguyen HH, Yamagishi J, Echizen I (2019) Capsule-forensics: using capsule networks to detect forged images and videos. In: IEEE international conference on acoustics, speech and signal processing (ICASSP), Brighton
189. Montserrat DM, Hao H, Yarlagadda SK, Baireddy S, Shao R, Horváth J, Bartusiak E, Yang J, Güera D, Zhu F, Delp EJ (2020) Deepfakes detection with automatic face weighting. In: IEEE/CVF conference on computer vision and pattern recognition workshops, Seattle
190. Choi DH, Lee HJ, Lee S, Kim JU, Ro YM (2020) Fake video detection with certainty-based attention network. In: IEEE international conference on image processing (ICIP), Abu Dhabi
191. Chintha A, Thai B, Sohrawardi SJ, Bhatt K, Hickerson A, Wright M, Ptucha R (2020) Recurrent convolutional structures for audio spoof and video deepfake detection. IEEE J Sel Top Signal Process 14(5):1024–1037
192. Hu J, Wang S, Li X (2021) Improving the generalization ability of deepfake detection via disentangled representation learning. In: IEEE international conference on image processing (ICIP), Anchorage
193. Hu J, Liao X, Wang W, Qin Z (2021) Detecting compressed deepfake videos in social networks using frame-temporality two-stream convolutional network. IEEE Trans Circuits Syst Video Technol (Early Acces) 32(3):1089–1102
194. Han B, Han X, Zhang H, Li J, Cao X (2021) Fighting fake news: two stream network for deepfake detection via learnable SRM. IEEE Trans Biometrics Behav Ident Sci 3(3):320–331
195. Kim M, Tariq S, Woo SS (2021) FReTAL: generalizing deepfake detection using knowledge distillation and representation learning. In: IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW), Nashville
196. Zhao H, Wei T, Zhou W, Zhang W, Chen D, Yu N (2021) Multi-attentional deepfake detection. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR), Nashville
197. Sun Z, Han Y, Hua Z, Ruan N, Jia W (2021) Improving the efficiency and robustness of deepfakes detection through precise geometric features. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR), Nashville
198. Tariq S, Lee S, Woo SS (2021) One detector to rule them all. In: Proceedings of the web conference 2021, New York
199. Wang R, Juefei-Xu F, Huang Y, Guo Q, Xie X, Ma L, Liu Y (2020) DeepSonar: towards effective and robust detection of AI-synthesized fake voices. In: Proceedings of the 28th ACM international conference on multimedia, Seattle
200. Balamurli B, Lin KE, Lui S, Chen J-M, Herremans D (2019) Toward robust audio spoofing detection: a detailed comparison of traditional and learned features. In: IEEE Access
201. Saranya MS, Padmanabhan R, Murthy HA (2018) Replay attack detection in speaker verification using non-voiced segments and decision level feature switching. In: International conference on signal processing and communications (SPCOM), Bangalore
202. Witkowski M, Kacprzak S, Zelasko P, Kowalczyk K, Gałka J (2017) Audio replay attack detection using high-frequency features. In: INTERSPEECH, Stockholm
203. AlBadawy EA, Lyu S, Farid H (2019) Detecting AI-synthesized speech using bispectral analysis. In: IEEE/CVF Conference on computer vision and pattern recognition (CVPR), Long Beach
204. Patil HA, Kamble MR (2018) A survey on replay attack detection for automatic speaker verification (ASV) system. In: Proceedings of the APSIPA Annual Summit and Conference 2018, Hawai
205. Wijethunga R, Matheesha D, Noman AA, Silva KD, Tissera M, Rupasinghe L (2020) Deepfake audio detection: a deep learning

based solution for group conversations. In: International conference on advancements in computing (ICAC), Malabe
206. Chen T, Kumar A, Nagarsheth P, Sivaraman G, Khoury E (2020) Generalization of audio deepfake detection. In: Odyssey 2020 the speaker and language recognition workshop, Tokyo
207. Shim H-J, Jung J-W, Heo H-S, Yoon S-H, Yu H-J (2018) Replay spoofing detection system for automatic speaker verification using multi-task learning of noise classes. In: Conference on technologies and applications of artificial intelligence (TAAI), Taichung
208. Yang J, Das RK (2020) Long-term high frequency features for synthetic speech detection. Digital Signal Process 97:102622
209. Malik H (2019) Securing voice-driven interfaces against Fake (Cloned) Audio Attacks. In: IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose
210. Gunendradasan T, Irtza S, Ambikairajah E, Epps J (2019) Transmission line cochlear model based AM-FM features for replay attack detection. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton
211. Borrelli C, Bestagini P, Antonacci F, Sarti A, Tubaro S (2021) Synthetic speech detection through short-term and long-term prediction traces. EURASIP J Inf Secur, 2
212. Lai C-I, Abad A, Richmond K, Yamagishi J, Dehak N, King S (2019) Attentive filtering networks for audio replay attack detection. In: IEEE international conference on acoustics, speech and signal processing (ICASSP), Brighton
213. Huang L, Pun C-M (2019) Audio replay spoof attack detection using segment-based hybrid feature and DenseNet-LSTM network. In: IEEE international conference on acoustics, speech and signal processing (ICASSP), Brighton
214. Gomez-Alanis A, Peinado AM, Gonzalez JA, Gomez AM (2019) A light convolutional GRU-RNN deep feature extractor for ASV spoofing detection. In: INTERSPEECH, Graz
215. Gomez-Alanis A, Peinado AM, Gonzalez JA, Gomez AM (2021) A gated recurrent convolutional neural network for robust spoofing detection. IEEE/ACM Trans Audio Speech Lang Process 27(12):1985–1999
216. Huang L, Pun C-M (2020) Audio replay spoof attack detection by joint segment-based linear filter bank feature extraction and attention-enhanced DenseNet-BiLSTM Network. IEEE/ACM Trans Audio Speech Lang Process 28:1813–1825
217. Wu Z, Das RK, Yang J, Li H (2020) Light convolutional neural network with feature genuinization for detection of synthetic speech attacks. In: INTERSPEECH, Shanghai
218. Wang Z, Cui S, Kang X, Sun W, Li Z (2021) Densely connected convolutional network for audio spoofing detection. In: Asia-Pacific signal and information processing association annual summit and conference (APSIPA ASC), Auckland
219. You CH, Yang J (2020) Device feature extraction based on parallel neural network training for replay spoofing detection. IEEE/ACM Trans Audio Speech Lang Process 28:2308–2318
220. Luo A, Li E, Liu Y, Kang X, Wang ZJ (2021) A capsule network based approach for detection of audio spoofing attacks. In: IEEE International conference on acoustics, speech and signal processing (ICASSP), Toronto
221. Ren Y, Liu W, Liu D, Wang L (2021) Recalibrated bandpass filtering on temporal waveform for audio spoof detection. In: IEEE International conference on image processing (ICIP), Anchorage
222. Huang L, Zhao J (2021) Audio replay spoofing attack detection using deep learning feature and long-short-term memory recurrent neural network. In: The second international conference on artificial intelligence, information processing and cloud computing, Hangzhou
223. Ouyang M, Das RK, Yang J, Li H (2021) Capsule network based end-to-end system for detection of replay attacks. In: International symposium on chinese spoken language processing (ISCSLP), Hong Kong
224. Li X, Li N, Weng C, Liu X, Su2 D, Yu D, Meng H (2021) Replay and synthetic speech detection with Res2Net architecture. In: IEEE International conference on acoustics, speech and signal processing (ICASSP), Toronto
225. Li Y, Yang X, Sun P, Qi H, Lyu S (2020) Celeb-DF: a large-scale challenging dataset for deepfake forensics. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle
226. Dolhansky B, Bitton J, Pflaum B, Lu J, Howes R, Wang M, Ferrer C C (2020) The deepfake detection challenge (DFDC) dataset. In: http://arxiv.org/2006.07397v4
227. Korshunov P, Marcel S (2018) DeepFakes: a new threat to face recognition? Assessment and Detection. In: http://arxiv.org/1812.08685v1
228. Rossler A, Cozzolino D, Verdoliva L, Riess C, Thies J, Nießner M (2018) FaceForensics: a large-scale video dataset for forgery detection in human faces. http://arxiv.org/1803.09179v1
229. Khodabakhsh A, Ramachandra R, Raja K, Wasnik P, Busch C (2018) Fake face detection methods: can they be generalized? In: International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt
230. Dolhansky B, Howes R, Pflaum B, Baram N, Ferrer CC (2019) The deepfake detection challenge (DFDC) preview dataset. In: http://arxiv.org/1910.08854v2
231. Contributing Data to Deepfake Detection Research, (2019). https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html
232. Jiang L, Li R, Wu W, Qian C, Loy CC (2020) DeeperForensics-1.0: a large-scale dataset for real-world face forgery detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle
233. Zi B, Chang M, Chen J, Ma X, Jiang Y-G (2020) WildDeepfake: a challenging real-world dataset for deepfake detection. In: Proceedings of the 28th ACM international conference on multimedia, Seattle
234. Dong X, Bao J, Chen D, Zhang W, Yu N, Chen D, Wen F, Guo B (2020) Identity-driven deepfake detection. In http://arxiv.org/2012.03930v1
235. Huang J, Wang X, Du B, Du P, Xu C (2021) DeepFake MNIST+: a DeepFake facial animation dataset. In: IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal
236. Kominek J, Black AW (2004) The CMU Arctic speech databases. In: Fifth ISCA Workshop on Speech Synthesis
237. Panayotov V, Chen G, Povey D, Khudanpur S (2015) Librispeech: an ASR corpus based on public domain audio books. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane
238. Wu Z, Kinnunen T, Evans N, Yamagishi J, Hanilc C¸ Sahidullah IM, Sizov A (2015) ASVspoof 2015: the first automatic speaker verification spoofing and countermeasures challenge. In: InterSpeech, Dresden
239. Ito K, Johnson L (2017) The LJ speech dataset, LibriVox project. https://keithito.com/LJ-Speech-Dataset/. Accessed 28 July 2021
240. Delgado H, Todisco1 M, Sahidullah M, Evans N, Kinnunen T, Lee KA, Yamagishi J (2018) ASVspoof 2017 Version 2.0: metadata analysis and baseline enhancements. In: Odyssey 2018—the speaker and language recognition workshop, Les Sables
241. Chung JS, Nagrani A, Zisserman A (2018) VoxCeleb2: Deep speaker recognition. In: *INTERSPEECH*, Hyderabad
242. Veaux C, Yamagishi J, MacDonald K (2019) CSTR VCTK Corpus: English multi-speaker Corpus for CSTR voice cloning toolkit. The Centre for Speech Technology Research (CSTR), University of Edinburgh
243. Reimao R, Tzerpos V (2019) FoR: a dataset for synthetic speech detection. In: International conference on speech technology and human-computer dialogue (SpeD), Timisoara

244. Nagrani A, Chung JS, Xie W, Zisserman A (2020) Voxceleb: Large-scale speaker verification in the wild. Comput Speech Lang 60:101027S
245. GMAIL. The M-AILABS Speech dataset, Caito, https://www.caito.de/2019/01/the-m-ailabs-speech-dataset/. Accessed 28 July 2021].
246. Wang X, Yamagishi J, Todisco M, Delgado H, Nautsch A, Evans N, Sahidullah M, Vestman V, Kinnunen T, Lee KA, Juvela L, Alku P, Peng Y-H, Hwang H-T, Tsao Y, Wang H-M, Maguer SL, Becker M, Henderson F, Clark R, Zhang Y, Wang Q, Jia Y, Onuma K, Mushika K, Kaneda T, Jiang Y, Liu L-J, Wu Y-C, Huang W-C, Toda T, Tanaka K, Kameoka H, Steiner I, Matrouf D, Bonastre J-F, Govender A, Ronanki S, Zhang J-X, Ling Z-H (2020) ASVspoof 2019: a large-scale public database of synthesized, converted and replayed speech. Comput Speech Lang 64:101114
247. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Advances in neural information processing systems, Montreal
248. DepFaceLab, GitHub, [Online]. Available: https://github.com/iperov/DeepFaceLab. Accessed 6 April 2021
249. Deepfakes web, [Online]. Available: https://deepfakesweb.com/. Accessed 6 April 2021
250. FaceApp, [Online]. Available: https://www.faceapp.com/. Accessed 1 April 2021
251. Zao, [Online]. Available: https://zaodownload.com/. Accessed 6 April 2021
252. MachineTube, [Online]. Available: https://www.machine.tube/. Accessed 6 April 2021
253. Doublicat, [Online]. Available: https://reface.app/about/. Accessed 7 April 2021
254. Resemble AI, [Online]. Available: https://www.resemble.ai/. Accessed 28 08 2021
255. Rudrabha/Wav2Lip, github, [Online]. Available: https://github.com/Rudrabha/Wav2Lip
256. Thies J, Zollhöfer M, Theobalt C, Stamminger M, Nießner M (2018) Headon: real-time reenactment of human portrait videos. ACM Trans Gr 37(4):1–13
257. Nguyen TT, Nguyen CM, Nguyen DT, Nguyen DT, Nahavandi S (2019) Deep learning for deepfakes creation and detection http://arxiv.org/1909.11573v1


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

This article has been accepted for publication in IEEE Transactions on Network Science and Engineering. 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/TNSE.2022.3188881



# A N2CNN-based Anomaly Detection Method for Cardiovascular data in Cyber-Physical System

Raju Pal, Ashish Kumar Tripathi, Avinash Chandra Pandey, Mohammad Ayoub Khan, Varun G. Menon and Himanshu Mittal

Abstract—Internet-of-medical-things is the new means of monitoring patient health remotely. However, the real-time detection of anomalies in the patient data is a challenging task, especially on ECG-data. To ease the same, a novel method, NSGA-II based convolution neural network, is presented in this paper for efficient anomaly detection. In the proposed method, non-dominated sorting genetic algorithm-II is employed to obtain optimal hyper-parameters of CNN by evaluating three objective functions namely, accuracy, precision, and recall. Further, the performance validation of the proposed method is conducted on two public datasets and compared against seven state-of-the-art methods. Experimental results affirm that the proposed method outperforms the considered methods with an accuracy of 94.83% and 94.96% on MIT-BIH arrhythmia dataset and INCART dataset, respectively. Therefore, it can be claimed that the proposed method is an efficient alternative for anomaly detection.

Index Terms—Cyber-physical system, Remote health monitoring, Convolutional neural network, Elitist non-dominated sorting genetic algorithm, Anomaly Detection

## I. Introduction

Cyber-physical systems are connecting the world intelligently by interconnecting devices with people through Internet of Things (IoT). This has facilitated different domains of human lifestyle such as entertainment, shopping, and health [1]. With the ability to capture real-time information by sensors, there has been a boom in the IoT-based applications. One of the prominent applications is healthcare. In healthcare, patients and physicians are closely associated through IoT, which advantages in better health monitoring. Moreover, this collaboration of healthcare with IoT has initiated the era of remote health monitoring. In this, patients are geared with a number of health monitoring sensors to capture data like, pulse rate, calories burned, etc. While working with real-time data, this helps in intelligent health management by classifying alarming conditions in a pervasive manner. Fig. 1 demonstrates a remote health monitoring in cyber-physical system from capturing the sensory data to analyzing it for anomaly detection. For efficient analysis, this paper presents a new anomaly detection method for effective health monitoring in the cyber-physical system.

In general, health monitoring can be viewed as a classification process of labeling health signals to different health conditions. One of the most common devices in practice for health monitoring is the electrocardiogram (ECG). ECG captures cardiovascular movements in the form of electrical signals by attaching sensors to the human chest. Generally, ECG signals resemble a pattern of heart movement, which is analyzed to determine the heart condition. However, this analysis requires expertise and is a challenging task for a naive person. In literature, deep-learning models have been quite efficient, especially convolutional neural networks (CNN). However, CNN demands an optimal value for large number of hyper-parameters such as, number of layers, number of nodes, filter size, etc. Moreover, incorrect initialization of CNN hyper-parameters have resulted in local optimal solutions and determination of optimal values for them is a non-trivial and time-consuming task. Recent works have demonstrated that meta-heuristic algorithms are efficient approach to handle the same.

Meta-heuristic based methods are the mathematical models of the nature behaviour for attaining optimization [2]. To attain optimal solution, these algorithms employ intensification (exploitation) and diversification (exploitation). In exploration, existing solutions are updated with a large step-size to examine the search region, while exploitation performs the local search by taking small step-size. This searching behavior helps in finding an optimal solution. Generally, meta-heuristic algorithms consider single-objective function to attain the solution. However, single objective functions considers single perspective of the data which is not efficient for CNN hyper-parameters optimization [3]. On the contrary, multi-objective meta-heuristic algorithms consider multiple criteria such as accuracy, recall, precision, etc. and have exhibited better performance in optimization of CNN hyper-parameters. The considered objectives are generally highly correlated where multi-objective algorithms performs better than single-objective algorithms. Certainly, elitist non-dominated sorting genetic algorithm-II (NSGA-II) is one of the successful multi-objective meta-heuristic

Raju Pal is with Dept. of Computer Science Engineering and Information Technology, JIIT, Noida, India, e-mail:raju3131.pal@gmail.com

Ashish Kumar Triapthi is with Dept. of Computer Science and Engineering, Malaviya National Institute Jaipur, India, e-mail: ashish.cse@mnit.ac.in

A.C. Pandey is with Dept. of Computer Science Engineering, IIITDM, Jabalpur, India, e-mail:avish.nsit@gmail.com

Varun G. Menon is with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam 682024, India (e-mail: varunmenon@ieee.org)

Mohammad Ayoub Khan is with the College of Computing and Information Technology, University of Bisha, Bisha, Saudi Arabia (e-mail: ayoub.khan@ieee.org)

H. Mittal is with the Dept. of Computer Science Engineering and Information Technology, JIIT, Noida, India, e-mail: himanshu.mittal224@gmail.com.

© 2022 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 July 15,2022 at 09:26:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network Science and Engineering. 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/TNSE.2022.3188881



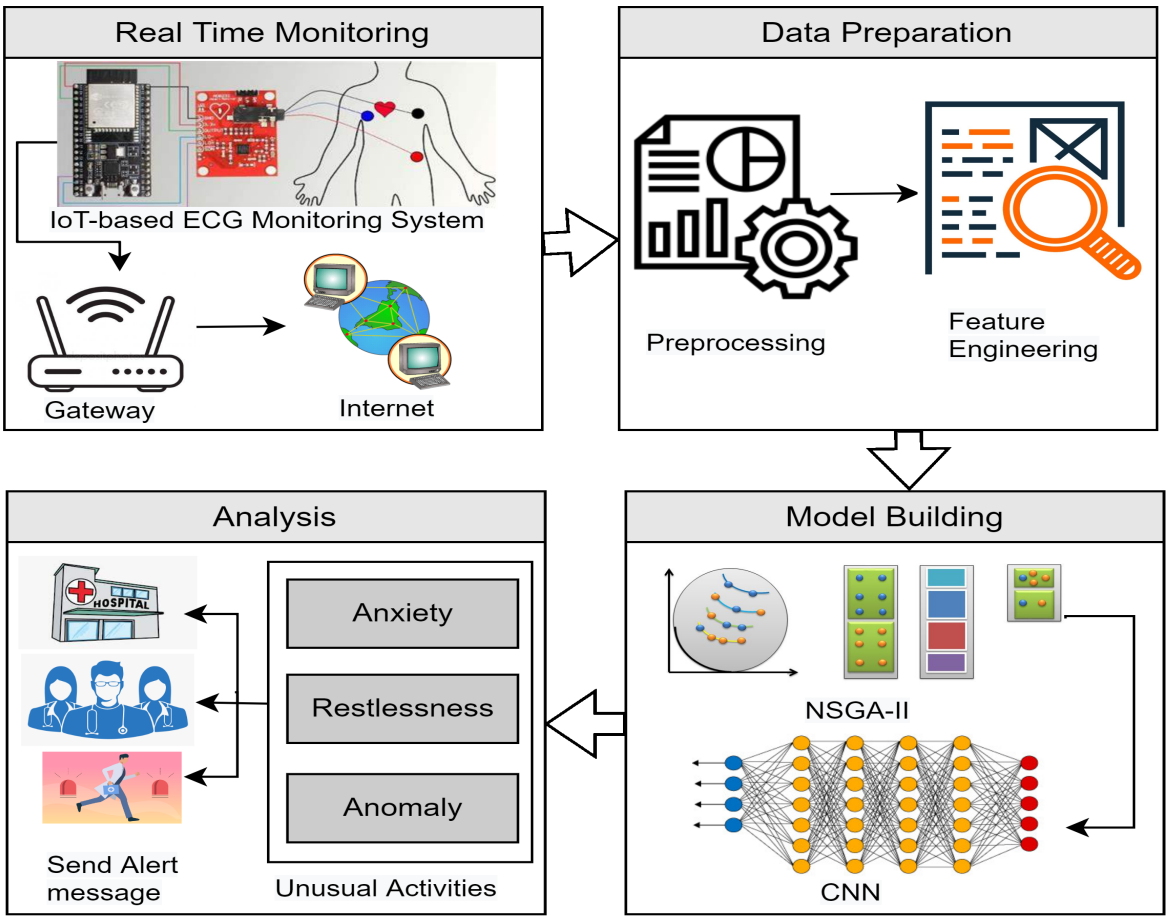



Fig. 1. Framework of the remote health monitoring in cyber-physical system

algorithms. It uses the concept of non-domination sorting for organizing the solutions into different fronts according to multiple objectives. The perseverance of elite solutions and maintaining the diversity in the next generation solutions using crowding distance operators are the main advantages of NSGA-II. Therefore, this paper presents a new automated method to generate an optimized CNN model by leveraging the advantages of NSGA-II.

Thus, the overall contribution of this paper is two folds; (i) Multi-objective NSGA-II is leveraged to optimize the hyper-parameters of CNN model and (ii) The new optimized CNN architecture is employed for automated categorization of sensor generated health signals. An extensive experimental analysis of the proposed method, NSGA-II based CNN, is performed on two public datasets, namely MIT-BIH arrhythmia dataset and INCART dataset. The experimental results are compared against seven state-of-the-art classification based methods.

The organization of the paper is as follows; Section II briefs the related work on the ECG-based methods and optimization of CNN. Section III details the proposed method for the classification of ECG-data. Section IV discusses the considered datasets, while evaluation metrics are presented in Section V. Section VI performs the ablation study on the proposed model and Section VII analysis the experimental performance of the proposed method. Lastly, the conclusion is drawn in Section VIII.

## II. Related Work

In literature, several ECG-based classification methods are presented, which can be mainly classified into machine learning, deep learning, and metaheuristic-based methods [4]. Support vector machine (SVM), linear regression, and linear discriminant are the popular machine learning algorithms used for cardiac arrhythmic classification [5], [6]. Park et al. [7] employed SVM to classify the heartbeats into normal and abnormal categories. Garcia et al. [8] detected the abnormal heart movements using a support vector machine and temporal vector cardiogram. Lannoy et al. [9] used weighted SVM to classify the MIT-BIH imbalance database into respective categories. Huang et al. [10] employed a hierarchical SVM with a maximum voting approach and found considerable improvements in the accuracy of the classification tasks.

Literature witnessed that SVM-based approaches showed better results on balanced datasets. However, they show poor performance for datasets having imbalanced classes [11]. To eliminate the class imbalance problem, a two-stage approach has been proposed. The proposed method first employs the synthetic minority over-sampling

© 2022 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 July 15,2022 at 09:26:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network Science and Engineering. 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/TNSE.2022.3188881



technique (SMOTE) [12] to balance the skewed datasets and then uses CNN for heartbeat classification [13]. Besides, many approaches based on deep learning have been presented to classify heartbeats into normal and abnormal categories [14] [15]. Hammad et al. [16] developed a rule-based ECG classification model on ECG-extracted features through Pan–Tompkins algorithm. In [17], Acharya et al. presented a CNN-based classification method to analyze ECG records. Similarly, Amrani et al. [18] employed a CNN model with Q-Gaussian multi-class SVM for heart monitoring. Further, Sannino et al. [14] applied a deep neural network after performing pre-processing on the ECG data. Mousavi et al. [19] combined the recurrent neural network with CNN to develop the ECG classification method. Moreover, Kachuee et al. [20] employed the concept of transfer learning by using a trained CNN on the ECG dataset. In [21], Wang et al. employed a dual fully connected neural network for the detection of abnormalities in the heart. Ebrahimi et al. [22] presented a concise review of CNN for ECG-based signal classification.

Deep-learning models, particularly CNN, have proven to classify heartbeats effectively. However, it requires an optimal value for many hyper-parameters, such as the number of layers, nodes, and filter size, which is a tedious task to determine. Therefore, several meta-heuristic algorithms have been extensively leveraged for the optimization of CNN to solve numerous real-time problems. Goel et al. [23] presented an automatic CNN-based COVID-19 diagnosis system in which hyper-parameter values are optimized through a grey-wolf optimizer. Zhang et al. [24] fine-tuned the convolutional neural network with a whale optimization algorithm to diagnose skin cancer. Wang et al. [25] employed particle swarm optimization for the hyper-parameter's optimization of CNN. Likewise, Navaneeth and Suchetha [26] optimized the architecture of 1D-CNN with a support vector machine by employing particle swarm optimization and classified chronic kidney diseases. However, multi-objective meta-heuristic algorithms have gained increasing attention for CNN optimization due to their ability to achieve optimal solutions with conflicting objectives. Azad et al. [27] utilized a multi-objective genetic algorithm on a neural network to optimize power generation and water desalination in a solar chimney system.

Xie et al. [28] optimized the neural network using multi-objective swarm approach based on decomposition for finding optimal feed rate of the spindle in computerized numerical control milling. Bakar et al. [29] presented NSGA-II based CNN to design the low Reynolds number airfoil. Recently, Goel et al. [30] performed COVID-19 diagnosis by presenting a two-layer deep neural network which is optimized through multi-objective grasshopper optimization algorithm. Moreover, Kanwal et al. [31] evolved the convolutional auto-encoder to obtain optimal architecture by using multi-objective particle swarm optimization. Singh et al. [32] optimized the hyper-parameters of CNN with multi-objective DE (differential evolution) for medical image classification. A comprehensive survey on existing multi-objective meta-heuristic algorithms for hyper-parameter optimization of different deep neural networks is studied in [33], [34].

## III. Proposed method

In the proposed approach, we are using convolution neural network (CNN) optimized with enhanced NSGA-II with three objective functions, namely accuracy, specificity, and sensitivity. Moreover, three health monitoring datasets are leveraged to validate the efficacy of the proposed N2CNN model. A simplified architecture of the proposed approach is depicted in Fig. 2. Further, each component of the N2CNN is described in the following sections.

### A. Convolution Neural Network

CNN is one of the popular deep learning model, equipped with a number of distinct layers. In this paper, a fully connected nonlinear CNN variation with nine layers including input, multiple convolutions, softmax, batch normalization, and a class output has been used. The first input layer of the CNN takes the image data. This means the size of the input layer has to be the same as that of the input data. The convolution function of the convolution layer comprises of sliding a kernel (or filter) over the input to capture different kinds of local information [35]. The values of the filters are chosen by learners during the training stage [36]. The size of the produced feature map depends on different factors such as the number of filters, size of the filter, stride of the filter process and the padding (i.e., adding additional zero value to fit the filter on bordering elements). The feature map produced by the entire convolution layer can be said to be a map of $n+1$ dimension (where $n$ denotes the input dimension) [37]. The output of the entire convolution layer can be given by

$$C_i = \epsilon(k_i \circledast f + b_i) \tag{1}$$

here, for every $i$ (kernel number in the convolution layer), $C$ is the feature map produced, $\circledast$ denotes the convolution operation, $k$ represents the kernel (or filter), $f$ denotes the input data, whereas bias is represented by $b$. The non linear layer also makes use of the RelU (rectified linear unit) function to induce non linearity into the network. This is accomplished by replacing all negative values with zeros, which increases the classification accuracy [38]. The function is defined as

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \tag{2}$$

here, x represents the input given to the function.

A commonly used method for reducing the dimensionality of the input data is achieved using a pooling layer. In this paper, the max pooling operation has been employed since it is known to give better results. However, an

© 2022 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 July 15,2022 at 09:26:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network Science and Engineering. 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/TNSE.2022.3188881



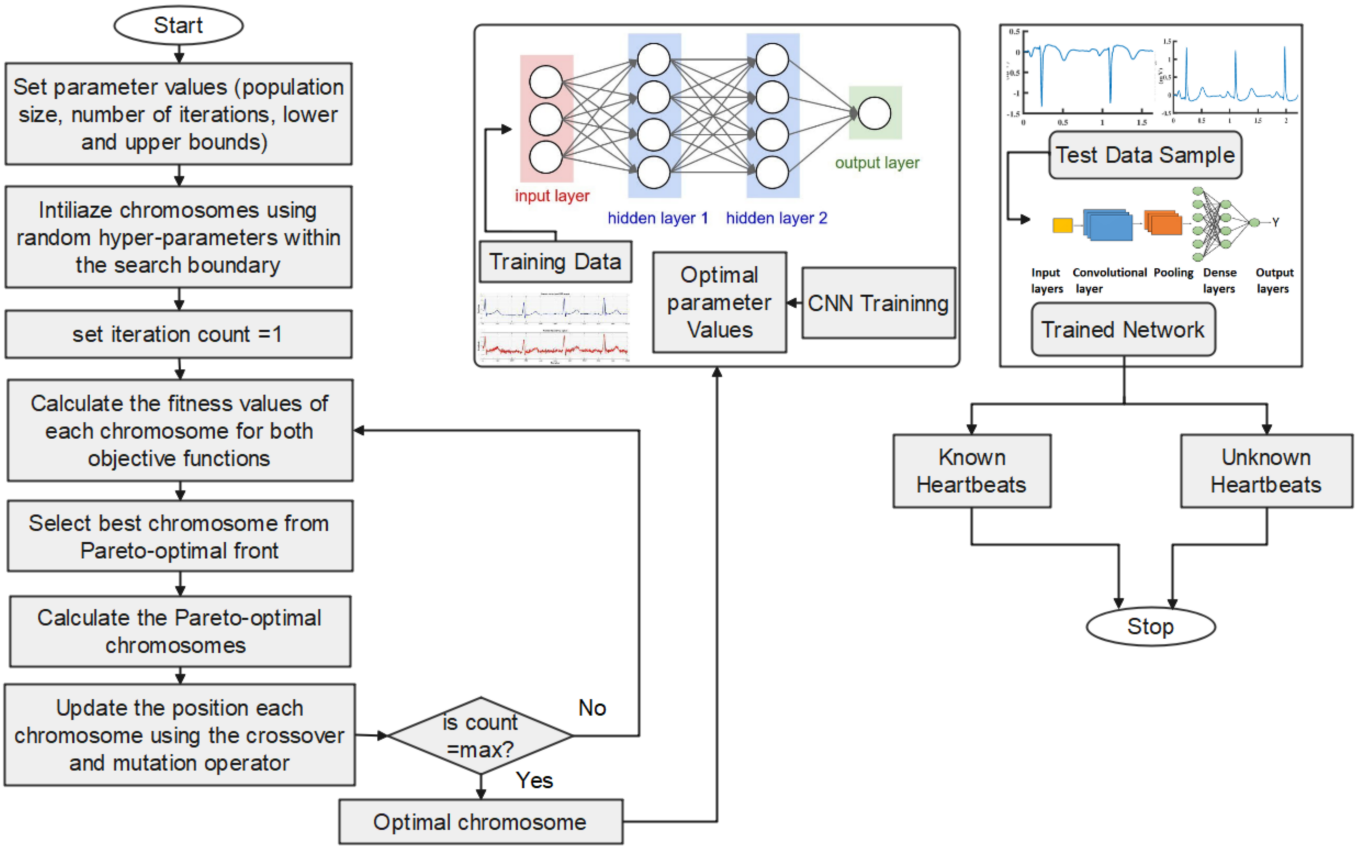



Fig. 2. Flow diagram of N2CNN-based Anomaly Detection Method

average operation can also be used as an alternative [39], [40].

$$f(x) = \max(x) \tag{3}$$

here, x defines the input given to this layer. Further, to reduce the sensitivity of network initialization and enhance the training speed, a batch normalization layer has been added. Then the fully connected layer is meant to help the model learn about the nonlinear parameters. This layer consists of neurons that are connected to every other neuron. The function for each neuron can be given as follows:

$$f(x) = \epsilon\left(\sum_{i}^{n} x_i \times w_i\right) + b \tag{4}$$

Here, $\epsilon$ denotes the nonlinear function, and $i$ is the input number, input is defined by $x$, whereas weight and bias are represented by $w$ and $b$ respectively. The fully connected layer is followed by a normalized exponential function (also known as the SoftMax layer). This layer is used for categorical classification. It creates a probabilistic distribution over $n$ different possible outcomes. The function can be given by Eq. (5).

$$f_{softmax}(x) = \frac{e^x}{\sum_i^n e^{x_i}} \tag{5}$$

Finally, the desired class for the given input is generated by the class out layer.

## B. NSGA-II

Elitist non-dominated sorting genetic algorithm (NSGA-II) [41] is an enhanced and fast version of NSGA algorithm. It is one of the popular multi-objective optimization algorithm. Mathematically, a multi-objective optimization problem are defined as follows:

$$Minimize/Maximize \quad fun_m(X), \tag{6}$$

where $m = 1, 2, \cdots, M$. Whereas:

$$e_j(X) \geq 0, \tag{7}$$

$$ie_k(X) = 0, \tag{8}$$

$$X = (x_1, x_2, \cdots, x_i, \cdots, x_n)^T \tag{9}$$

where, $j = 1, 2, \cdots, J$ denotes inequality constraints, $k = 1, 2, \cdots, andK$ denotes equality constraints, $X^{lower} \leq X_i \leq X^{Upper}$ (search bounds),and $i = 1, 2, \cdots, n$. $X^{lower}$ and $X^{upper}$ are the minimum and maximum dimension values respectively. NSGA-II is applied to the problems formulated as defined above and find the Pareto-optimal solutions which are considered as the best solutions considering all the objective functions. The key features of NSGA-II are as follows:

- Elite preserving property: The best solution of current population is carry forwarded to the next iteration. This may sometimes helps the algorithm to escape from local optima.

© 2022 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 July 15,2022 at 09:26:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network Science and Engineering. 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/TNSE.2022.3188881



- Maintaining the diversity among the Pareto-optimal fronts by the use of crowding distance.
- More focus on the non-dominated solutions in each generation.

The step by step process of NSGA-II is defined as given below:

1) Let $P_t$ and $Q_t$ denotes the child and offspring populations (having $N$ solutions each) and their combination is called $R_t$ which can be defined as $R_t = P_t\ U\ Q_t$ (having $2N$ solutions).
2) Apply the non-domination sorting [41] on $R_t$ and ensure the elite solutions.
3) The best solutions are arranged into different fronts as per their ranks. The best solutions are resides in Front $F_1$.
4) Now, to formulate new population $P_{t+1}$, all solutions resides in $F_1$ are selected. If $F_1 < N$, then the solutions from $F_2$ fronts will be chosen and so on.
5) From selecting the best solution from the last front, crowding distance sorting is applied within that front and less crowded solution will be picked to maintain the diversity in the population.
6) Once $P_{t+1}$ is formulated, all the GA operators can be applied on it.

The above set of steps are applied for the predefined number of iterations or till the sufficient error precision is not achieved. Finally, the resultant solutions are called the Pareto-optimal solutions. Out of these solutions, one solution is selected as per application need.

### C. NSGA-II for CNN hyper-parameter optimization

The performance of CNN can be improved by optimizing its hyper-parameters. In this paper, two hyper-parameters of CNN have been optimized namely, number of layers and filter size. To perform optimization, Elitist Non-dominated sorting genetic algorithm (NSGA-II) [41], a popular multi-objective optimization, has been employed in the proposed method. This benefits over single objective optimization as it is suitable for optimization with multiple objectives [3]. We have considered three objectives for the optimization process, namely accuracy(acc), recall (rec), and precision (prc) which are defined by Eq. (10), (11), and (12) respectively.

$$Accuracy(Acc) = \frac{\sum_{i=1}^{k} \frac{TP_i+TN_i}{TP_i+TN_i+FP_i+FN_i}}{k} \tag{10}$$

$$Recall(rec) = \frac{\sum_{i=1}^{k} TP_i}{\sum_{i=1}^{k}(TP_i + FN_i)} \tag{11}$$

$$Precision(prc) = \frac{\sum_{i=1}^{k} TP_i}{\sum_{i=1}^{k}(TP_i + FP_i)} \tag{12}$$

Where, $i$ represent any class and there are total $k$ classes, $TP_i$ represents the true positive samples or number of hits for any class $i$, $TN_i$, $FP_i$, and $FN_i$ denote the true negative, false positive, false negative samples for respective class $i$.

Let us assume there is a population of $N$ solutions. Each solution consists of $x$ hyper-parameters generated randomly between upper bound ($x_i^U$) and lower bound ($x_i^L$). Let there are $k$ objective functions (in this paper, $k = 3$), then vector $Z$ of objective functions can be represented as Eq. (13).

$$Z(x) = (Z_1(x), Z_2(x), \cdots, Z_k(x)) \tag{13}$$

subject to $j$ constraint functions which are denoted by vector $f$ in Eq. (14).

$$f(x) = (f_1(x), f_2(x), \cdots, f_j(x)) \tag{14}$$

Fig. 2 presents the overall workflow of the proposed method. The inputs to each layer are represented by $in$, and outputs are represented by $out$. Each generation, having a population $P_t$ is composed of chromosomes that are generated using mutation and crossover of hyper-parameters. Once training and testing is completed for the current population, the next population ($P_t + 1$) is created using Pareto fronts and crowding distance in objective space with three objectives $Acc$, $rec$, and $prc$. The following steps can be used to summarize the implemented technique:

Step 1: A population $P_0$ containing $N$ generated randomly chromosomes.

Step 2: Each chromosome represents the hyper-parameters of CNN networks. This CNN has a fixed input and output layer by default regardless of the structure decided by the genetic algorithm (GA). The GA has to select the number of layers between fixed layers, kernel size, layer type, pooling size, stride, and count of neurons of a fully connected layer.

Step 3: After the hyper-parameters have been generated, the weights of the network are trained by ADAM [42] using two-fold approach and applied on the considered dataset. To use as objective parameters, the values of Acc, rec, and pre are averaged.

Step 4: The parent population is sorted by the objective parameters as a non-domination sorting, where $P = P_0$.

Step 5: Further, $N$ new offspring population are created using crossover and mutation [43].

Step 6: Combined population, $R_t$ (equals 2N) consists of $Q_t$ and $P_t$.

Step 7: The combined population is further sorted by fast non-dominated sorting, and solutions are organized into different fronts ($Fr_i$).

Step 8: If $\mid Fr_i \mid \leq N$, then include all solutions of $i^{th}$ front to the next generation else go to step 10.

Step 9: $i = i + 1$ and go to step 8.

Step 10: Define the crowding distance as illustrated by Deb et al. [41].

Step 11: A crowded comparison operator is used in case of population size being greater than $N$. This operator is leveraged in decreasing order to develop the population size of size $N$.

Step 12: Repeat the process from step 5 until termination conditions are fulfilled (i.e., 50 generations).

© 2022 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 July 15,2022 at 09:26:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network Science and Engineering. 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/TNSE.2022.3188881



TABLE I
Distribution of heartbeats among five classes

| | MIT-BIH Dataset | INCART dataset |
|---|---|---|
| Class | Total samples | Total samples |
| N | 90,589 | 1,53,618 |
| S | 2,779 | 1,959 |
| V | 7,236 | 20,041 |
| F | 803 | 219 |
| Q | 8,039 | 7 |
| Number of heartbeats | 1,01,407 | 1,75,844 |

TABLE II
Distribution of heartbeats among five classes

| | MIT-BIH Dataset | INCART dataset |
|---|---|---|
| Class | Total samples | Total samples |
| Known heartbeats | 101407 | 175837 |
| Unknown heartbeats | 8039 | 7 |
| Number of heartbeats | 109446 | 175844 |

## IV. Considered Dataset

1) MIT-BIH arrhythmia database: The MIT-BIH database [44] is a freely available for public access, which has been widely used in comparable studies. Two ECG recording channels recorded at 360 Hz by 47 human male and female volunteers ranging in age from 23 to 89 years are included in this collection. Each set has 22 recordings which is separated into training (DS1) and testing (DS2) parts. Two cardiologists classified each heartbeat into one of 19 categories.
2) INCART dataset: INCART is a popular dataset that is also available for free on the Physionet website [45]. There are 75 annotated recordings in this collection, which have been derived from 32 Holter recordings. Each record lasts 30 minutes and features 12 standard leads. Every lead is sampled at a rate of 257 Hz, with gains ranging from 250 to 1,100 ADC units per millivolt. There are over 175,000 beat annotations in the reference annotation files. Each heartbeat is divided into 19 categories. The class annotations of the dataset have been generated by an automated algorithm.

Initially, ECG heartbeats were classified into 19 groups. However, AAMI reclassified these categories into five primary groups. As per the AAMI standard, ECG heartbeats are divided in 5 classes namely, unknown (Q), atrial premature and supraventricular ectopic (S), fusion of ventricular (F), normal (N), premature ventricular contraction and ventricular ectopic (V). Furthermore, Table I also indicates the count of heartbeats in each of the five classes (N, S, V, F, and Q) for the considered datasets. Out of five heartbeat categories, N, S, V, and F belong to the known heartbeat, while Q belongs to the unknown heartbeat. Therefore, to vindicate the efficacy of the proposed N2CNN based method, heartbeats are grouped into two categories, namely known beats and unknown beats as shown in Table II. The known type comprises heartbeats of N, S, V, and F classes, and the unknown type contains the heartbeat of the Q class. It is clearly visible in the table that the class distribution of the datasets are imbalanced. For a fair analysis, modified synthetic minority oversampling technique (MSMOTE) has been used to enhance the frequency of the minority class. To test the efficacy of the proposed approach, 70% of randomly chosen data samples has been employed for training and the rest data is utilized for testing.

## V. Evaluation Metrics

The most popular parameter for evaluating the performance and efficacy of the ECG classification approach is accuracy. However, the heartbeat categories are very unbalanced; accuracy alone cannot be enough to investigate the efficacy of the proposed N2CNN. Therefore, to assess the efficacy of the proposed approach, four fundamental metrics, namely F1-score, precision, and recall, are also utilized along with accuracy. Further, to make a uniform environment for comparing results, the AAMI standard recommends that the value of F1-score, precision, and recall should be high in successful ECG heartbeat classifications.

## VI. Ablation Study of the Proposed Method

In this paper, ablation study is conducted to analyze the effect of various components on the efficiency of the proposed method. This process is analogy to the brain ablation procedure [46] and has been considered a viable approach to determine the robustness of the deep neural network [47]. To perform this study, MIT-BIH dataset has been considered according to the settings mentioned in Section IV.

Table III depicts the quantitative verification of the optimization of the proposed model with NSGA-II in terms of the considered evaluation metrics. In the table, the proposed model with NSGA-II has performed much better than without it which clearly validates the performance advantages of optimizing the CNN with NSGA-II. Further, the functional verification of the ADAM algorithm in the proposed model (N2CNN) is explored in Table IV. For comparison, gradient descent (GD) algorithm has been considered. It is clearly visible that the use of ADAM algorithm has resulted in efficient performance, especially in terms of accuracy.

TABLE III
Ablation study of N2CNN on MIT-BIH dataset.

| CNN | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| With NSGA-II | 94.70 | 79.72 | 85.42 | 82.26 |
| W/o NSGA-II | 89.02 | 65.24 | 77.58 | 75.36 |

TABLE IV
Ablation study of ADAM algorithm in N2CNN on MIT-BIH dataset.

| N2CNN | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| With ADAM | 94.70 | 79.72 | 85.42 | 82.26 |
| With GD | 91.86 | 76.40 | 81.9 | 77.92 |

© 2022 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 July 15,2022 at 09:26:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network Science and Engineering. 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/TNSE.2022.3188881



TABLE V
Confusion matrices returned by all cosidered methods for MIT-BIH dataset and INCART dataset

| Methods | | MIT-BIH dataset | | INCART dataset | |
|---|---|---|---|---|---|
| | Classes | KH | UH | KH | UH |
| N2CNN | KH | 97654 | 2046 | 166572 | 2 |
| | UH | 3753 | 5993 | 9265 | 5 |
| MOGA-CNN | KH | 95321 | 3103 | 159224 | 4 |
| | UH | 6086 | 4936 | 16613 | 3 |
| MOPSO-CNN | KH | 94897 | 3257 | 158132 | 3 |
| | UH | 6510 | 4782 | 17705 | 4 |
| MOGWO-CNN | KH | 94925 | 3188 | 158673 | 3 |
| | UH | 6482 | 4851 | 17164 | 4 |
| MOWOA-CNN | KH | 94879 | 3136 | 158714 | 4 |
| | UH | 6528 | 4903 | 17123 | 3 |

## VII. Experimental Analysis

1) Confusion Matrix: In this work, four standard multi-objective optimization methods other than proposed NSGA-II based CNN (N2CNN) method, are used in the anomaly detection. These are Multi-objective Genetic Algorithm (MOGA) [48], Multi-objective particle swarm optimization (MOPSO) [49], Multi-objective Grey-wolf Optimization (MOGWO) [50], and Multi-objective whale optimization algorithm (MOWOA) [51]. The confusion matrices are generated to analyze the classification performance of proposed N2CNN methods against other considered methods and depicted in Table V. For MIT-BIH dataset, out of 109446 heartbeat samples, 97654 are correctly identified as known heartbeat (KH). Further, 3753 samples are false positives which are considered as the anomalies. Moreover, for INCART datset, 166572 samples out of 175844 samples are correctly identified as the known heartbeats by the proposed N2CNN method. The values of false positives, true positives, true negatives, and false negatives for all the considered methods are shown in Table V. The false positive cases are considered as detected anomalies. It can be envisioned from the table that the N2CNN based approach shows better results for all the considered performance parameters.

2) Performance Evaluation: The deep learning networks are trained over ECG signals. From the overall dataset, 70% signals are leveraged for training and rest are used for validation. The performance is measured in terms of and F1-measure, precision and recall and the values for these parameters are provide in Table VI and VII for the considered multi-objective optimization methods for MIT-BIH and INCAR datasets respectively. The values of these parameters indicates that the NSGA-II optimized CNN provides the best accuracy on both of the considered datasets. Moreover, Fig. 3 and 4 show the radar charts to validate the efficacy of N2CNN against other methods for BIT-MIH and INCART datasets respectively. The proposed N2CNN method covers the larger area in the radar chart for both the datasets which clearly depicts its efficacy over other considered methods. Therefore, the proposed network can be reliably used to detect the anomaly in real-time applications.

3) Comparative Analysis with state-of-the-art methods: The proposed N2CNN has also been validated against four other state-of-the-art approaches as shown in Table VIII and IX. In Table VIII, results of the N2CNN and other methods over the MIT-BIH dataset are tabulated, while in Table IX, performance metrics of the proposed approach and state-of-the-art methods over INCART dataset has been presented.

From the results of Table VIII and IX, It can be observed that the N2CNN outperforms other state-of-the-art approaches in terms of accuracy, precision, and F1 score. The F1-score is a useful performance metric, especially when dealing with unbalanced data. It can be seen from the above analysis that the proposed N2CNN surpassed the existing state-of-the-art approaches for the considered metrics, including the F1 score, which confirm

TABLE VI
Performance analysis with existing multi-objective optimization methods on MIT-BIH dataset

| Method | Accuracy | Precision | Recall | F1-measure |
|---|---|---|---|---|
| MOGA-CNN | 91.60 | 70.82 | 77.70 | 73.60 |
| MOPSO-CNN | 91.08 | 69.52 | 76.53 | 72.29 |
| MOGWO-CNN | 91.16 | 69.78 | 76.98 | 72.62 |
| MOWOA-CNN | 91.17 | 69.85 | 77.28 | 72.76 |
| N2CNN | 94.70 | 79.72 | 85.42 | 82.26 |

TABLE VII
Performance analysis with existing multi-objective optimization methods on INCART dataset

| Method | Accuracy | Precision | Recall | F1-measure |
|---|---|---|---|---|
| MOGA-CNN | 90.55 | 50.01 | 66.70 | 47.54 |
| MOPSO-CNN | 89.93 | 50.01 | 73.54 | 47.37 |
| MOGWO-CNN | 90.24 | 50.01 | 73.69 | 47.46 |
| MOWOA-CNN | 90.26 | 50.01 | 66.56 | 47.46 |
| N2CNN | 94.73 | 50.03 | 83.08 | 48.70 |

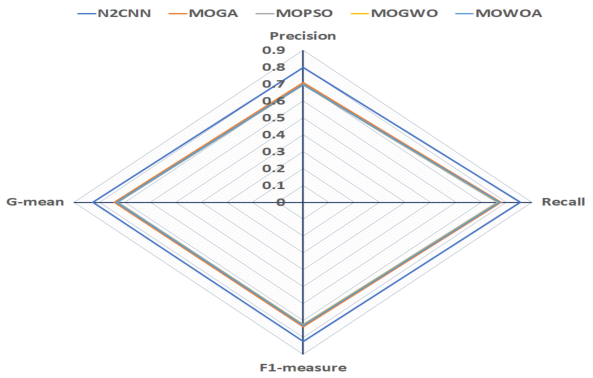



Fig. 3. Radar Chart for MIT-BIH dataset

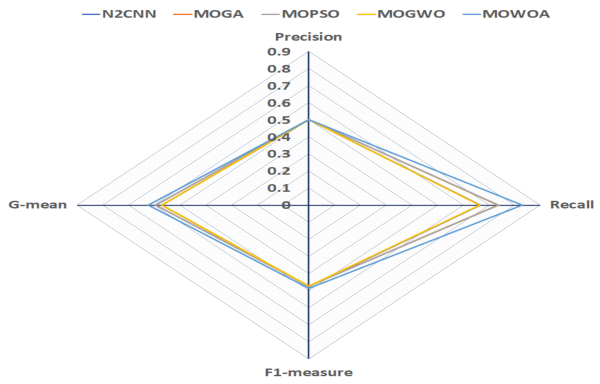



Fig. 4. Radar Chart for INCART dataset

© 2022 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 July 15,2022 at 09:26:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network Science and Engineering. 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/TNSE.2022.3188881



TABLE VIII
Comparison of the proposed approach with state-of-the-art techniques over MIT-BIH dataset

| Authors | Methods | Accuracy (%) | F1 score (%) | Precision (%) | Recall (%) |
|---|---|---|---|---|---|
| Martis et al. [52], 2014 | DWT and SVM | 94.52 | – – | – – | – – |
| Li et al. [53], 2016 | WPE and random forest | 94.61 | – – | – – | – – |
| Garcia et al. [8], 2017 | SVM and swarm optimization (PSO) | 92.40 | – – | – – | 81.10 |
| M. Kachuee et al. [20], 2018 | Deep residual CNN | 93.40 | – – | – – | – – |
| Ramirez et al. [54], 2019 | Hybrid neural network | 93.80 | – – | – – | – – |
| Marinho et al. [55], 2019 | NB, SVM, MLP, and OPF | 94.30 | – – | – – | – – |
| Bidias et al. [56], 2020 | Conditional entropy of ordinal patterns | 93.62 | – – | – – | – – |
| Proposed approach, 2021 | N2CNN | 94.70 | 94.77 | 94.70 | 94.70 |

TABLE IX
Comparison of the proposed approach with state-of-the-art techniques over INCART dataset

| Authors | Methods | Accuracy (%) | F1 score (%) | Precision (%) | Recall (%) |
|---|---|---|---|---|---|
| Chazul et al. [57], 2004 | Heartbeat Interval Features | 84.70 | 77.05 | 77.05 | 77.05 |
| Zhang et al [58], 2014 | Disease-specific feature and SVM | 86.24 | 79.36 | 79.36 | 79.36 |
| Huang et al. [10], 2014 | Random projections and RR intervals | 87.43 | 81.14 | 81.14 | 81.14 |
| Chu et al. [59], 2019 | CNN and LSTM | 88.56 | 81.17 | 82.80 | 82.84 |
| Proposed Method, 2021 | N2CNN | 94.73 | 94.74 | 94.73 | 94.73 |

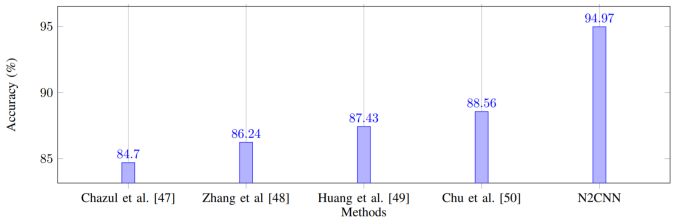



Fig. 5. Comparison of accuracy of the proposed approach and other state-of-the-arts over INCART dataset

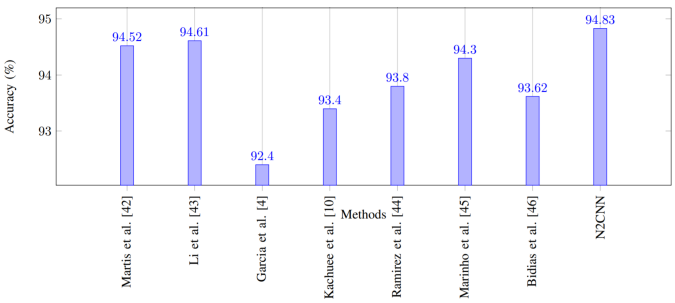



Fig. 6. Comparison of accuracy of the proposed approach and other state-of-the-arts over MIT-BIH dataset

the efficacy of the proposed N2CNN based approach. Furthermore, for better visualization, the accuracy of the proposed and seven other considered methods for both datasets is also depicted in Fig. 5 and 6, respectively. It is visible from the figure that the N2CNN has clearly surpassed the other methods on both datasets. Thus, it can be established from the experimental results that the proposed N2CNN based approach is superior to seven other state-of-the-art methods in terms of accuracy.

## VIII. Conclusion

This paper introduces a new NSGA-II based convolution neural network (N2CNN) for anomaly monitoring in remote patient. The proposed method utilizes the strength of multi-objective NSGA-II for hyperparameters optimization of CNN layers. The classification accuracy of the proposed method is tested on two public datasets, namely, INCART, and MIT-BIH. The experimental results witnessed that the proposed N2CNN surpassed the seven other state-of-the-art methods on both datasets. The proposed method has achieved 94.70% accuracy in the MIT-BIH dataset, whereas 94.73% accuracy in INCART dataset, whereas, runner method has given 91.60% and 90.55% accuracy on both the datasets respectively. Thus, it can be elucidated that the proposed N2CNN model can be efficiently utilized to monitor patients' health conditions in a real-time environment. The employment of NSGA-II on CNN has optimized the network in terms of number of layers and filter size. This has resulted in better performance of proposed N2CNN. Moreover, the proposed N2CNN method can detect the any tempering or anomaly in the cardiovascular data of the patients. Thus, the security of the remote patient data can be efficiently handled used the proposed N2CNN based method. In the future, the proposed method can be employed on massive real-time datasets using Big-data methods. Moreover, new wearable sensors may be used for creating data to ensure the health of patients.

## References

[1] F. J. Dian, R. Vahidnia, and A. Rahmati, "Wearables and the internet of things (iot), applications, opportunities, and challenges: A survey," IEEE Access, vol. 8, pp. 69 200–69 211, 2020.

[2] H. Mittal and M. Saraswat, "An optimum multi-level image thresholding segmentation using non-local means 2d histogram and exponential kbest gravitational search algorithm," Engineering applications of artificial intelligence, vol. 71, pp. 226–235, 2018.

[3] S. S. Mostafa, N. Horta, A. G. Ravelo-Garcia, and F. Morgado-Dias, "Analog active filter design using a multi objective genetic algorithm," AEU-International Journal of Electronics and Communications, vol. 93, pp. 83–94, 2018.

[4] E. J. d. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "Ecg-based heartbeat classification for arrhythmia detection: A survey," Computer methods and programs in biomedicine, vol. 127, pp. 144–164, 2016.

[5] C. M. Bishop and N. M. Nasrabadi, Pattern recognition and machine learning. Springer, 2006, vol. 4, no. 4.

[6] D. Bhamare and P. Suryawanshi, "Review on reliable pattern recognition with machine learning techniques," Fuzzy Information and Engineering, vol. 10, no. 3, pp. 362–377, 2018.

[7] K. Park, B. Cho, D. Lee, S. Song, J. Lee, Y. Chee, I. Kim, and S. Kim, "Hierarchical support vector machine based heartbeat classification using higher order statistics and hermite basis

© 2022 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 July 15,2022 at 09:26:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network Science and Engineering. 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/TNSE.2022.3188881



function," in 2008 Computers in Cardiology. IEEE, 2008, pp. 229–232.
[8] G. Garcia, G. Moreira, D. Menotti, and E. Luz, "Inter-patient ecg heartbeat classification with temporal vcg optimized by pso," Scientific reports, vol. 7, no. 1, pp. 1–11, 2017.
[9] G. d. Lannoy, D. François, J. Delbeke, and M. Verleysen, "Weighted svms and feature relevance assessment in supervised heart beat classification," in International joint conference on biomedical engineering systems and technologies. Springer, 2010, pp. 212–223.
[10] H. Huang, J. Liu, Q. Zhu, R. Wang, and G. Hu, "A new hierarchical method for inter-patient heartbeat classification using random projections and rr intervals," Biomedical engineering online, vol. 13, no. 1, pp. 1–26, 2014.
[11] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," Intelligent data analysis, vol. 6, no. 5, pp. 429–449, 2002.
[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," Journal of artificial intelligence research, vol. 16, pp. 321–357, 2002.
[13] H. Zhang, L. Huang, C. Q. Wu, and Z. Li, "An effective convolutional neural network based on smote and gaussian mixture model for intrusion detection in imbalanced dataset," Computer Networks, vol. 177, p. 107315, 2020.
[14] G. Sannino and G. De Pietro, "A deep learning approach for ecg-based heartbeat classification for arrhythmia detection," Future Generation Computer Systems, vol. 86, pp. 446–455, 2018.
[15] S. K. Pandey and R. R. Janghel, "Classification of electrocardiogram signal using an ensemble of deep learning models," Data Technologies and Applications, 2021.
[16] M. Hammad, A. Maher, K. Wang, F. Jiang, and M. Amrani, "Detection of abnormal heart conditions based on characteristics of ecg signals," Measurement, vol. 125, pp. 634–644, 2018.
[17] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. San Tan, "A deep convolutional neural network model to classify heartbeats," Computers in biology and medicine, vol. 89, pp. 389–396, 2017.
[18] M. Amrani, M. Hammad, F. Jiang, K. Wang, and A. Amrani, "Very deep feature extraction and fusion for arrhythmias detection," Neural Computing and Applications, vol. 30, no. 7, pp. 2047–2057, 2018.
[19] S. Mousavi and F. Afghah, "Inter-and intra-patient ecg heartbeat classification for arrhythmia detection: a sequence to sequence deep learning approach," in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019, pp. 1308–1312.
[20] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "Ecg heartbeat classification: A deep transferable representation," in IEEE International Conference on Healthcare Informatics (ICHI). IEEE, 2018, pp. 443–444.
[21] H. Wang, H. Shi, K. Lin, C. Qin, L. Zhao, Y. Huang, and C. Liu, "A high-precision arrhythmia classification method based on dual fully connected neural network," Biomedical Signal Processing and Control, vol. 58, p. 101874, 2020.
[22] Z. Ebrahimi, M. Loni, M. Daneshtalab, and A. Gharehbaghi, "A review on deep learning methods for ecg arrhythmia classification," Expert Systems with Applications, vol. 7, p. 100033, 2020.
[23] T. Goel, R. Murugan, S. Mirjalili, and D. K. Chakrabartty, "Optconet: an optimized convolutional neural network for an automatic diagnosis of covid-19," Applied Intelligence, vol. 51, no. 3, pp. 1351–1366, 2021.
[24] N. Zhang, Y.-X. Cai, Y.-Y. Wang, Y.-T. Tian, X.-L. Wang, and B. Badami, "Skin cancer diagnosis based on optimized convolutional neural network," Artificial intelligence in medicine, vol. 102, p. 101756, 2020.
[25] Y. Wang, H. Zhang, and G. Zhang, "cpso-cnn: An efficient pso-based algorithm for fine-tuning hyper-parameters of convolutional neural networks," Swarm and Evolutionary Computation, vol. 49, pp. 114–123, 2019.
[26] B. Navaneeth and M. Suchetha, "Pso optimized 1-d cnn-svm architecture for real-time detection and classification applications," Computers in biology and medicine, vol. 108, pp. 85–92, 2019.
[27] A. Azad, E. Aghaei, A. Jalali, and P. Ahmadi, "Multi-objective optimization of a solar chimney for power generation and water desalination using neural network," Energy Conversion and Management, vol. 238, p. 114152, 2021.
[28] J. Xie, P. Zhao, P. Hu, Y. Yin, H. Zhou, J. Chen, and J. Yang, "Multi-objective feed rate optimization of three-axis rough milling based on artificial neural network," The International Journal of Advanced Manufacturing Technology, vol. 114, no. 5, pp. 1323–1339, 2021.
[29] A. Bakar, K. Li, H. Liu, Z. Xu, M. Alessandrini, and D. Wen, "Multi-objective optimization of low reynolds number airfoil using convolutional neural network and non-dominated sorting genetic algorithm," Aerospace, vol. 9, no. 1, p. 35, 2022.
[30] T. Goel, R. Murugan, S. Mirjalili, and D. K. Chakrabartty, "Multi-covid-net: Multi-objective optimized network for covid-19 diagnosis from chest x-ray images," Applied Soft Computing, vol. 115, p. 108250, 2022.
[31] S. Kanwal, I. Younas, and M. Bashir, "Evolving convolutional autoencoders using multi-objective particle swarm optimization," Computers & Electrical Engineering, vol. 91, p. 107108, 2021.
[32] D. Singh, V. Kumar, M. Kaur et al., "Classification of covid-19 patients from chest ct images using multi-objective differential evolution–based convolutional neural networks," European Journal of Clinical Microbiology & Infectious Diseases, vol. 39, no. 7, pp. 1379–1389, 2020.
[33] E.-G. Talbi, "Automated design of deep neural networks: A survey and unified taxonomy," ACM Computing Surveys (CSUR), vol. 54, no. 2, pp. 1–37, 2021.
[34] M. Feurer and F. Hutter, "Hyperparameter optimization," in Automated machine learning. Springer, Cham, 2019, pp. 3–33.
[35] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press, 2016.
[36] J. Ren and L. Xu, "On vectorization of deep convolutional neural networks for vision tasks," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 29, no. 1, 2015.
[37] D. Stutz, "Understanding convolutional neural networks," 2014.
[38] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in Icml, 2010.
[39] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in International conference on artificial neural networks. Springer, 2010, pp. 92–101.
[40] H. Mittal and M. Saraswat, "A new fuzzy cluster validity index for hyperellipsoid or hyperspherical shape close clusters with distant centroids," IEEE Transactions on Fuzzy Systems, vol. 29, no. 11, pp. 3249–3258, 2020.
[41] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," IEEE transactions on evolutionary computation, vol. 6, no. 2, pp. 182–197, 2002.
[42] D. Kingma and J. Ba, "Dp kingma and j. ba, adam: A method for stochastic optimization," arXiv preprint arxiv:1412.6980.
[43] K. Deb, "Multi-objective optimisation using evolutionary algorithms: an introduction," in Multi-objective evolutionary optimisation for product design and manufacturing. Springer, 2011, pp. 3–34.
[44] G. B. Moody and R. G. Mark, "The mit-bih arrhythmia database on cd-rom and software for use with it," in Proceedings Computers in Cardiology. IEEE, 1990, pp. 185–188.
[45] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," circulation, vol. 101, no. 23, pp. e215–e220, 2000.
[46] J. Wolf, "Speech recognition: Invited papers presented at the 1974 ieee symposium," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 25, no. 2, pp. 207–207, 1977.
[47] R. Meyes, M. Lu, C. W. de Puiseau, and T. Meisen, "Ablation studies in artificial neural networks," arXiv preprint arXiv:1901.08644, 2019.
[48] R. Pal, A. Malik, S. Yadav, and R. Karnwal, "Optimal cluster head election in industrial wsns using the multi-objective genetic algorithm," in Applications of Advanced Optimization Techniques in Industrial Engineering. CRC Press, 2022, pp. 169–180.
[49] C. C. Coello and M. S. Lechuga, "Mopso: A proposal for multiple objective particle swarm optimization," in Proceedings of the

© 2022 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 July 15,2022 at 09:26:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Network Science and Engineering. 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/TNSE.2022.3188881



2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600), vol. 2. IEEE, 2002, pp. 1051–1056.
[50] S. Mirjalili, S. Saremi, S. M. Mirjalili, and L. d. S. Coelho, "Multi-objective grey wolf optimizer: a novel algorithm for multi-criterion optimization," Expert Systems with Applications, vol. 47, pp. 106–119, 2016.
[51] J. Wang, P. Du, T. Niu, and W. Yang, "A novel hybrid system based on a new proposed algorithmâ€"multi-objective whale optimization algorithm for wind speed forecasting," Applied energy, vol. 208, pp. 344–360, 2017.
[52] R. J. Martis, U. R. Acharya, C. M. Lim, K. Mandana, A. K. Ray, and C. Chakraborty, "Application of higher order cumulant features for cardiac health diagnosis using ecg signals," International journal of neural systems, vol. 23, no. 04, p. 1350014, 2013.
[53] T. Li and M. Zhou, "Ecg classification using wavelet packet entropy and random forests," Entropy, vol. 18, no. 8, p. 285, 2016.
[54] E. Ramirez, P. Melin, and G. Prado-Arechiga, "Hybrid model based on neural networks, type-1 and type-2 fuzzy systems for 2-lead cardiac arrhythmia classification," Expert Systems with Applications, vol. 126, pp. 295–307, 2019.
[55] L. B. Marinho, N. de MM Nascimento, J. W. M. Souza, M. V. Gurgel, P. P. Reboucas Filho, and V. H. C. de Albuquerque, "A novel electrocardiogram feature extraction approach for cardiac arrhythmia classification," Future Generation Computer Systems, vol. 97, pp. 564–577, 2019.
[56] J. B. B. à Mougoufan, J. A. E. Fouda, M. Tchuente, and W. Koepf, "Adaptive ecg beat classification by ordinal pattern based entropies," Communications in Nonlinear Science and Numerical Simulation, vol. 84, p. 105156, 2020.
[57] P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ecg morphology and heartbeat interval features," IEEE transactions on biomedical engineering, vol. 51, no. 7, pp. 1196–1206, 2004.
[58] Z. Zhang, J. Dong, X. Luo, K.-S. Choi, and X. Wu, "Heartbeat classification using disease-specific feature selection," Computers in biology and medicine, vol. 46, pp. 79–89, 2014.
[59] J. Chu, H. Wang, and W. Lu, "A novel two-lead arrhythmia classification system based on cnn and lstm," Journal of Mechanics in Medicine and Biology, vol. 19, no. 03, p. 1950004, 2019.

Dr. Raju Pal is an Assistant Professor at Jaypee Institute of Information Technology, Noida, India. Dr. Pal has obtained his Ph.D. in Computer Science & Engineering from JIIT, Noida, India. He has more than 8 years of teaching and research experience. The keen research areas of Dr. Pal are medical image analysis, meta-heuristic algorithms, and wireless sensor networks. Also, he has been the co-principal investigator of the research project on histopathological image analysis funded by the SERB, Department of Science and Technology, India. He served as a program committee member of various international conferences and reviewer of many international journals of good repute.

Dr. Ashish Kumar Tripathi (Member, IEEE) received M.Tech., and Ph.D. degrees from the Department of Computer Science and Engineering, Delhi Technological University, Delhi, India, in 2013, and 2019 respectively. He is currently working as an Assistant Professor with the Department of Computer Science & Engineering, Malaviya National Institute of Technology (MNIT), Jaipur, India. His research interests include big data analytics, social media analytics, Internet of Things, Vedio and Image data processing.

Dr. Avinash Chandra Pandey (Member, IEEE) is currently working as Assistant Professor in the department of Computer Science & Engineering of Indian Institute of Information Technology Design and Manufacturing, Jabalpur. He has more than 8 years of teaching and research experience. He has guided many M. Tech. dissertations and B. Tech. projects. He has published more than 20 journal and conference papers in the area of text mining, data mining, and soft computing. His research areas include data analytics, text mining, and soft computing.

Dr Mohammad Ayoub khan (Senior Member, IEEE) received a Ph. D (Computer Engineering) from Jamia Millia Islamia, New Delhi, India and a Master of Technology (Computer Science and Engineering) from Guru Gobind Singh Indraprastha, New Delhi, India. Presently, he is an Associate Professor at the University of Bisha, Saudi Arabia with interests in the Internet of Things, RFID, wireless sensor networks, adhoc networks, smart cities, industrial IoT, signal processing, NFC, routing in network-on-chip, real time and embedded systems. He has more than 14 years of experience in his research areas. He has published many research papers and books in reputed journals and international IEEE conferences. He contributes to the research community by undertaking various volunteer activities in the capacity of editor for many journals and as a conference chair.

Dr. Varun G. Menon is currently Professor and Head of the Department of Computer Science Engineering at SCMS School of Engineering and Technology, India. He is a Distinguished Speaker of the Association of Computing Machinery (ACM) and a Senior Member of IEEE. He has completed his Ph.D. in Computer Science and Engineering and holds an M. Tech degree in Computer and Communication with University First Rank. He also holds an M.Sc. in Applied Psychology, an MBA in Human Resource Management and a Diploma in Training and Development. His research interests include Sensor Technologies, Internet of Things, Green IoT, Wireless Communication, Fog Computing and Networking.

Dr. Himanshu Mittal is an Assistant Professor in the Computer Science Department of Jaypee Institute of Information Technology (JIIT) , India. He received his Ph.D. in the field of computer vision under the supervision of Dr. Mukesh Saraswat. The keen research areas of Dr. Mittal are image analysis, machine learning, and evolutionary algorithms. He has an excellent academic record as well as research background with the papers in reputed journals including IEEE Transactions. Also, he has been the co-principal investigator of the research project funded by the Science and Engineering Research Board, Department of Science and Technology, India.

© 2022 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 July 15,2022 at 09:26:13 UTC from IEEE Xplore. Restrictions apply.

Ricerche di Matematica
https://doi.org/10.1007/s11587-022-00720-6

# A nonlinear SAIR epidemic model: Effect of awareness class, nonlinear incidences, saturated treatment and time delay

**Kanica Goel**[1] · **Nilam**[1]

Received: 7 January 2022 / Revised: 28 May 2022 / Accepted: 20 June 2022
© Università degli Studi di Napoli "Federico II" 2022

**Abstract**
Awareness plays a vital role in informing and educating people about infection risk during an outbreak and hence helps to reduce the epidemic's health burden by lowering the peak incidence. Therefore, this paper studies a susceptible-aware-infected-recovered (SAIR) epidemic model with the novel combinations of Michaelis-Menten functional type nonlinear incidence rates for unaware and aware susceptible with the inclusion of time delay as a latent period and a saturated treatment rate for infected people. The model is analyzed mathematically to describe disease transmission dynamics in two obtained equilibria: disease-free and endemic. We derive the basic reproduction number $R_0$ and investigate the local and global stability behavior of obtained equilibria for the time delay $\varpi \geq 0$. A bifurcation analysis is performed using center manifold theory when there is no time delay, revealing the forward bifurcation when $R_0$ varies from unity. Moreover, the presence of Hopf bifurcation around EE is shown depending on the bifurcation parameter time delay. Lastly, the numerical simulations validate the analytical findings.

**Keywords** Epidemic model with awareness · Time-delay · Michaelis-Menten incidence rates · Saturated treatment rate · Stability · Bifurcations

**Mathematics Subject Classification** 34D20 · 92B05 · 37M05

✉ Nilam
rathi.nilam@gmail.com; nilam.research18@gmail.com

Kanica Goel
kanica.dtu2016@gmail.com

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

Published online: 11 July 2022

## 1 Introduction

Epidemics remain a significant challenge for human survival. Every year, more than 8.7 million people die due to infectious diseases such as diarrhea, measles, dengue, etc. These diseases are the reason for high mortality and morbidity around the world every year. The compartmental modeling of infectious diseases is valuable in enhancing the understanding of the mechanisms of disease transmission. By analyzing the model, we can foresee future outbreaks up to a large extent which helps in evaluating control methodologies. A compartmental model describes the disease transmission mechanism by dividing the total population into distinct subclasses according to their epidemiological status. In 1927, Kermack and McKendrick [28] presented a deterministic epidemic model by considering three epidemiological compartments (classes): the susceptible compartment ($S(t)$), which measures the type of those individuals who can catch the disease and become infectives, the infective compartment ($I(t)$) that measures those individuals who are infected and can spread it to others, and the removed compartment ($R(t)$), that defines the class of those individuals who have recovered from the disease. Many epidemic models have been formulated with different compartments, such as SIS, SIR, SIRS, SEIS, SEIR, SVIRS, SFIR, etc. [8, 11, 13, 17, 20, 25, 32, 45, 50, 51].

A significant factor in the dynamic study of infectious disease is the incidence rate by which infection transmits to susceptible individuals. The incidence rate measures the frequency of occurrence of new infection per unit time. In 1927, Kermack Mckendrick assumed the bilinear incidence rate of the form $g(I)S = kIS$, where $g(I) = kI$ is unbounded for $I \geq 0$. This incidence rate is real for a small population of infected individuals, but it is unfeasible for a high density of infected populations. Therefore, numerous authors are keen to deliberate nonlinear incidence rates to study the transmission dynamics of infectious diseases (e.g., [2, 6, 19, 20, 29, 31, 36, 48, 51]). In the present article, we consider the saturating Michaelis-Menten functional response type incidence rate in which the number of adequate contacts per infective in unit time grows less rapidly as the total population increases. Michaelis-Menten contact rate is of the form $g(I) = \beta\sigma I/(1 + I)$. It combines the bilinear and standard incidence rates methods, by assuming that if the number of infectives $I$ is suitably low, the number of actual per capita infectives $g(I)$ is proportional to $I$, whereas, in the case of a high density of infected individuals $I$, there is a saturation effect which makes the number of actual infectives constant [3, 4]. Consideration of time delay in the epidemic model emphasizes that a person may not be infectious until some time after becoming infected, which has a significant role in determining the number of infected individuals during an outbreak. Also, the inclusion of time delay in the epidemic models can induce oscillations and periodic solutions as the delay increases. Therefore, many authors focused on considering the time delay in the epidemic [24, 25, 41, 50]. Motivated by the work [3–5], in the present article, we incorporate the incidence rate of Michaelis-Menten type functional response for both susceptible and aware classes individuals with time delay representing latent period.

Affordable and safe medical treatments are also necessary, which reduce and prevent the increase in the number of infected individuals. Treatment is a powerful method that fights against the infection by stopping it from reproducing or killing bacteria.

Wang and Ruan considered a constant treatment rate in the SIR epidemic model [47]. However, a constant treatment rate is appropriate when humans have rich and better treatment resources. Therapeutic amenities and subsequent treatments might need some time to develop and implement; therefore, the choice of a reasonable treatment rate is crucial. Due to the limited medical resources, providing treatment to all the infectives puts a tremendous burden on public health associations. Hence, the saturating treatment rate, which tends to a finite limit as the number of infected individuals increases, makes the epidemic model more realistic. Therefore, Wang [46] considered a SIR epidemic model with a piecewise function type treatment rate given by

$$h(I) = \begin{cases} rI, & 0 < I \leq I_0 \\ rI_0, & I > I_0 \end{cases}$$

In this form, the treatment rate and the number of infected individuals are proportional if the maximal treatment capacity is not reached. Later, Dubey et al. deliberated different nonlinear treatment rates such as Holling Type II, Holling Type III, and Holling Type IV and studied the disease dynamics [11, 13, 14]. In the present study, for a more realistic situation of treatment availability of infected individuals, we employ a saturated functional type treatment rate given by

$$h(I) = \frac{aI^2}{bI^2 + cI + 1},$$

where, $a$ is the maximum cure rate of infected individuals, $b$ is the rate of limitation in treatment availability and $c$ is the saturation constant in the absence of inhibitory effect.

Awareness about the spread of a disease is a valuable ally in affecting susceptibles' behavior and mitigating further infection. Awareness leads to sharing necessary information about the condition to the general population, getting thought, making the individuals familiar with the disorder, and providing the most substantial protection against infectious diseases. During the 2003 SARS outbreak, the Chinese Southern Weekend newspaper spread the instant message, "There is a fatal flu in Guangzhou" 126 million times in Guangzhou alone, which affected people's behavior to take necessary preventive measures [42]. This figure remains a distinct difference from the nearly low number of 5,327 cases recorded in the entire China [49]. If no centralized information is available, people can be aware of infection risk through word of mouth, personal communications, and social media like Twitter, Facebook, and other online tools by which people search out and analyze prescriptions to obtain methods for fast healing. The disease spread can also be controlled by vaccination, but immunization is costly, and sometimes vaccinations are temporary, and it is difficult to vaccine all the individuals due to various limitations. Even some fatal diseases like AIDS, Malaria, Chikungunya, Plague, and Dengue have no vaccination; only a person's awareness can prevent the spread of these diseases efficiently and effectively. For instance, the habit of mosquito nets and mosquito coils helps in preventing Dengue and Chikungunya [15, 40]. Many authors have studied the concept of awareness in their epidemic models [1, 8, 9, 37, 38]. Funk et al., Misra et al., and Dubey et al. [12, 16, 39] deliberate the

influence and significance of awareness plans on the transmission and control of the outbreak via nonlinear mathematical models. Kumar et al. [30] incorporated the alert individuals class into the SIR epidemic model and studied the effect of alertness in infectious disease transmission dynamics. Goel et al. [18] also deliberate the influence of the full and partial awareness about the diseases spread among susceptibles with Holling type II incidences and treatment rates. A multiplicity of studies involving homogeneous spread qualitative models have been developed to provide a better understanding of the complex potential provided by awareness for the containment of epidemics. For example, in 2018, Just et al. [27] introduce awareness in the epidemic model through a specific class of aware susceptibles and investigated a reactive SAIS model in which awareness is imparted via a special class of aware susceptibles. They demonstrated that in these models, even in the presence of awareness decay, a constant fraction of aware individuals may be maintained, leading to an elevation of the epidemic threshold above the basic reproduction number. In Lacitignola et al. [34] model, awareness is introduced through a system variable ruled by an assigned evolution equation. Das et al. [10] investigate the influence of social awareness spread by media in TB transmission dynamics and gave the optimal strategy for the prevalence of tuberculosis. Lacitignola et al. [33] used the Z-control approach to gain insight into the role of awareness in the management of epidemics through the SEIR epidemic model, where they considered awareness as a time-dependent variable whose dynamics are not assigned a priori and used it as an indirect control on the class of infective individuals, exploiting its potential to produce social-distancing and self-isolation among susceptibles.

Human awareness during an outbreak is a critical factor that profoundly influences the transmission pattern of infectious diseases. Therefore, in this study, we present an infectious disease transmission compartmental model comprising four subpopulations: fully susceptible, aware susceptible population, infected, and recovered subpopulations, and formulate a mathematical time-delayed epidemic model that incorporates two explicit nonlinear incidences with a latent period and a nonlinear treatment rate for the infected individuals. We have considered the Michaelis-Menten type nonlinear incidence rate that prevents the unboundedness of the infected people and a saturating treatment rate of infected people, including the limited accessibility of treatment resources. At the beginning of the infectious disease, there is a time known as the latency period, before an infected person can transmit the infection to another person. Therefore, the inclusion of the latent period in the incidence pattern makes the present model more realistic. After formulating the model, we perform a mathematical analysis that allows long-term qualitative predictions of outbreaks and the persistence of the disease. We derive the basic reproduction number and estimate how an infection can spread in a population. Using $R_0$, the local and global stability behavior of disease-free and endemic equilibria is investigated, revealing the persistence or eradication of infection. The global stability behavior of both the equilibria has been proved by using Lyapunov functionals and the Lyapunov-LaSalle invariance approach. Moreover, the oscillatory and periodic solutions near-endemic equilibrium has been seen via Hopf bifurcation by considering time delay as a bifurcation parameter. The numerical experiments show the significance of the model's variables and parameters and suggest strategies that could prevent infection.

The remainder manuscript is ordered as follows: In Sect. 2, a novel time-delayed SAIR epidemic model is presented. In Sect. 3, the basic properties of the model are presented. In Sect. 4, the local stability behavior of the disease-free equilibrium and the endemic equilibrium is investigated for the time-delay $\varpi \geq 0$ by driving the basic reproduction number $R_0$. The forward bifurcation for the undelayed system is shown when the $R_0$ varies from unity. Further, the appearance of the periodic solutions is shown near-endemic equilibrium through Hopf bifurcation. Furthermore, the global stability of both the equilibria is investigated by employing Lyapunov functionals' direct Lyapunov method. In Sect. 5, numerical simulations are presented to show the significance of the analytical findings. To conclude, a discussion of the present study is given in Sect. 6.

## 2 Model derivation

Let $P$ denotes the total population and the transmission of infectious diseases involves four types of subpopulations: Susceptible individuals $X_p(t)$, Aware individuals $A_p(t)$, Infected individuals $I_p(t)$, and Recovered individuals $R_p(t)$. That is, $P = X_p(t) + A_p(t) + I_p(t) + R_p(t)$, which means that the individuals categorized in $X_p(t)$, $A_p(t)$, $I_p(t)$ and $R_p(t)$ may vary with time $t$ and $P$ is a fixed population. It is assumed that each subpopulation of the SAIR model is well mixed and interact homogeneously with each other [20].

Let $\kappa$ denotes the constant recruitment rate of susceptibles. $\xi$ is the awareness rate in susceptible individuals, and thus the term $\xi X_p$ enters the class $A_p(t)$. We consider Michaelis-Menten type two explicit nonlinear incidence rates with the following interpretation: the term $\Psi(X_p(t-\varpi), I_p(t-\varpi)) = \frac{\beta\sigma X_p(t-\varpi)I_p(t-\varpi)}{1+I_p(t-\varpi)}$ represents the incidence rate when susceptible individuals catch the infection from infected individuals, and the term $\Lambda(A_p(t-\varpi), I_p(t-\varpi)) = \frac{\gamma\sigma A_p(t-\varpi)I_p(t-\varpi)}{1+I_p(t-\varpi)}$ represents the incidence rate when aware individuals catch the infection from infectives. Here, $\beta$ and $\gamma$ denote the force of infection among susceptible and aware individuals, respectively, and $\sigma$ denotes the average number of contact partners. We assume that $\gamma < \beta$, as the aware individuals are at a lower risk of getting infected than fully susceptible individuals. The parameter $\varpi$ is the time delay which represents the latent phase having a fixed duration. The period from the time of infection to becoming infectious is called the latent period. During the latent period, a host may or may not show symptoms, but the host cannot infect other hosts in both cases. Thus, the latent period significantly influences the spreading dynamics of an infectious disease or epidemic. Since the aware individuals can also become infected, perhaps at a lower rate than fully susceptibles, they also have some behavioral responses and have a latency phase due to immunological reasons. Thus, the time delay $\varpi$ is the constant latency time and represents the time taken by the fully susceptible and aware individuals, that, infected at a time $t$ can infect other susceptible and aware individuals at time $t + \varpi$ only. The parameters $\vartheta$, $d$, and $\theta$ represent the natural death rate, disease-induced death rate, and the recovery rate of infected individuals, respectively. The nonlinear term $h(I_p(t)) = \frac{aI_p^2}{bI_p^2+cI_p+1}$

**Table 1** Symbolizations of variables and parameters

| Notation | Description |
|---|---|
| $P$ | Total constant population |
| $X_p(t)$ | Susceptible subpopulation (full vulnerable) |
| $A_p(t)$ | Aware subpopulation |
| $I_p(t)$ | Infected subpopulation |
| $R_p(t)$ | Recovered subpopulation |
| $\varpi$ | Time delay |
| $\kappa$ | Susceptibles' recruitment rate |
| $\beta$ | Rate of transmission of susceptible class to infected class |
| $\gamma$ | Rate of transmission of aware class to infected class |
| $\sigma$ | Average contact partners |
| $\xi$ | Awareness rate in susceptibles |
| $\vartheta$ | Natural mortality rate |
| $d$ | Death rate due to disease |
| $\theta$ | Recovery rate |
| $a$ | Cure rate |
| $b$ | Limitations rate in treatment availability |
| $c$ | Saturation constant |
| $\Psi(X_p(t), I_p(t))$ | Michaelis-Menten incidence rate among susceptibles |
| $\Lambda(X_p(t), I_p(t))$ | Michaelis-Menten incidence rate among aware individuals |
| $h(I_p(t))$ | Saturated treatment rate of infected individuals |

represents the saturated treatment rate of infectives, where $a$ denotes the maximum treatment (cure) rate, $b$ denotes the rate of limitations in treatment availability, and $c$ denotes the saturation constant. The description of the parameters is given in Table 1 and the transition diagram of the model (1) is shown in Fig. 1.

The resulting mathematical disease-transmission and control model based on the above assumptions is presented under the following system of delay differential equations:

$$\begin{aligned}
\frac{\mathrm{d}X_p}{\mathrm{d}t} &= \kappa - \vartheta X_p - \frac{\beta\sigma X_p(t-\varpi)I_p(t-\varpi)}{1+I_p(t-\varpi)} - \xi X_p,\\
\frac{\mathrm{d}A_p}{\mathrm{d}t} &= \xi X_p - \vartheta A_p - \frac{\gamma\sigma A_p(t-\varpi)I_p(t-\varpi)}{1+I_p(t-\varpi)},\\
\frac{\mathrm{d}I_p}{\mathrm{d}t} &= \frac{\beta\sigma X_p(t-\varpi)I_p(t-\varpi)}{1+I_p(t-\varpi)} + \frac{\gamma\sigma A_p(t-\varpi)I_p(t-\varpi)}{1+I_p(t-\varpi)}\\
&\quad - (\vartheta + d + \theta)I_p - \frac{aI_p^2}{bI_p^2 + cI_p + 1},\\
\frac{\mathrm{d}R_p}{\mathrm{d}t} &= \theta I_p - \vartheta R_p + \frac{aI_p^2}{bI_p^2 + cI_p + 1}.
\end{aligned} \tag{1}$$

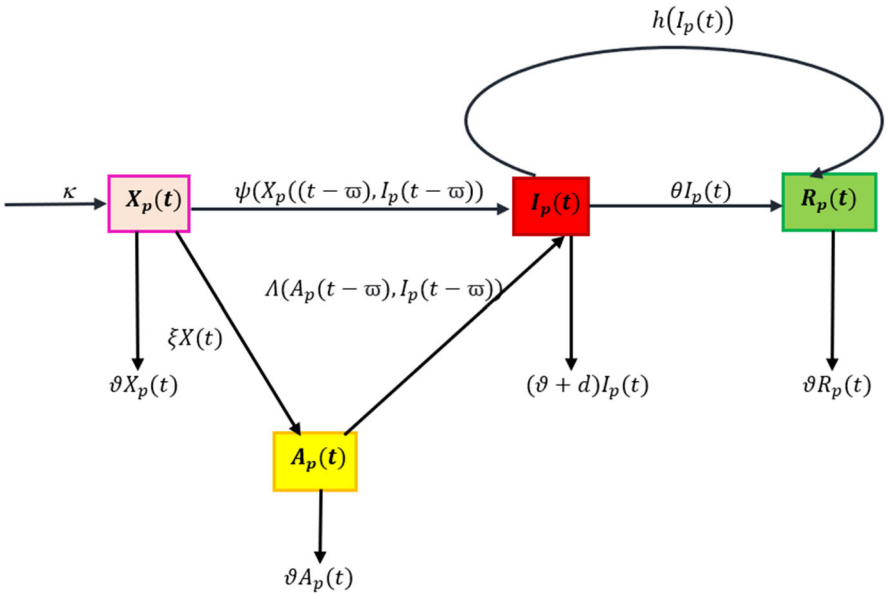



**Fig. 1** Block diagram of the model (1)

subject to the initial conditions $\phi = (\phi_1, \phi_2, \phi_3, \phi_4)$ defined in the Banach space

$$C_+ = \{\phi \in C([-\varpi, 0], R_+^4) : \phi_1(\Omega) = X_p(\Omega), \phi_2(\Omega) = A_p(\Omega), \\ \phi_3(\Omega) = I_p(\Omega), \phi_4(\Omega) = R_p(\Omega)\}, \tag{2}$$

where $R_+^4 = \{(X_p, A_p, I_p, R_p) \in R^4 : X_p \geq 0, A_p \geq 0, I_p \geq 0, R_p \geq 0\}$. Biologically, it is assumed that $\phi_i > 0 \quad (i = 1, 2, 3, 4)$.
We observe that the incidence functions $\Psi(X_p(t-\varpi), I_p(t-\varpi))$, $\Lambda(A_p(t-\varpi), I_p(t-\varpi))$ and the treatment rate function $h(I_p(t))$ are continuously differentiable, positive, and monotonically increasing for all $X_p(t),\ A_p(t),\ I_p(t) > 0$. That is, the following postulates hold:

**P1.** $\Psi(X_p(t), I_p(t)) > 0,\ \Psi'_{X_p}(X_p(t), I_p(t)) > 0,\ \Psi'_{I_p}(X_p(t), I_p(t)) > 0$ for $X_p(t) > 0$,
$\Lambda(A_p(t), I_p(t)) > 0,\ \Lambda'_{A_p}(A_p(t), I_p(t)) > 0,\ \Lambda'_{I_p}(A_p(t), I_p(t)) > 0$ for $A_p(t) > 0$ and $I_p(t) > 0$.

**P2.** $\Psi(X_p(t), 0) = \Psi(0, I_p(t)) = 0,\ \Psi'_{X_p}(X_p(t), 0) = 0,\ \Psi'_{I_p}(X_p(t), 0) > 0$ for $X_p(t) > 0, I_p(t) > 0$ and,
$\Lambda(A_p(t), 0) = \Lambda(0, I_p(t)) = 0,\ \Lambda'_{A_p}(A_p(t), 0) = 0,\ \Lambda'_{I_p}(A_p(t), 0) > 0$ for $A_p(t) > 0, I_p(t) > 0$.

**P3.** $h(0) = 0, h'(0) > 0$ for $I_p(t) \geq 0$.

## 3 Basic properties

Due to biological reasons, all the model's parameters $\kappa,\ \vartheta,\ \beta,\ \sigma,\ \xi,\ \gamma,\ d,\ \theta,\ a,\ b,\ c$ are assumed to be positive.

**Positivity:** We see that the positivity of the above initial conditions for $X_p$, $A_p$, $I_p$ and $R_p$ in $[-\varpi, 0]$ implies positivity for all solutions $(X_p(t), A_p(t), I_p(t), R_p(t))$, $t > 0$ of model (1). We further note that $X_p(t)$ can never vanish since at each time $t > 0$ where $X_p(t)$ vanishes, it is $\frac{\mathrm{d}X_p}{\mathrm{d}t} = \kappa > 0$.
We prove the following lemma.

**Lemma 1** *The compact set*

$$D = \left\{ \left(X_p, A_p, I_p, R_p\right) \in R_+^4 : X_p(t) + A_p(t) + I_p(t) + R_p(t) \leq \frac{\kappa}{\vartheta} \right\}$$

*is globally attractive and invariant for the solutions of* (1).

***Proof*** The model (1) is well-posed as the right-hand side of the model (1), and its derivatives are continuous. Addition of model's equations results to

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}(X_p(t) + A_p(t) + I_p(t) + R_p(t)) &= \kappa - \vartheta(X_p + A_p + I_p + R_p) - dI \\ &\leq \kappa - \vartheta(X_p + A_p + I_p + R_p). \end{aligned} \tag{3}$$

Thus, we obatin

$$\begin{aligned} 0 &< X_p(t) + A_p(t) + I_p(t) + R_p(t) \\ &\leq \frac{\kappa}{\vartheta} - \left(X_p(0) + A_p(0) + I_p(0) + R_p(0) - \frac{\kappa}{\vartheta}\right) e^{-\vartheta t} \end{aligned} \tag{4}$$

Thus, the solutions of model (1) exists in the invariant region

$$0 < \lim_{t\to\infty} (X_p(t) + A_p(t) + I_p(t) + R_p(t)) \leq \frac{\kappa}{\vartheta}. \tag{5}$$

Thus, the solutions of the model (1) are closed and bounded. □

## 4 Mathematical analysis

This section obtains the disease-free equilibrium (DFE), then the basic reproduction number $R_0$, and investigates the system's stability at DFE. We show the existence of the endemic equilibrium (EE) and examine its stability and show the presence of Hopf bifurcation around it. Also, the global stability behavior of the DFE and EE is investigated with the help of $R_0$.
We observe that the first three equations of the model (1) are free from $R(t)$; therefore, without loss of generality, we can restrict our analysis to the following reduced system

of delay differential equation:

$$\begin{aligned}
\frac{\mathrm{d}X_p}{\mathrm{d}t} &= \kappa - \vartheta X_p - \frac{\beta\sigma X_p(t-\varpi)I_p(t-\varpi)}{1+I_p(t-\varpi)} - \xi X_p,\\
\frac{\mathrm{d}A_p}{\mathrm{d}t} &= \xi X_p - \vartheta A_p - \frac{\gamma\sigma A_p(t-\varpi)I_p(t-\varpi)}{1+I_p(t-\varpi)},\\
\frac{\mathrm{d}I_p}{\mathrm{d}t} &= \frac{\beta\sigma X_p(t-\varpi)I_p(t-\varpi)}{1+I_p(t-\varpi)} + \frac{\gamma\sigma A_p(t-\varpi)I_p(t-\varpi)}{1+I_p(t-\varpi)}\\
&\quad - (\vartheta + d + \theta)I_p - \frac{aI_p^2}{bI_p^2 + cI_p + 1}.
\end{aligned} \tag{6}$$

Taking the system (6) into rest, we find that the system (6) has two equilibria:

(i) The disease-free equilibrium (DFE): $E_0\,(X_0, A_0, I_0) = E_0\left(\frac{\kappa}{\vartheta+\xi}, \frac{\kappa\xi}{\vartheta(\vartheta+\xi)}, 0\right)$.
(ii) The positive or endemic equilibrium (EE): $E_e(X_p^*, A_p^*, I_p^*)$, where $X_p^*$, $A_p^*$ and $I_p^*$ are obtained in Subsect. 4.2.

### 4.1 Local stability of disease-free equilibrium

The characteristic equation of the system (6) at DFE $E_0(X_0, A_0, I_0) = E_0\left(\frac{\kappa}{\vartheta+\xi}, \frac{\kappa\xi}{\vartheta(\vartheta+\xi)}, 0\right)$ is obtained as

$$(-\vartheta - \xi - \lambda)(-\vartheta - \lambda)\left(\frac{\kappa\sigma \mathrm{e}^{-\lambda\varpi}}{\vartheta(\vartheta+\xi)}(\beta\vartheta + \xi\gamma) - (\vartheta + d + \theta + \lambda)\right) = 0. \tag{7}$$

The roots of the Eq. (7) are $\lambda_1 = -\vartheta - \xi$, $\lambda_2 = -\vartheta$, and the solutions of the transcendental equation

$$\frac{\kappa\sigma}{\vartheta(\vartheta+\xi)}(\beta\vartheta + \xi\gamma)\mathrm{e}^{-\lambda\varpi} - (\vartheta + d + \theta) - \lambda = 0. \tag{8}$$

Assume that

$$g(\lambda) := \lambda + \vartheta + d + \theta - \frac{\kappa\sigma(\beta\vartheta + \xi\gamma)\mathrm{e}^{-\lambda\varpi}}{\vartheta(\vartheta+\xi)} = 0. \tag{9}$$

We define the term $\frac{\kappa\sigma(\beta\vartheta+\xi\gamma)\mathrm{e}^{-\lambda\varpi}}{\vartheta(\vartheta+\xi)(\vartheta+d+\theta)}$ at $\varpi = 0$ as the basic reproduction number $R_0$ of the system (6). The basic reproduction number is defined as "the average number of secondary infections caused by a single infected agent, during his/her entire infectious period, in an entirely susceptible population" [43]. Thus, the system (6) has $R_0$ of the form

$$R_0 = \frac{\kappa\sigma(\beta\vartheta + \xi\gamma)}{\vartheta(\vartheta+\xi)(\vartheta+d+\theta)}.$$

#### 4.1.1 Analysis for $R_0 \neq 1$

The roots $\lambda_1$ and $\lambda_2$ of Eq. (7) preserve negative signs. Therefore, the analysis is now based on the Eq. (9). We see that

$$g(0) = \vartheta + d + \theta - \frac{\kappa\sigma(\beta\vartheta + \xi\gamma)}{\vartheta(\vartheta + \xi)} = (\vartheta + d + \theta)(1 - R_0). \tag{10}$$

If $R_0 > 1$, then $g(0) < 0$. Also,

$$g'(\lambda) = 1 + \varpi \frac{\kappa\sigma(\beta\vartheta + \xi\gamma)\mathrm{e}^{-\lambda\varpi}}{\vartheta(\vartheta + \xi)} > 0. \tag{11}$$

Hence, $g(0) < 0$ and $g'(\lambda) > 0$ imply that when $R_0 > 1$, then the root of $g(\lambda) = 0$ is unique, real and positive.
Note that,

$$\begin{aligned} \mathrm{Re}\,\lambda &= \frac{\kappa\sigma(\beta\vartheta + \xi\gamma)\cos(Im\lambda)\varpi}{\vartheta(\vartheta + \xi)}\mathrm{e}^{-(\mathrm{Re}\,\lambda)\varpi} - (\vartheta + d + \theta) \\ &< \frac{\kappa\sigma(\beta\vartheta + \xi\gamma)}{\vartheta(\vartheta + \xi)} - (\vartheta + d + \theta) \\ &< 0, \quad \text{provided } R_0 < 1. \end{aligned} \tag{12}$$

Therefore, $R_0 < 1$ implies that Eq. (7) has a root $\lambda$ with negative real part. Thus, the following theorem is stated:

**Theorem 1** *For $\varpi \geq 0$, the disease-free equilibrium (DFE) $E_0$ is*

1. *locally asymptotically stable if $R_0 < 1$*
2. *unstable if $R_0 > 1$.*

#### 4.1.2 Analysis at $R_0 = 1$

Now we analyze the system (6) at $E_0$ and $R_0 = 1$ for $\varpi > 0$ and $\varpi = 0$, seperately.

**Case (i) $\varpi > 0$**
When $R_0 = 1$, then Eq. (9) has a simple characteristic root $\lambda = 0$.
It is noticed that $R_0 = 1$ gives $\kappa\sigma(\beta\vartheta + \xi\gamma) = \vartheta(\vartheta + \xi)(\vartheta + d + \theta)$.
Let $\lambda = p + \mathrm{i}q$ be the other solution of Eq. (9), then we get:

$$p + \mathrm{i}q + d + \vartheta + \theta - (\cos q\varpi - \mathrm{i}\sin q\varpi)(d + \vartheta + \theta)\mathrm{e}^{-p\varpi} = 0. \tag{13}$$

Applying Euler's formula and then splitting real part and imaginary part of Eq. (13) yields

$$\begin{aligned} p + d + \vartheta + \theta &= \mathrm{e}^{-p\varpi}\ (d + \vartheta + \theta)\ \cos q\varpi, \\ q &= -(d + \vartheta + \theta)\ \mathrm{e}^{-p\varpi}\ \sin q\varpi. \end{aligned} \tag{14}$$

A root that satisfies both equations of (14) must be a solution to the equation attained by squaring and adding these two equations. Hence, we get

$$(p + d + \vartheta + \theta)^2 + q^2 = (d + \vartheta + \theta)^2 \, \mathrm{e}^{-2p\varpi}. \tag{15}$$

For Eq. (15) to hold, we must have $p \leq 0$. Thus, $E_0$ is linearly neutrally stable. Thus, the following theorem is stated:

**Theorem 2** *The disease-free equilibrium $E_0$ at $R_0 = 1$ of the system (6) is linearly nuetrally stable for $\varpi > 0$.*

**Case (ii) $\varpi = 0$**

In this case, we study the qualitative behavior of the undelayed system (6) (i.e., $\varpi = 0$) through the stability analysis near critical points, i.e., at $E_0$ and $R_0 = 1$, using the bifurcation theory approach [7], depending upon the center manifold theory [21]. Here, we are keen on evaluating if there is a stable coexistence endemic equilibrium that bifurcates from $E_0$, and $E_0$ converts its behavior from stable to unstable, which is known as forward bifurcation.

For simplicity, let $X_p = x_1$, $A_p = x_2$, and $I_p = x_3$. So, the system (6) reduces to

$$\begin{aligned}
\frac{\mathrm{d}x_1}{\mathrm{d}t} &= \kappa - \vartheta x_1 - \frac{\beta\sigma x_1(t)x_3(t)}{1 + x_3(t)} - \xi x_1 \equiv f_1, \\
\frac{\mathrm{d}x_2}{\mathrm{d}t} &= \xi x_1 - \vartheta x_2 - \frac{\gamma\sigma x_2(t)x_3(t)}{1 + x_3(t)} \equiv f_2, \\
\frac{\mathrm{d}x_3}{\mathrm{d}t} &= \frac{\beta\sigma x_1(t)x_3(t)}{1 + x_3(t)} + \frac{\gamma\sigma x_2(t)x_3(t)}{1 + x_3(t)} - (\vartheta + d + \theta)x_3 - \frac{ax_3^2}{bx_3^2 + cx_3 + 1} \equiv f_3.
\end{aligned} \tag{16}$$

Observe that $R_0 = 1 \iff$ the bifurcation parameter $\sigma = \sigma^* = \frac{\vartheta(\vartheta+\xi)(\vartheta+d+\theta)}{\kappa(\beta\vartheta+\xi\gamma)}$.

The Jacobian matrix $J(E_0, \sigma^*)$ of (16) obtained at $E_0$ and $\sigma^*$ is

$$J(E_0, \sigma^*) = \begin{bmatrix} -\vartheta - \xi & 0 & -\frac{\beta\sigma^*\kappa}{\vartheta+\xi} \\ \xi & -\vartheta & -\frac{\gamma\sigma^*\kappa\xi}{\vartheta(\vartheta+\xi)} \\ 0 & 0 & 0 \end{bmatrix}.$$

$J(E_0, \sigma^*)$ has eigenvalues $\lambda_1 = -\vartheta - \xi$, $\lambda_2 = -\vartheta$, and $\lambda_3 = 0$. Since $\lambda_3$ is a simple zero eigenvalue of $J(E_0, \sigma^*)$. Hence, when $R_0 = 1$, the DFE $E_0$ is a non-hyperbolic equilibrium.

The right eigenvector $v = (v_1, v_2, v_3)$ corresponding to $\lambda_3 = 0$ of the Jacobian matrix $J(E_0,\ \sigma^*)$ is obtained as

$$\begin{aligned} v_1 &= -\frac{\beta\kappa\sigma^*}{(\vartheta+\xi)^2}, \\ v_2 &= -\frac{\kappa\vartheta\xi\sigma^*(\beta+\gamma)+\gamma\kappa\xi^2\sigma^*}{\vartheta^2(\vartheta+\xi)^2}, \\ v_3 &= 1. \end{aligned}$$

The left eigenvector $w = (w_1, w_2, w_3)$ of the Jacobian matrix $J(E_0,\ \sigma^*)$ corresponding to $\lambda_3 = 0$ is obtained as

$$\begin{aligned} w_1 &= 0, \\ w_2 &= 0, \\ w_3 &= 1. \end{aligned}$$

Assume that the right-hand side of the system (16) is denoted by $f_k$. Then, the bifurcation cofficients $a_1$ and $b_1$ defined in Theorem 4.1 of [7] are given by:

$$\begin{aligned} a_1 &= \sum_{k,i,j=1}^{3} w_k v_i v_j \left(\frac{\partial^2 f_k}{\partial x_i \partial x_j}\right)_{E_0}, \\ b_1 &= \sum_{k,i=1}^{3} w_k v_i \left(\frac{\partial^2 f_k}{\partial x_i \partial \sigma^*}\right)_{E_0}. \end{aligned}$$

The non-zero partial derivative of the functions $f_k$'s calculated at $E_0$ are evaluated as

$$\left(\frac{\partial^2 f_3}{\partial x_3 \partial x_1}\right)_{E_0} = \beta\sigma^*, \left(\frac{\partial^2 f_3}{\partial x_3 \partial x_2}\right)_{E_0} = \gamma\sigma^*, \left(\frac{\partial^2 f_3}{\partial x_1 \partial x_3}\right)_{E_0} = \beta\sigma^*,$$
$$\left(\frac{\partial^2 f_3}{\partial x_2 \partial x_3}\right)_{E_0} = \gamma\sigma^*, \left(\frac{\partial^2 f_3}{\partial x_3^2}\right)_{E_0} = -2a - 2\beta\sigma^*\left(\frac{\kappa}{\vartheta+\xi}\right) - 2\gamma\sigma^*\left(\frac{\kappa\xi}{\vartheta(\vartheta+\xi)}\right),$$
$$\text{and } \left(\frac{\partial^2 f_3}{\partial x_3 \partial \sigma^*}\right)_{E_0} = \frac{\beta\kappa\vartheta+\gamma\kappa\xi}{\vartheta(\vartheta+\xi)}.$$

Thus, the coefficients $a_1$ and $b_1$ are computed as

$$a_1 = -\frac{2(a\vartheta^2(\vartheta+\xi)^2+\kappa\sigma^*(\beta^2\,\vartheta^2\sigma^*+\gamma\xi(\vartheta+\xi)(\vartheta+\gamma\,\sigma^*)+\beta\vartheta(\vartheta^2+\vartheta\xi+\gamma\xi\,\sigma^*)))}{\vartheta^2(\vartheta+\xi)^2}, \tag{17}$$

$$b_1 = \frac{\beta\kappa\vartheta+\gamma\kappa\xi}{\vartheta(\vartheta+\xi)}. \tag{18}$$

By Theorem 4.1 given by Castillo-Chavez and Song [7], the local dynamics of system (16) around $E_0$ at $R_0 = 1$ is completely determined by the signs of bifurcation

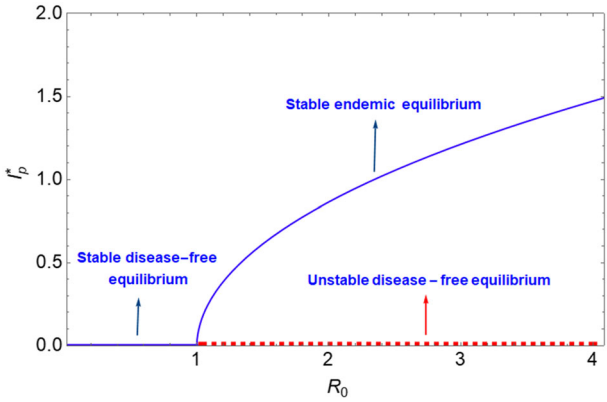



**Fig. 2** $R_0$ vs. $I_p^*$

coefficients $a_1$, and $b_1$. More precisely, if $a_1 < 0$ and $b_1 > 0$, then the bifurcation is forward; if $a_1 > 0$ and $b_1 > 0$ then the bifurcation is backward. Since, the bifurcation coefficients obtained in Eqs. (17-18) shows that $a_1$ is always negative, and $b_1$ is always positive, therefore, the bifurcation is forward. "In the forward bifurcation [22], when $R_0 < 1$, a small influx of infected individuals will not generate large outbreaks, and the disease dies out in time, and the corresponding disease-free equilibrium is asymptotically-stable. On the other hand, when $R_0$ exceeds unity, then the disease will persist with a stable endemic equilibrium. This phenomenon, where the disease-free equilibrium loses its stability and a stable endemic equilibrium appears as the basic reproduction number increases through one, is known as forward bifurcation" The main characteristics of forward bifurcation are :

1. There is no endemic equilibrium near disease-free equilibrium when $R_0 < 1$. That is, the disease-free equilibrium is often the only equilibrium when $R_0 < 1$.
2. A low level of endemicity when $R_0$ is slightly above unity.

Thus, the following theorem is concluded:

**Theorem 3** *The disease-free equilibrium $E_0$ at $R_0 = 1$ of the system (6) exhibits a forward bifurcation at $E_0$ and $R_0 = 1$ when $\varpi = 0$.*

Figure 2 illustrates the result of Theorem 3 for the following numerical experminetal data: $\kappa = 2$, $\beta = 0.09$, $\xi = 0.12$, $\sigma = 0.08$, $\vartheta = 0.01$, $\gamma = 0.009$, $d = 0.08$, $\theta = 0.03$, $a = 0.005$, $b = 0.01$, $c = 0.03$. At these values of parameters, bifurcation constants $a_1 = -0.543519 < 0$ and $b_1 = 3.04615 > 0$, which satisfies the theoretical condition for the occurrence of forward bifurcation. This figure provides the qualitative description of infectives when $R_0$ varies from unity. It depicts that when $R_0$ crosses unity from below, a small positive asymptotically stable endemic equilibrium exists, and $E_0$ changes its stability from stable to unstable.

### 4.2 Existence and local stability of the endemic equilibrium

Now, we determine the conditions for endemic equilibrium existence. For that, assuming that $X_p,\ A_p,\ I_p\ \neq 0$ and setting the system (6) to zero, we get:

$$\kappa - \vartheta X_p - \frac{\beta\sigma X_p I_p}{1+I_p} - \xi X_p = 0, \tag{19}$$

$$\xi X_p - \vartheta A_p - \frac{\gamma\sigma A_p I_p}{1+I_p} = 0, \tag{20}$$

$$\frac{\beta\sigma X_p I_p}{1+I_p} + \frac{\gamma\sigma A_p I_p}{1+I_p} - (\vartheta + d + \theta)I_p - \frac{aI_p^2}{bI_p^2 + cI_p + 1} = 0. \tag{21}$$

On obtaining $X_p$ from Eq. (19) and $A_p$ from Eq. (20), and then substituting it in (21), the following quartic equation in $I_p$ is obtained:

$$F(I_p) := K_0 + K_1 I_p + K_2 I_p^2 + K_3 I_p^3 + K_4 I_p^4 = 0, \tag{22}$$

where,

$$\begin{aligned}
K_0 &= \vartheta(\vartheta+\xi)(d+\vartheta+\theta)(1-R_0),\\
K_1 &= a\vartheta(\xi+\vartheta) + (d+\theta+\vartheta)\left(\sigma\vartheta(\beta+\gamma) + \gamma\xi\sigma + (\xi+\vartheta)\left(c\left(1-R_0\right)\vartheta + \vartheta\right)\right)\\
&\quad + (\vartheta(\xi+\vartheta)(d+\theta+\vartheta) - \kappa\sigma(\beta(\gamma\sigma+\vartheta)+\gamma\xi)),\\
K_2 &= a\sigma(\beta\vartheta + \gamma(\xi+\vartheta)) + 2a\vartheta(\xi+\vartheta) + b\left(1-R_0\right)\vartheta(\xi+\vartheta)(d+\theta+\vartheta)\\
&\quad + (d+\theta+\vartheta)\left(\beta\gamma\sigma^2 + (c+1)\sigma(\beta\vartheta+\gamma(\xi+\vartheta)) + (c+1)\vartheta(\xi+\vartheta)\right)\\
&\quad + c(\vartheta(\xi+\vartheta)(d+\theta+\vartheta) - \kappa\sigma(\beta(\gamma\sigma+\vartheta)+\gamma\xi)),\\
K_3 &= a(\gamma\sigma+\vartheta)(\beta\sigma+\xi+\vartheta) + (d+\theta+\vartheta)\left(\sigma(b+c)(\beta\vartheta+\gamma(\xi+\vartheta))\right.\\
&\quad \left.+\vartheta(b+c)(\xi+\vartheta) + \beta c\gamma\sigma^2\right)\\
&\quad + b(\vartheta(\xi+\vartheta)(d+\theta+\vartheta) - \kappa\sigma(\beta(\gamma\sigma+\vartheta)+\gamma\xi)),\\
K_4 &= b(\sigma\gamma+\vartheta)(d+\theta+\vartheta)(\sigma\beta+\vartheta+\xi).
\end{aligned} \tag{23}$$

For the positive root $I_p^*$ of polynomial $F(I_p)$, we can make

$$X_p^* = \frac{(1+I_p^*)\kappa}{\sigma I_p^*\beta + (1+I_p^*)(\vartheta+\xi)} > 0. \tag{24}$$

and

$$A_p^* = \frac{(1+I_p^*)^2\kappa\xi}{(\sigma I_p^*\gamma + \vartheta + I_p^*\vartheta)(\sigma I_p^*\beta + (1+I_p^*)(\vartheta+\xi))} > 0. \tag{25}$$

So, $E_e(X_p^*, A_p^*, I_p^*)$ is an endemic equilibrium of (6).

In this paper, we consider the case of the existence of a unique endemic equilibrium only.

**Theorem 4** *When $R_0 < 1$, the system (6) has no endemic equilibrium if*

$$(\vartheta(\xi+\vartheta)(d+\theta+\vartheta) > \kappa\sigma(\beta(\gamma\sigma+\vartheta)+\gamma\xi)).$$

***Proof*** Let $R_0 < 1$, which implies $1 - R_0 > 0$, and so $K_0 > 0$. Also, $K_4 > 0$. The coefficients $K_2$, $K_3$, and $K_4$ are positive if

$$(\vartheta(\xi+\vartheta)(d+\theta+\vartheta) - \kappa\sigma(\beta(\gamma\sigma+\vartheta)+\gamma\xi)) > 0$$

Or

$$(\vartheta(\xi+\vartheta)(d+\theta+\vartheta) > \kappa\sigma(\beta(\gamma\sigma+\vartheta)+\gamma\xi)).$$

Hence, by Descartes' rule of signs [44], Eq. (22) has no positive root under the condition $(\vartheta(\xi+\vartheta)(d+\theta+\vartheta) > \kappa\sigma(\beta(\gamma\sigma+\vartheta)+\gamma\xi))$ and so the system (6) has no positive equilibrium under this condition. □

**Theorem 5** *When $R_0 > 1$, there exists either a unique or three endemic equilibria if all equilibria are simple roots.*

***Proof*** Let $R_0 > 1$. Then, the coefficient $K_0 < 0$. Also, $K_4 > 0$. From Eq. (22), we have

$$F(I_p) := K_0 + K_1 I_p + K_2 I_p^2 + K_3 I_p^3 + K_4 I_p^4 = 0.$$

The conditions $\mathbf{U}_1$–$\mathbf{U}_8$, given below, define the possibilities for the signs of $K_1$, $K_2$, and $K_3$ as follows:

$$\begin{aligned}
\mathbf{U}_1 &: K_1 > 0,\ K_2 > 0,\ \text{and } K_3 > 0,\\
\mathbf{U}_2 &: K_1 < 0,\ K_2 < 0,\ \text{and } K_3 > 0,\\
\mathbf{U}_3 &: K_1 < 0,\ K_2 > 0,\ \text{and } K_3 > 0,\\
\mathbf{U}_4 &: K_1 < 0,\ K_2 < 0,\ \text{and } K_3 < 0,\\
\mathbf{U}_5 &: K_1 > 0,\ K_2 > 0,\ \text{and } K_3 < 0,\\
\mathbf{U}_6 &: K_1 > 0,\ K_2 < 0,\ \text{and } K_3 > 0,\\
\mathbf{U}_7 &: K_1 > 0,\ K_2 < 0,\ \text{and } K_3 < 0,\\
\mathbf{U}_8 &: K_1 < 0,\ K_2 > 0,\ \text{and } K_3 < 0.
\end{aligned} \tag{26}$$

By Descartes' rule of signs [44], under the conditions $\mathbf{U}_1$–$\mathbf{U}_8$, the equation $F(I_p)$ can have either a unique or three positive roots. Under any of the conditions $\mathbf{U}_1$–$\mathbf{U}_4$, there exists a unique endemic equilibrium, and under $\mathbf{U}_5$–$\mathbf{U}_8$, there exsits either a unique or three endemic equilibria.
Thus, we state the following theorem: □

**Theorem 6** *Assume that any of the conditions **H1:** ($\mathbf{U}_1$–$\mathbf{U}_8$), and $R_0 > 1$ hold, then the system (6) has a unique endemic equilibrium.*

Next, we examine the stability behavior of the unique endemic equilibrium. At $E_e$, the characteristic equation of the system (6) is obtained as

$$\lambda^3+L_2\lambda^2+L_1\lambda+L_0+\left(M_2\lambda^2+M_1\lambda+M_0\right)\mathrm{e}^{-\lambda\varpi}+(N_1\lambda+N_0)\,\mathrm{e}^{-2\lambda\varpi}=0, \quad (27)$$

where,

$$L_2 = d + \frac{aI_p^*(2+cI_p^*)}{(1+I_p^*(c+bI_p^*))^2} + \theta + 3\vartheta + \xi,$$

$$L_1 = 2\theta\vartheta + 3\vartheta^2 + \theta\xi + 2\vartheta\xi + d(2\vartheta+\xi) + \frac{a(2\vartheta+\xi)I_p^*(2+cI_p^*)}{(1+I_p^*(c+bI_p^*))^2},$$

$$L_0 = \vartheta\left(d + \frac{aI_p^*(2+cI_p^*)}{(1+I_p^*(c+bI_p^*))^2} + \theta + \vartheta\right)(\vartheta+\xi),$$

$$M_2 = \frac{(-X_p^*\beta - A_p^*\gamma + I_p^*(1+I_p^*)(\beta+\gamma))\sigma}{(1+I_p^*)^2},$$

$$M_1 = \frac{aI_p^{*2}(2+cI_p^*)(\beta+\gamma)\sigma}{(1+I_p^*)(1+I_p^*(c+bI_p^*))^2} + \frac{I_p^*((\beta+\gamma)(d+\theta+2\vartheta)+\gamma\xi)\sigma}{1+I_p^*} - \frac{(X_p^*\beta + A_p^*\gamma)(2\vartheta+\xi)\sigma}{(1+I_p^*)^2},$$

$$M_0 = \frac{I_p^*(d+\theta+\vartheta)(\beta\vartheta+\gamma(\vartheta+\xi))\sigma}{1+I_p^*} + \frac{aI_p^{*2}(2+cI_p^*)(\beta\vartheta+\gamma(\vartheta+\xi))\,\sigma}{(1+I_p^*)(1+I_p^*(c+bI_p^*))^2} - \frac{(X_p^*\beta + A_p^*\gamma)\,\vartheta\,(\vartheta+\xi)\,\sigma}{(1+I_p^*)^2},$$

$$N_1 = \frac{I_p^*(-A_p^* - X_p^* + I_p^* + I_p^{*2})\beta\gamma\sigma^2}{(1+I_p^*)^3},$$

$$N_0 = \frac{I_p^*\beta\gamma(-(A_p^*+X_p^*)\vartheta + I_p^*(1+I_p^*)(d+\theta+\vartheta))\sigma^2}{(1+I_p^*)^3} + \frac{aI_p^{*3}(2+cI_p^*)\beta\gamma\sigma^2}{(1+I_p^*)^2(1+I_p^*(c+bI_p^*))^2}.$$

Multiplying Eq. (27) by $\mathrm{e}^{\lambda\varpi}$, we get

$$M_2\lambda^2+M_1\lambda+M_0+\left(\lambda^3+L_2\lambda^2+L_1\lambda+L_0\right)\mathrm{e}^{\lambda\varpi}+(N_1\lambda+N_0)\,\mathrm{e}^{-\lambda\varpi}=0. \quad (28)$$

When $\varpi = 0$, Eq. (28) becomes

$$\lambda^3 + L_{00}\lambda^2 + L_{01}\lambda + L_{02} = 0, \quad (29)$$

where,

$$\begin{aligned} L_{00} &= M_2 + L_2, \\ L_{01} &= M_1 + L_1 + N_1, \\ L_{02} &= M_0 + L_0 + N_0. \end{aligned}$$

By Routh-Hurwitz criterion, Eq. (29) has all the roots with negative real parts under **H2:**, given below:

$$\textbf{H2:} \quad L_{00} > 0, \quad L_{02} > 0, \quad \text{and} \quad L_{00}L_{01} > L_{02}. \tag{30}$$

Hence, we state the following Theorem:

**Theorem 7** *Assume that **H2** holds. Then, the endemic equilibrium $E_e$ is locally asmptotically stable when $\varpi = 0$.*

Now, for $\varpi > 0$, let Eq. (28) has a root $\mathrm{i}\omega$ ($\omega > 0$). Then, replacing $\lambda$ with $\mathrm{i}\omega$ ($\omega > 0$) in Eq. (28) and splitting real and imaginary parts, we obtain

$$B_1(\omega)\cos\omega\varpi - B_2(\omega)\sin\omega\varpi = B_3(\omega), \tag{31}$$

$$B_4(\omega)\sin\omega\varpi + B_5(\omega)\cos\omega\varpi = B_6(\omega). \tag{32}$$

where,

$$\begin{aligned} B_1(\omega) &= -L_2\omega^2 + L_0 + N_0, \\ B_2(\omega) &= \omega(L_1 - N_1) - \omega^3, \\ B_3(\omega) &= M_2\omega^2 - M_0, \\ B_4(\omega) &= -L_2\omega^2 + L_0 - N_0, \\ B_5(\omega) &= (L_1 + N_1)\omega - \omega^3, \\ B_6(\omega) &= -M_1\omega. \end{aligned} \tag{33}$$

From Eqs. (31) and (32), we obtain

$$\sin\omega\varpi = \frac{B_{01}(\omega)}{B_{00}(\omega)}, \tag{34}$$

$$\cos\omega\varpi = \frac{B_{02}(\omega)}{B_{00}(\omega)}, \tag{35}$$

where,

$$\begin{aligned} B_{00} &= L_0^2 - N_0^2 + (L_1^2 - 2L_0L_2 - N_1^2)\omega^2 + (-2L_1 + L_2^2)\omega^4 + \omega^6, \\ B_{01} &= (L_1M_0 - L_0M_1 - M_1N_0 + M_0N_1)\omega + (-M_0 + L_2M_1 - L_1M_2 - M_2N_1)\omega^3 + M_2\omega^5, \\ B_{02} &= -L_0M_0 + M_0N_0 + (L_2M_0 - L_1M_1 + L_0M_2 - M_2N_0 + M_1N_1)\omega^2 + (M_1 - L_2M_2)\omega^4. \end{aligned} \tag{36}$$

On squaring and then adding Eqs. (34) and (35), we obtain

$$B_{01}^2(\omega) + B_{02}^2(\omega) - B_{00}^2(\omega) = 0. \tag{37}$$

Let **(H3:)** There exists at least one positive root $\omega_0$ of Eq. (37), i.e., Eq. (28) has a pair of purely imaginary roots $\pm i\omega_0$.

For $\omega_0$, the corresponding critical value of the time delay $\varpi_k$ is obtained as follows:

$$\varpi_k = \frac{1}{\omega_0} \arccos\left(\frac{B_{02}(\omega_0)}{B_{00}(\omega_0)}\right) + \frac{2k\pi}{\omega_0}, \qquad k = 0, 1, 2, 3, \ldots. \tag{38}$$

Let $\varpi_0 = \min\{\varpi_k,\ k = 0, 1, 2, 3, \ldots\}$.

To estabish Hopf bifurcation, we must have $Re\left[\frac{d\lambda}{d\varpi}\right]^{-1}\Big|_{\lambda=i\omega_0} \neq 0$.

Differentiating Eq. (28) with respect to $\varpi$ yields

$$\left[\frac{d\lambda}{d\varpi}\right]^{-1} = \frac{2\lambda M_2 + M_1 + N_1 e^{-\lambda\varpi} + e^{\lambda\varpi}(3\lambda^2 + 2\lambda L_2 + L_1)}{\lambda\left((N_1\lambda + N_0)e^{-\lambda\varpi} - (\lambda^3 + L_2\lambda^2 + L_1\lambda + L_0)\right)e^{\lambda\varpi}} - \frac{\varpi}{\lambda}.$$

From Eq. (28), we have

$$-(\lambda^3 + L_2\lambda^2 + L_1\lambda + L_0)e^{\lambda\varpi} = M_2\lambda^2 + M_1\lambda + M_0 + (N_1\lambda + N_0)\, e^{-\lambda\varpi}.$$

Thus, we obtain

$$\left[\frac{d\lambda}{d\varpi}\right]^{-1} = \frac{2\lambda M_2 + M_1 + N_1 e^{-\lambda\varpi} + e^{\lambda\varpi}(3\lambda^2 + 2\lambda L_2 + L_1)}{\lambda\left(2(N_1\lambda + N_0)e^{-\lambda\varpi} + (M_2\lambda^2 + M_1\lambda + M_0)\right)} - \frac{\varpi}{\lambda}.$$

So,

$$Re\left[\frac{d\lambda}{d\varpi}\right]^{-1}\Big|_{\lambda=i\omega_0} = \frac{Y}{P^2 + Q^2},$$

where,

$$\begin{aligned}
Y =& \omega_0(-M_1^2\omega_0 + 2M_0M_2\omega_0 - 2N_1^2\omega_0 - 2M_2^2\omega_0^3 + \omega_0(-L_1M_1 + 4M_2N_0 \\
&- 3M_1N_1 + 3M_1\omega_0^2 \\
&+ 2L_2(M_0 - M_2\omega_0^2))\cos\varpi\omega + 2\omega(2L_2N_0 - L_1N_1 \\
&+ 3N_1\omega_0^2)\cos 2\varpi\omega + (-M_0N_1 - 3M_0\omega_0^2 \\
&+ 5M_2N_1\omega_0^2 + 3M_2\omega_0^4 + 2M_1(N_0 + L_2\omega_0^2) + L_1(M_0 - M_2\omega_0^2))\sin\varpi\omega + 2(L_1N_0 \\
&+ (-3N_0 + 2L_2N_1)\omega_0^2)\sin 2\varpi\omega), \\
P =& -M_1\omega_0^2 - 2N_1\omega_0^2\cos\varpi\omega_0 + 2N_0\omega_0\sin\varpi\omega_0, \\
Q =& M_0\omega_0 - M_2\omega_0^3 + 2N_0\omega_0\cos\varpi\omega_0 + 2N_1\omega_0^2\sin\varpi\omega_0.
\end{aligned}$$

Obviously, if **H4:** $Y \neq 0$, then $Re\left[\frac{d\lambda}{d\varpi}\right]^{-1}\Big|_{\lambda=i\omega_0} \neq 0$.

Thus, the following theorem is stated:

**Theorem 8** *Suppose* ***(H1-H4)*** *holds. Then, the endemic equilibrium* $E_e(X_p^*, A_p^*, I_p^*)$ *of the system* (6) *is*

1. *locally asmptotically stable when* $\varpi \in [0, \varpi_0)$,
2. *undergoes a Hopf bifurcation when* $\varpi = \varpi_0$, *and a family of periodic solutions bifurcate from* $E_e(X_p^*, A_p^*, I_p^*)$ *when* $\varpi$ *crosses* $\varpi_0$.

### 4.3 Global stability

In this subsection, we establish the global stability behavior of the DFE and EE.

#### 4.3.1 Global stability behavior of disease-free equilibrium

We examine the global stability behavior of DFE $E_0\,(X_0, A_0, I_0) = E_0\left(\frac{\kappa}{\vartheta+\xi}, \frac{\kappa\xi}{\vartheta(\vartheta+\xi)}, 0\right)$ for $R_0 \leq 1$ by constructing a suitable Lyapunov function. For this, the following postulates are proposed:

**P4.** $\Psi'_{I_p}(X_p(t), 0)$ is increasing for $X_p(t) > 0$ and $\Lambda'_{I_p}(A_p(t), 0)$ is increasing for $A_p(t) > 0$.

**P5.** $\frac{\Psi'_{I_p}(X_0,0)}{\Psi'_{I_p}(X_p(t),0)} < 1$ for $X_p(t) > X_0$, $\frac{\Psi'_{I_p}(X_0,0)}{\Psi'_{I_p}(X_p(t),0)} > 1$ for $X_p(t) \in (0, X_0)$,
$\frac{\Lambda'_{I_p}(A_0,0)}{\Lambda'_{I_p}(A_p(t),0)} < 1$ for $A_p(t) > A_0$, $\frac{\Lambda'_{I_p}(X_0,0)}{\Lambda'_{I_p}(A_p(t),0)} > 1$ for $A_p(t) \in (0, A_0)$.

**P6.**

$$\begin{aligned}&\Phi(X_p(t), I_p(t)) + \Lambda(A_p(t), I_p(t))\\ &\leq I_p(t)\left(\left(\frac{\partial\Phi(X_p(t), I_p(t))}{\partial I_p}\right)_{(X_0,0)} + \left(\frac{\partial\Lambda(A_p(t), I_p(t))}{\partial I_p}\right)\right)_{(A_0,0)} - \left(\frac{\partial H(I_p)}{\partial I_p}\right)_{I_p=0}\\ &\quad + H(I_p(t)),\ \text{where } H(I_p(t)) = (\vartheta + d + \zeta)I_p(t) + h(I_p(t)) \text{ and } I_p(t) > 0.\end{aligned}$$

**P7.** $\frac{X_p}{X_0} > \frac{A_p}{A_0}$ if $\frac{X_p}{X_0} > 1$ and $\frac{X_p}{X_0} < \frac{A_p}{A_0}$ if $\frac{X_p}{X_0} < 1$.

Under these postulates, the following theorem is being proposed:

**Theorem 9** *Suppose that (P1.)–(P7.) and* $R_0 \leq 1$ *hold. Then, DFE* $E_0(X_0, A_0, 0)$ *of the system* (6) *is globally asymptotically stable for* $\varpi \geq 0$.

The proof has been given in Appendix.

#### 4.3.2 Global stability behavior of endemic equilibrium

Now, we examine the global stability behavior of $E_e(X^*, A^*, I^*)$ of the system (6) by constructing a Lyapunov functional and employing the Lyapunov direct method. To proceed, we propose the following postulates:

**P8.**

$$\left(\frac{\Phi(X_p^*, I_p^*)}{\Phi(X_p(t), I_p^*)} - \frac{I_p^*}{I_p(t)}\right) \leq 0; \left(\frac{\Phi(X_p(t), I_p(t))}{\Phi(X_p^*, I_p^*)} - 1\right) \leq 0;$$
$$\left(\frac{\Phi(X_p(t), I_p^*)}{\Phi(X_p(t), I_p(t))} - \frac{I_p(t)}{I_p^*}\right) \leq 0;$$
$$\left(\frac{\Lambda(A_p^*, I_p^*)}{\Lambda(A_p(t), I_p^*)} - \frac{I_p^*}{I_p(t)}\right) \leq 0; \left(\frac{\Lambda(A_p(t), I_p(t))}{\Lambda(A_p^*, I_p^*)} - 1\right) \leq 0;$$
$$\left(\frac{\Lambda(A_p(t), I_p^*)}{\Lambda(A_p(t), I_p(t))} - \frac{I_p(t)}{I_p^*}\right) \leq 0 \text{ for } I_p \geq I_p^*.$$

**P9.**

$$\left(\frac{h(I_p^*)}{h(I_p(t))} - \frac{I_p^*}{I_p(t)}\right)\left(\frac{I_p(t)}{I^*} - 1\right) \leq 0 \text{ for } I_p \geq I_p^*.$$

**P10.**

$$\frac{X_p}{X_p^*} - \frac{A_p}{A_p^*} > 0 \text{ for } \frac{X_p}{X_p^*} > 1 \text{ and } \frac{X_p}{X_p^*} - \frac{A_p}{A_p^*} < 0 \text{ for } \frac{X_p}{X_p^*} < 1.$$

The following theorem is proposed under these postulates:

**Theorem 10** *Suppose that (P1.)–(P3.), (P8.)–(P10.), and $R_0 > 1$ hold. Then, the EE $E_e(X_p^*, A_p^*, I_p^*)$ of the system (6) is globally asymptotically stable for $\varpi \geq 0$.*

The proof has been given in Appendix.

## 5 Numerical simulation

This section aims to validate the theoretical findings which we have discussed in previous sections and show the significance of various parameters by considering numerical experimental data. For this, we assume that all the parameters are positive. Two examples have been considered herein:-

1. Example 1 illustrates the existence of a unique endemic equilibrium and its stability. Also, it shows the significance of considering various parameters such as time delay, treatment rate, awareness rate, and the corporation of the aware compartment in the SIR epidemic model.
2. Example 2 illustrates the occurrence of Hopf bifurcation via oscillatory and periodic solutions.

***Example 1*** We consider the following experimental data: $\kappa = 2$, $\beta = 0.09$, $\xi = 0.12$, $\sigma = 1$, $\vartheta = 0.01$, $\gamma = 0.009$, $d = 0.08$, $\theta = 0.03$, $a = 0.005$, $b =$

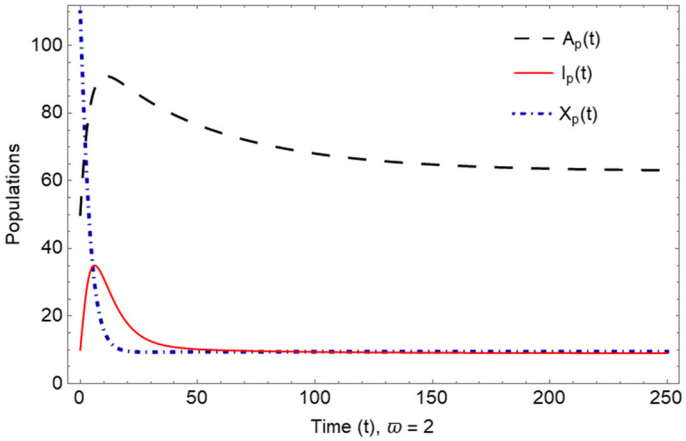



**Fig. 3** The temporal behavior of susceptible, alert, and infected populations for time-delay $\varpi = 2$

0.01, $c = 0.03$. At these values of parameters, we calculate that $R_0 = 25.3846 > 1$ and the coefficients of the polynomial $F(I_p) := K_0 + K_1 I_p + K_2 I_p^2 + K_3 I_p^3 + K_4 I_p^4 = 0$, given in Eq. (22) have the value $K_0 = -0.003804 < 0$, $K_1 = -0.00512722 < 0$, $K_2 = 0.000336322 > 0$, $K_3 = -0.000014248 < 0$, and $K_4 = 5.016 * 10^{-6} > 0$. So, one of the conditions $U_8$ given in Eq. (26) is satisfied. Further, the solutions of the polynomial equation $F(I_p) = 0$ are $I_p = -2.73341 - 10.5534\text{i}$, $-2.73341 + 10.5534\text{i}$, $-0.707827$ and $9.01515$. So, the only positive root of $F(I_p) = 0$ is $I_p^* = 9.01515$. Thus, the conditions of Theorem 6 are satisfied and there exists a unique endemic equilibrium $E_e(9.47806,\ 62.8333,\ 9.01515)$.

Figure 3 shows the qualitative behavior of susceptible, alert, and infected populations for a time delay $\varpi = 2$. It is depicted that the susceptible population decreases over time and a large population gets aware of the disease, and as time passes, they become less serious, and finally both the populations settle down to a steady state. Also, it elucidates that infectives increase at a high rate and then start decreasing and eventually reach their steady state.

Figure 4 indicates the impact of latent period $\varpi$ on infected individuals $I_p(t)$. We can see the variation in the number of infectives for higher values of time delay. This figure confirms that the longer the latent period (time delay) of infection, the higher its spread, which shows the importance of considering time delay in studying infectious disease's dynamical behavior.

Figure 5 shows the influence of transmission rates $\beta$ of susceptibles and $\gamma$ of aware individuals on infectives. The higher the transmission rates, the higher the possibility of spreading infection. Therefore, it is imperative to minimize the interaction rate of susceptible and aware people with infected individuals.

Figure 6 shows the power of cure rate $(a)$ in declining the number of infected individuals. An increased cure rate can reduce the infection level efficiently. Thus, the accessibility of treatment resources and adequate treatment is significant in declining the cases of infected individuals.

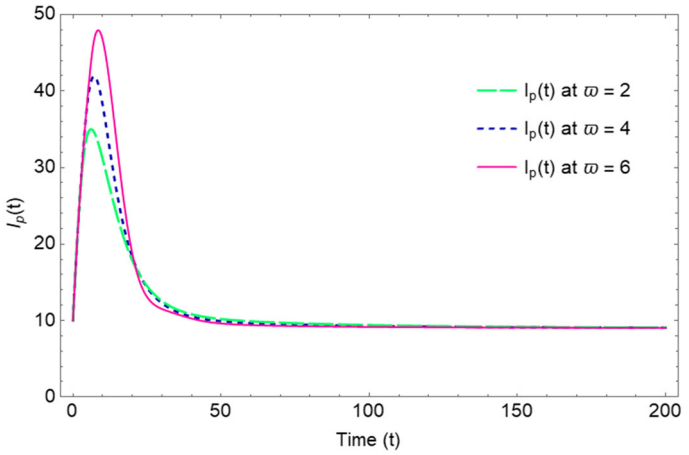
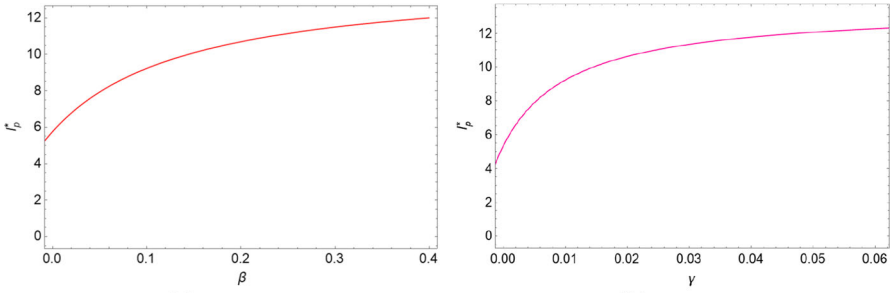



**Fig. 4** Infected population $I_P(t)$ when time delay $\varpi$ varies




**(a)** $I_p^*$ vs. $\beta$ **(b)** $I_p^*$ vs. $\gamma$

**Fig. 5** Variation in the number of infected individuals when transmission rates $\beta$ and $\gamma$ vary

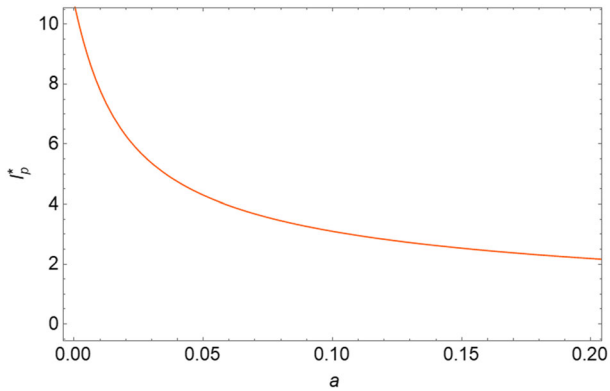



**Fig. 6** Effect of cure rate $(a)$ on $I_p^*$

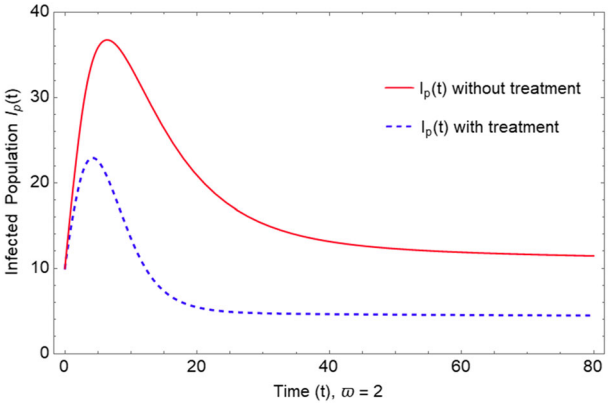



**Fig. 7** Effect of saturating treatment rate $(H(I_p(t)))$ on infected population $(I_p(t))$

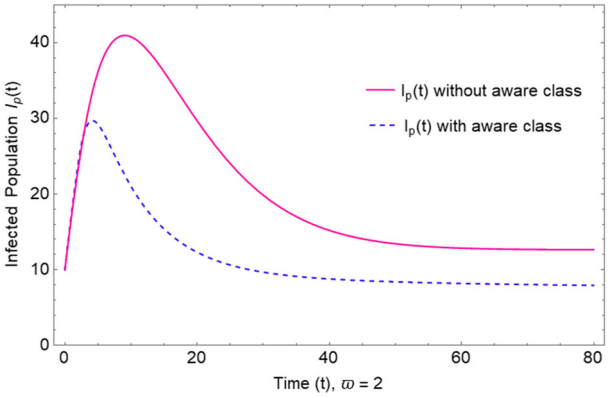



**Fig. 8** Effect of incorporating awareness class on the infected population $I_p(t)$ for $\varpi = 2$

Figure 7 demonstrates the impact of nonlinear saturating treatment rate on infected individuals. The infected population, drawn for the $a = 0.05$, reveals the significance of the availability of treatment to infected people. If the health system has sufficient treatment facilities, then the spread of infection can be controlled on a large scale.

Figure 8 shows the difference in the number of infected individuals with the inclusion and exclusion of aware individuals' compartment, deliberating that unaware individuals are becoming infected faster than those individuals who are familiar with the disease spread. It shows the relationship between human awareness and the spread of infection. The graph of infected individuals with awareness class is drawn for the awareness rate $\xi = 0.6$, revealing that a considerable value of awareness rate causes more individuals to be safe from illnesses. The disease spread awareness alerts people and helps them to take necessary protection measures against disease which reduce the occurrence of infection on a large scale. Thus, the significance of incorporating an alert compartment in the SIR model is vital.

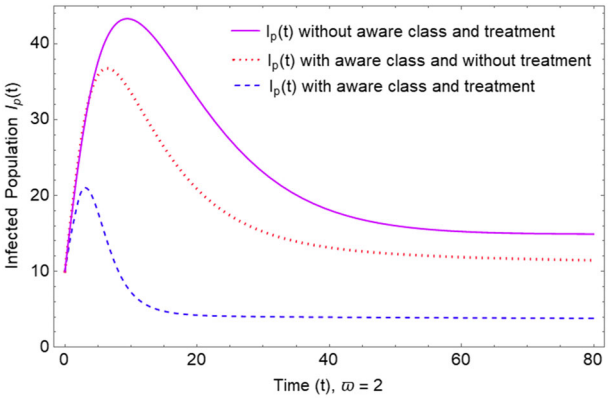



**Fig. 9** Effect of both awareness class $A_p(t)$ and saturated treatment rate on $I_p(t)$ for $\varpi = 2$

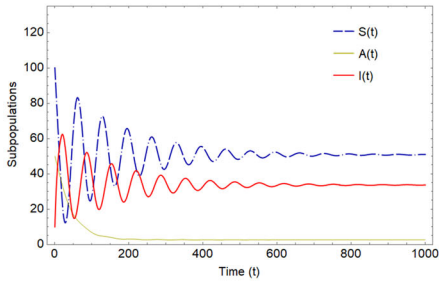
Figure 9 depicts the number of the infected population for the cases: neither awareness nor treatment is available (shown by the solid purple line); people are aware of the disease, but treatment for infectives is not available (indicated by the dotted red line); and when both treatment and awareness are present (indicated thru the dashed blue line). This figure captures the impact of both awareness, and treatment on infectives. When awareness and treatment are absent, the infected population stabilizes at a high level. If the treatment is not available, then the awareness among people reduces the spread of infection with a big difference. For eradicating disease or lowering it to the lowest level, the presence of both awareness among susceptibles and sufficient availability of treatment resources has a vital role.

***Example 2*** The following numerical experimental data is considered to validate the occurrence of Hopf bifurcation:

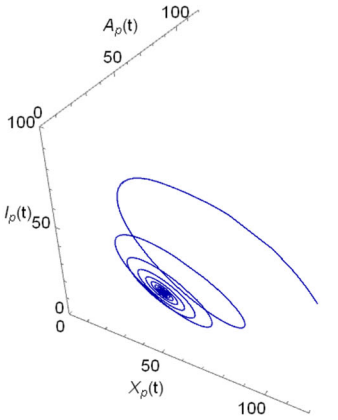
$$\kappa = 5,\ \xi = 0.001,\ \vartheta = 0.01,\ \beta = 0.009,\ \gamma = 0.001,\ \sigma = 10.$$
$$d = 0.08,\ \theta = 0.03,\ a = 0.005,\ b = 0.01,\ c = 0.03.$$

At these values of parameters, it has been verified that Eq. (37) has a unique positive root $\omega_0 = 0.08524468136097717$. For this value of $\omega_0$, the corresponding critical value of the time delay $\varpi_0$ given in the formula of Eq. (38), is obtained as $\varpi_0 = 19.7113$.

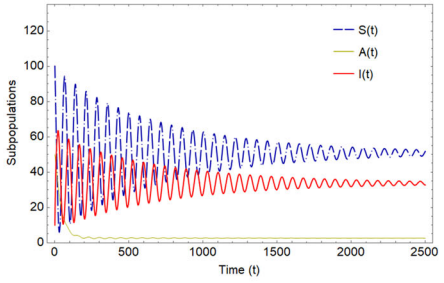
Figure 10 shows the time series solution and respective phase portraits of susceptible, aware and infected populations. Figures. 10(a)-10(d) have been plotted for $\varpi = 17$, and $\varpi = 19$, which reveals that the endemic equilibrium is stable when the time delay is less than its critical value $\varpi_0 = 19.7113$. On the other hand, Figs. 10(e) and 10(f) have been plotted for the time delay $\varpi = 19.8$ which reveals that the orbits spiral goes away from endemic equilibrium as $\varpi$ crosses its critical value $\varpi_0 = 19.7113$, and the endemic equilibrium turns into unstable. These figures confirm the result of the Theorem 8.



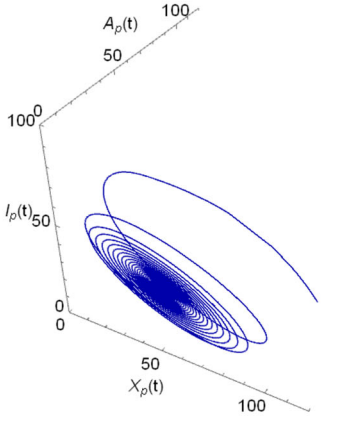

**(a)** $X_p - A_p - I_p$ at $\varpi = 17$.



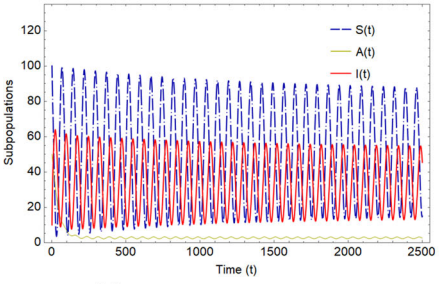

**(b)** $X_p - A_p - I_p$ at $\varpi = 17$.



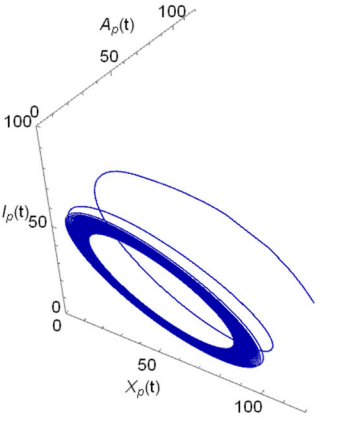

**(c)** $X_p - A_p - I_p$ at $\varpi = 19$.




**(d)** $X_p - A_p - I_p$ at $\varpi = 19$.




**(e)** $X_p - A_p - I_p$ at $\varpi = 19.8$.




**(f)** $X_p - A_p - I_p$ at $\varpi = 19.8$.

**Fig. 10** The time series solutions and respective phase portraits of susceptible, aware and infected subpopulations

## 6 Discussion

During an outbreak, awareness about the transmission routes and interventions of a disease can alert individuals regarding the infection risk, resulting in a change in human behavior and disease transmission patterns. Therefore, the present article studies a mathematical epidemic model with awareness effects to study disease transmission and control dynamics. We comprised four dynamical variables in our model: susceptible, aware, infected, and recovered individual; and proposed a nonlinear time-delayed SAIR epidemic model by incorporating Michaelis-Menten type incidences with latent period and a saturating treatment rate. We assume that aware individuals can also catch the infection, probably at a lower rate than the fully vulnerable population. We analyze the model mathematically, revealing that it has two equilibria: the disease-free equilibrium (complete eradication of infection) and the endemic equilibrium (persistence of disease at a certain level). We obtain the model's threshold quantity, the basic reproduction number $R_0$, and perform the stability analysis to determine whether the disease eliminates or persists. The basic reproduction number determines the potential for an infectious agent to start an outbreak, the degree of transmission without control measures, and the capacity of control measures to diminish spread. The delayed system analysis reveals that the disease-free equilibrium is locally asymptotically stable when $R_0 < 1$, unstable when $R_0 > 1$, and neutrally linearly stable when $R_0 = 1$. However, using the center manifold theory approach, we show that the undelayed system exhibits a forward bifurcation at $R_0 = 1$, meaning that reducing $R_0$ below unity is enough to eradicate society's infection. Further, we investigate the endemic equilibrium's local stability, and prove the existence of oscillating and periodic solutions near-endemic equilibrium through Hopf bifurcation, concerning time delay as the bifurcation parameter. Furthermore, the global stability behavior of the disease-free and endemic equilibria is also examined using the Lyapunov functionals by employing the Lyapunov method. It is shown that the disease-free equilibrium is globally asymptotically stable when $R_0 < 1$, and the endemic equilibrium is globally asymptotically stable when $R_0 > 1$.

The numerical simulations validate the effectiveness of theoretical findings and show the impact of the model's parameters. It is observed that the longer the time delay, the higher the number of individuals who catch the infection. The oscillatory solutions for various values of time delay establish Hopf bifurcations near-endemic equilibrium. Moreover, if the time delay crosses its critical value, then the trajectories of the solutions bifurcate from endemic equilibrium and destabilize the system. We show that the number of infected individuals is much higher in the SIR model (i.e., without awareness) in comparison to the number of infected in the SAIR model (i.e., with awareness). If susceptible individuals are aware of infection risk, they will be on high alert and choose not to go to crowded areas, avoid unnecessary contact with infected individuals, and implement other anti-epidemic inhibition measures which reduce the infection spread effectively. The numerical result shows the impact of saturating treatment rate, which reveals that adequate treatment availability is crucial in controlling infection spread. If the treatment facilities are not enough, individuals' awareness and their willingness to adopt anti-epidemic measures are the only way to

reduce infection. Individuals' awareness together with sufficient treatment facilities for infectives can reduce or even eradicate the infectious disease from society.

The present study consisting of nonlinear incidences of unaware and aware susceptibles with the latent period, and saturated treatment rate, signifies the substantial role of the latent period in the disease transmission process, susceptibles' behavior in preventing disease spread, and limitation in treatment facilities in curing infectives. The results indicate that awareness about the spread of infection in susceptible individuals is vital in preventing disease transmission and is a potential policy for controlling the disease spread in the absence of treatment availabilities. The public health authorities and the government have a significant contribution to raising awareness among people and encouraging them to adopt anti-epidemic measures. For example, the government is enforcing different non-pharmaceutical interventions to obstruct COVID-19 transmission due to the absence of proper therapeutics or vaccines. Several countries focus on raising awareness through media advertising campaigns to encourage people to maintain social distance, wear a face mask, adopt healthy sanitation practices, wear hand gloves, avoid touching surfaces, regular hand washing, etc. These behaviors urge people to adopt preventive measures, reduce contact with others, and so reduce disease spread, consequently suppressing disease spread burden. Thus, awareness about the disease with the encouragement of adopting preventive measures and appropriate treatment facilities for infectives can efficiently reduce disease spread.

**Acknowledgements** The authors are grateful to the Delhi Technological University, Delhi, India, for providing financial support to carry out this research work.

## Declarations

**Conflicts of interest** The authors declare that they have no conflict of interest.

## Appendix

### Proof of Theorem 9

***Proof*** (P1.) and (P2.) establish that $E_0(X_0, A_0, 0)$ is the only equilibrium of the system (6).
The Lyapunov functional is given as

$$V(t) = V_1(t) + V_2(t),$$

where,

$$V_1(t) = X_p(t) + A_p(t) - X_0 - A_0 - \int_{X_0}^{X_p(t)} \lim_{I_p \to 0^+} \frac{\Psi(X_0, I_p(t))}{\Psi(\varepsilon, I_p(t))}\, \mathrm{d}\varepsilon$$
$$- \int_{A_0}^{A_p(t)} \lim_{I_p \to 0^+} \frac{\Lambda(A_0, I_p(t))}{\Lambda(\varepsilon, I_p(t))}\, \mathrm{d}\varepsilon + I_p(t),$$
$$V_2(t) = \int_0^{\varpi} (\Psi(X_p(t-\rho), I_p(t-\rho)) + \Lambda(A_p(t-\rho), I_p(t-\rho)))\, \mathrm{d}\rho.$$

Using the postulates (P1.)–(P3.), it follows that $V_1(t)$ is well-defined and continuously differentiable function $\forall\ X_p(t) > 0,\ A_p(t) > 0,\ I_p(t) > 0$, and $V(t) = 0$ at $E_0(X_0, A_0, 0)$. Now, we show that $\frac{\mathrm{d}V(t)}{\mathrm{d}t} \leq 0$ for all $t \geq 0$. First, we compute $\frac{\mathrm{d}V_1(t)}{\mathrm{d}t}$ as follows:

$$\begin{aligned}
\frac{\mathrm{d}V_1(t)}{\mathrm{d}t} &= \left(1 - \lim_{I\to 0^+} \frac{\Psi(X_0, I(t))}{\Psi(X_p, I_p(t))}\right) X_p'(t) + \left(1 - \lim_{I_p\to 0^+} \frac{\Lambda(A_0, I_p(t))}{\Lambda(A_p, I_p(t))}\right) A_p'(t) + I_p'(t) \\
&= \left(1 - \lim_{I_p\to 0^+} \frac{\Psi(X_0, I_p(t))}{\Psi(X_p, I_p(t))}\right) \big(\kappa - \vartheta X_p - \Psi(X_p(t-\varpi), I_p(t-\varpi)) - \xi X_p\big) \\
&\quad + \left(1 - \lim_{I_p\to 0^+} \frac{\Lambda(A_0, I_p(t))}{\Lambda(A_p, I_p(t))}\right) \big(\xi X_p - \vartheta A_p - \Lambda(A_p(t-\varpi) I_p(t-\varpi))\big) \\
&\quad + \Psi(X_p(t-\varpi), I_p(t-\varpi)) \\
&\quad + \Lambda(A_p(t-\varpi) I_p(t-\varpi)) - (\vartheta + d + \theta) I_p - \frac{a I_p^2}{b I_p^2 + c I_p + 1}.
\end{aligned}$$

Since $\kappa - (\vartheta + \xi) X_p = (\vartheta + \xi)(X_0 - X_p)$, thus, we obtain:

$$\begin{aligned}
\frac{\mathrm{d}V_1(t)}{\mathrm{d}t} &= \left(1 - \lim_{I\to 0^+} \frac{\Psi(X_0, I_p(t))}{\Psi(X_p, I_p(t))}\right) \big((\vartheta + \xi)(X_0 - X_p) - \Psi(X_p(t-\varpi), I_p(t-\varpi))\big) \\
&\quad + \left(1 - \lim_{I_p\to 0^+} \frac{\Lambda(A_0, I_p(t))}{\Lambda(A_p, I_p(t))}\right) \big((\xi X_p - \vartheta A_p) - \Lambda(A_p(t-\varpi) I_p(t-\varpi))\big) \\
&\quad + \Psi(X_p(t-\varpi), I_p(t-\varpi)) \\
&\quad + \Lambda(A_p(t-\varpi) I_p(t-\varpi)) - H(I_p(t)).
\end{aligned}$$

We now obtain the derivative of $V_2(t)$ as below:

$$\begin{aligned}
\frac{\mathrm{d}V_2(t)}{\mathrm{d}t} &= -\Psi(X_p(t-\varpi), I_p(t-\varpi)) + \Psi(X_p(t), I_p(t)) \\
&\quad - \Lambda(A_p(t-\varpi), I_p(t-\varpi)) + \Lambda(A_p(t), I_p(t)).
\end{aligned}$$

Thus, the derivative of $V(t)$ is obtained as:

$$\begin{aligned}
\frac{\mathrm{d}V(t)}{\mathrm{d}t} &= \frac{\mathrm{d}V_1(t)}{\mathrm{d}t} + \frac{\mathrm{d}V_2(t)}{\mathrm{d}t} \\
&= \left(1 - \lim_{I_p\to 0^+} \frac{\Psi(X_0, I_p(t))}{\Psi(X_p, I_p(t))}\right) \big((\vartheta + \xi)(X_0 - X_p) - \Psi(X_p(t-\varpi), I_p(t-\varpi))\big) \\
&\quad + \left(1 - \lim_{I_p\to 0^+} \frac{\Lambda(A_0, I_p(t))}{\Lambda(A_p, I_p(t))}\right) \big((\xi X_p - \vartheta A_p) - \Lambda(A_p(t-\varpi) I_p(t-\varpi))\big) \\
&\quad + \Psi(X_p(t-\varpi), I_p(t-\varpi)) \\
&\quad + \Lambda(A_p(t-\varpi) I_p(t-\varpi)) - H(I_p(t)) - \Psi(X_p(t-\varpi), I_p(t-\varpi)) \\
&\quad + \Psi(X_p(t), I_p(t)) \\
&\quad - \Lambda(A_p(t-\varpi), I_p(t-\varpi)) + \Lambda(A_p(t), I_p(t)).
\end{aligned}$$

The postulates P4–P6 imply that

$$\begin{aligned}
\frac{\mathrm{d}V(t)}{\mathrm{d}t} &\le (\vartheta+\xi)\left(1-\frac{\Psi'_{I_p}(X_0,0)}{\Psi'_{I_p}(X_p(t),0)}\right)(X_0-X_p)\\
&\quad+\Psi(X_p(t-\varpi),I_p(t-\varpi))\left(\frac{\Psi'_{I_p}(X_0,0)}{\Psi'_{I_p}(X_p(t),0)}-1\right)\\
&\quad+\left(1-\frac{\Lambda'_{I_p}(A_0,0)}{\Lambda'_{I_p}(A_p(t),0)}\right)(\xi X_p-\vartheta A_p)\\
&\quad+\Lambda(A_p(t-\varpi),I_p(t-\varpi))\left(\frac{\Lambda'_{I_p}(A_0,0)}{\Lambda'_{I_p}(A_p(t),0)}-1\right)\\
&\quad+\frac{I_p(t)}{\vartheta+d+\theta}(R_0-1)\\
&=\vartheta\left(1-\frac{\Psi'_{I_p}(X_0,0)}{\Psi'_{I_p}(X_p(t),0)}\right)(X_0-X_p)+\xi X_0\left(1-\frac{X_p}{X_0}\right)\frac{A_0}{A_p}\frac{X_0}{X_p}\left(\frac{X_p}{X_0}-\frac{A_p}{A_0}\right)\\
&\quad+\Psi(X_p(t-\varpi),I_p(t-\varpi))\left(\frac{\Psi'_{I_p}(X_0,0)}{\Psi'_{I_p}(X_p(t),0)}-1\right)\\
&\quad+\vartheta\left(A_0-A_p\right)\left(1-\frac{\Lambda'_{I_p}(A_0,0)}{\Lambda'_{I_p}(A_p(t),0)}\right)\\
&\quad+\Lambda(A_p(t-\varpi),I_p(t-\varpi))\left(\frac{\Lambda'_{I_p}(A_0,0)}{\Lambda'_{I_p}(A_p(t),0)}-1\right)+\frac{I_p(t)}{\vartheta+d+\theta}(R_0-1).
\end{aligned}$$

Thus, $R_0 \le 1$ implies that $\frac{\mathrm{d}V(t)}{\mathrm{d}t} \le 0$ for all $t \ge 0$. Also, $\frac{\mathrm{d}V(t)}{\mathrm{d}t} = 0$ if $X_p(t) = X_0,\ A_p(t) = A_0,$ and $I_p(t) = 0$.
Hence, from the system (6), it follows that the largest invariant set $\left\{\left(X_p(t), A_p(t), I_p(t) \in R^3_+ \middle| \frac{\mathrm{d}V(t)}{\mathrm{d}t} = 0\right)\right\}$ is singleton set $\{E_0\}$. Using the Lyapunov-LaSalle asymptotic stability theorem [23, 26, 35], $E_0$ is the only equilibrium of the system (6) which is globally asymptotically stable. □

**Proof of Theorem 10**

***Proof*** For this, the Lyapunov functional is defined as

$$W(t) = W_1(t) + W_2(t),$$

where,

$$\begin{aligned}
W_1(t) &= X_p(t) - X_p^* - \int_{X_p^*}^{X_p(t)} \frac{\Psi(X_p^*, I_p^*)}{\Psi(\phi, I_p^*)}\,\mathrm{d}\phi + A_p(t) - A_p^*\\
&\quad - \int_{A_p^*}^{A_p(t)} \frac{\Lambda(A_p^*, I_p^*)}{\Lambda(\Lambda, I_p^*)}\,\mathrm{d}\phi + I_p(t) - I_p^* - I_p^* \ln\frac{I_p(t)}{I_p^*},\\
W_2(t) &= \Phi(X_p^*, I_p^*)\int_0^{\varpi}\left(\frac{\Phi(X_p(t-\theta), I_p(t-\theta))}{\Phi(X_p^*, I_p^*)} - 1 - \ln\frac{\Phi(X_p(t-\theta), I_p(t-\theta))}{\Phi(X_p^*, I^*)}\right)\mathrm{d}\theta
\end{aligned}$$

$$+\Lambda(A_p^*, I_p^*)\int_0^{\varpi}\left(\frac{\Lambda(A_p(t-\theta), I_p(t-\theta))}{\Lambda(A_p^*, I_p^*)} - 1 - \ln\frac{\Lambda(A_p(t-\theta), I_p(t-\theta))}{\Lambda(A_p^*, I_p^*)}\right)\mathrm{d}\theta.$$

(P1.)–(P3.) imply that, $W(t) = W_1(t) + W_2(t)$ is defined and continuously differentiable for all $X_p(t),\ A_p(t),\ I_p(t) > 0$ and $W(0) = 0$ at $E_e(X_p^*, A_p^*, I_p^*)$.
We compute the derivative of $W_1(t)$ as follows:

$$\begin{aligned}
\frac{\mathrm{d}W_1(t)}{dt} &= \left(1 - \frac{\Psi(X_p^*, I_p^*)}{\Psi(X_p(t), I_p^*)}\right)X_p'(t) + \left(1 - \frac{\Lambda(A_p^*, I^*)}{\Lambda(A_p(t), I_p^*)}\right)A_p'(t) + \left(1 - \frac{I_p^*}{I_p(t)}\right)I_p'(t)\\
&= \left(1 - \frac{\Psi(X_p^*, I_p^*)}{\Psi(X_p(t), I_p^*)}\right)\Big((\vartheta+\xi)(X_p^* - X_p(t))\\
&\quad + \Psi(X_p^*, I_p^*) - \Psi(X_p(t-\varpi), I_p(t-\varpi))\Big)\\
&\quad + \left(1 - \frac{\Lambda(A_p^*, I_p^*)}{\Lambda(A_p(t), I_p^*)}\right)\big(\xi X_p - \vartheta A_p - \Lambda(A_p(t-\varpi), I_p(t-\varpi))\big) + \left(1 - \frac{I_p^*}{I_p(t)}\right)\\
&\quad \times \Big(\Psi(X_p(t-\varpi), I_p(t-\varpi)) + \Lambda(A_p(t-\varpi), I_p(t-\varpi))\\
&\quad - \frac{\Big(\Psi(X_p^*, I_p^*) + \Lambda(A_p^*, I_p^*) - h(I_p^*)\Big)}{I_p^*}I_p(t) - h(I_p(t))\Big)\\
&= (\vartheta+\xi)\left(1 - \frac{\Psi(X_p^*, I_p^*)}{\Psi(X_p(t), I_p^*)}\right)(X_p^* - X_p(t)) + \Psi(X_p^*, I_p^*)\Big(1 - \frac{\Psi(X_p^*, I_p^*)}{\Psi(X_p(t), I_p^*)}\\
&\quad + \frac{\Psi(X_p(t-\varpi), I_p(t-\varpi))}{\Psi(X_p(t), I_p^*)}\Big) + \big(\xi X_p - \vartheta A_p\big)\left(1 - \frac{\Lambda(A_p^*, I_p^*)}{\Lambda(A_p(t), I_p^*)}\right)\\
&\quad - \Lambda(A_p(t-\varpi), I_p(t-\varpi))\frac{I_p^*}{I_p(t)}\\
&\quad + \frac{\Lambda(A_p(t-\varpi), I_p(t-\varpi))}{\Lambda(A_p(t), I_p^*)}\Lambda(A_p^*, I_p^*) - \frac{I_p^*}{I_p(t)}\Psi(X_p(t-\varpi), I_p(t-\varpi))\\
&\quad + \Psi(X_p^*, I_p^*)\left(1 - \frac{I_p}{I_p^*}\right)\\
&\quad + \Lambda(A_p^*, I_p^*)\left(1 - \frac{I_p}{I_p^*}\right) + \left(h(I_p^*) - \frac{h(I_p(t)I_p^*)}{I_p(t)}\right)\left(\frac{I_p(t)}{I_p^*} - 1\right).
\end{aligned}$$

Since $\xi X_p^* - \vartheta A_p^* - \Lambda(A_p^*, I_p^*) = 0$, or, $-\xi X_p^* + \vartheta A_p^* + \Lambda(A_p^*, I_p^*) = 0$, thus we have

$$\begin{aligned}
\frac{\mathrm{d}W_1(t)}{dt} &= (\vartheta+\xi)\left(1 - \frac{\Psi(X_p^*, I_p^*)}{\Psi(X_p(t), I_p^*)}\right)(X_p^* - X_p(t))\\
&\quad + \Psi(X_p^*, I_p^*)\left(1 - \frac{\Psi(X_p^*, I_p^*)}{\Psi(X_p(t), I_p^*)} + \frac{\Psi(X_p(t\varpi), I_p(t-\varpi))}{\Psi(X_p(t), I_p^*)}\right)\\
&\quad - \frac{I_p^*}{I_p(t)}\Psi(X_p(t-\varpi), I_p(t-\varpi)) + \Psi(X_p^*, I_p^*)\left(1 - \frac{I_p}{I_p^*}\right)
\end{aligned}$$

$$
\begin{aligned}
&+\xi(X_p - X_p^*)\left(1 - \frac{\Lambda(A_p^*, I_p^*)}{\Lambda(A_p(t), I_p^*)}\right)\\
&-\Lambda(A_p(t-\varpi), I_p(t-\varpi))\frac{I_p^*}{I_p(t)} + \Lambda(A_p^*, I_p^*)\left(1 - \frac{I_p}{I_p^*}\right)\\
&+\vartheta\left(A_p^* - A_p\right)\left(1 - \frac{\Lambda(A_p^*, I_p^*)}{\Lambda(A_p(t), I_p^*)}\right)\\
&+\Lambda(A_p^*, I_p^*)\left(1 - \frac{\Lambda(A_p^*, I_p^*)}{\Lambda(A_p(t), I_p^*)} + \frac{\Lambda(A_p(t-\varpi), I_p(t-\varpi))}{\Lambda(A_p(t), I_p^*)}\right).
\end{aligned}
$$

Now, we compute the derivative of $W_2(t)$ as follows:

$$
\begin{aligned}
\frac{\mathrm{d}W_2}{\mathrm{d}t} &= \Psi(X_p(t), I_p(t)) - \Psi(X_p(t-\varpi), I_p(t-\varpi))\\
&\quad+\Psi(X_p^*, I_p^*)\ln\frac{\Psi(X_p(t-\varpi), I_p(t-\varpi))}{\Psi(X_p(t), I_p(t))}\\
&\quad+\Lambda(A_p(t), I_p(t)) - \Lambda(A_p(t-\varpi), I_p(t-\varpi))\\
&\quad+\Lambda(A_p^*, I_p^*)\ln\frac{\Lambda(A_p(t-\varpi), I_p(t-\varpi))}{\Lambda(X_p(t), I_p(t))}.
\end{aligned}
$$

Thus, the derivative of $W(t)$ with respect to time $t$ is obtained as follows:

$$
\begin{aligned}
\frac{\mathrm{d}W}{\mathrm{d}t} &= \frac{\mathrm{d}W_1}{\mathrm{d}t} + \frac{\mathrm{d}W_2}{\mathrm{d}t}\\
&= (\vartheta + \xi)\left(1 - \frac{\Psi(X_p^*, I_p^*)}{\Psi(X_p(t), I_p^*)}\right)(X_p^* - X_p(t))\\
&\quad+\Psi(X_p^*, I_p^*)\left(1 - \frac{\Psi(X_p^*, I_p^*)}{\Psi(X_p(t), I_p^*)} + \frac{\Psi(X_p(t-\varpi), I_p(t-\varpi))}{\Psi(X_p(t), I_p^*)}\right)\\
&\quad-\frac{I_p^*}{I_p(t)}\Psi(X_p(t-\varpi), I_p(t-\varpi)) + \Psi(X_p^*, I_p^*)\left(1 - \frac{I_p}{I_p^*}\right)\\
&\quad+\xi(X_p - X_p^*)\left(1 - \frac{\Lambda(A_p^*, I_p^*)}{\Lambda(A_p(t), I_p^*)}\right)\\
&\quad-\Lambda(A_p(t-\varpi), I_p(t-\varpi))\frac{I_p^*}{I_p(t)} + \Lambda(A_p^*, I_p^*)\left(1 - \frac{I_p}{I_p^*}\right)\\
&\quad+\vartheta\left(A_p^* - A_p\right)\left(1 - \frac{\Lambda(A_p^*, I_p^*)}{\Lambda(A_p(t), I_p^*)}\right)\\
&\quad+\Lambda(A_p^*, I_p^*)\left(1 - \frac{\Lambda(A_p^*, I_p^*)}{\Lambda(A_p(t), I_p^*)} + \frac{\Lambda(A_p(t-\varpi), I_p(t-\varpi))}{\Lambda(A_p(t), I_p^*)}\right)
\end{aligned}
$$

$$
\begin{aligned}
&+ \Psi(X_p(t), I_p(t)) - \Psi(X_p(t-\varpi), I_p(t-\varpi)) \\
&+ \Psi(X_p^*, I_p^*) \ln \frac{\Psi(X_p(t-\varpi), I_p(t-\varpi))}{\Psi(X_p(t), I_p(t))} \\
&+ \Lambda(A_p(t), I_p(t)) - \Lambda(A_p(t-\varpi), I_p(t-\varpi)) \\
&+ \Lambda(A_p^*, I_p^*) \ln \frac{\Lambda(A_p(t-\varpi), I_p(t-\varpi))}{\Lambda(X_p(t), I_p(t))} \\
&= \vartheta(X_p^* - X_p)\left(1 - \frac{\Psi(X_p^*, I_p^*)}{\Psi(X_p(t), I_p^*)}\right) + \vartheta(A_p^* - A_p)\left(1 - \frac{\Lambda(A_p^*, I_p^*)}{\Lambda(A_p(t), I_p^*)}\right) \\
&+ \Psi(X_p^*, I_p^*)\left(1 - \frac{\Psi(X_p^*, I_p^*)}{\Psi(X_p(t), I_p^*)} + \ln \frac{\Psi(X_p^*, I_p^*)}{\Psi(X_p(t), I_p^*)}\right) \\
&+ \Psi(X_p^*, I_p^*)\left(1 - \frac{I_p}{I_p^*} + \ln \frac{I_p}{I_p^*}\right) \\
&+ \Lambda(A_p^*, I_p^*)\left(1 - \frac{\Lambda(A_p^*, I_p^*)}{\Lambda(A_p(t), I_p^*)} + \ln \frac{\Lambda(A_p^*, I_p^*)}{\Lambda(A_p(t), I_p^*)}\right) \\
&+ \Lambda(A_p^*, I_p^*)\left(1 - \frac{I_p}{I_p^*} + \ln \frac{I_p}{I_p^*}\right) \\
&+ \Psi(X_p^*, I_p^*)\Big(1 - \frac{\Psi(X_p(t-\varpi), I_p(t-\varpi))}{\Psi(X_p(t), I_p(t))} \frac{\Psi(X_p(t), I_p^*)}{\Psi(X_p^*, I_p^*)} \frac{I_p^*}{I_p(t)} \\
&+ \ln \frac{\Psi(X_p(t-\varpi), I_p(t-\varpi))}{\Psi(X_p(t), I_p(t))} \frac{\Psi(X_p(t), I_p^*)}{\Psi(X_p^*, I_p^*)} \frac{I_p^*}{I_p(t)}\Big) \\
&+ \Lambda(A_p^*, I_p^*)\Big(1 - \frac{\Lambda(A_p(t-\varpi), I_p(t-\varpi))}{\Lambda(A_p(t), I_p(t))} \frac{\Lambda(A_p(t), I_p^*)}{\Lambda(A_p^*, I_p^*)} \frac{I_p^*}{I_p(t)} \\
&+ \ln \frac{\Lambda(A_p(t-\varpi), I_p(t-\varpi))}{\Lambda(A_p(t), I_p(t))} \frac{\Lambda(A_p(t), I_p^*)}{\Lambda(A_p^*, I_p^*)} \frac{I_p^*}{I_p(t)}\Big) \\
&+ \Psi(X_p(t-\varpi), I_p(t-\varpi))\left(\frac{\Psi(X_p^*, I_p^*)}{\Psi(X_p(t), I_p^*)} - \frac{I_p^*}{I_p(t)}\right) \\
&+ \Psi(X_p^*, I_p^*)\left(\frac{\Psi(X_p(t), I_p(t))}{\Psi(X_p^*, I_p^*)} - 1\right) \\
&+ \Lambda(A_p^*, I_p^*)\left(\frac{\Lambda(A_p(t), I_p(t))}{\Lambda(A_p^*, I_p^*)} - 1\right) \\
&+ \Psi(X_p(t-\varpi), I_p(t-\varpi))\frac{I_p^*}{I_p(t)}\left(\frac{\Psi(X_p(t), I_p^*)}{\Psi(X_p(t), I_p(t))} - \frac{I_p(t)}{I_p^*}\right) \\
&+ \Lambda(A_p(t-\varpi), I_p(t-\varpi))\frac{I_p^*}{I_p(t)}\left(\frac{\Lambda(A_p(t), I_p^*)}{\Lambda(A_p(t), I_p(t))} - \frac{I_p(t)}{I_p^*}\right)
\end{aligned}
$$

$$+\Lambda(A_p(t-\varpi),I_p(t-\varpi))\Big(\frac{\Lambda(A_p^*,I_p^*)}{\Lambda(A_p(t),I_p^*)}-\frac{I_p^*}{I_p(t)}\Big)+\xi X_p^*\left(1-\frac{X_p}{X_p^*}\right)\left(\frac{X_p}{X_p^*}-\frac{A_p}{A_p^*}\right)\frac{A_p^*}{A_p}\frac{X_p^*}{X_p}+\left(\frac{h(I_p^*)}{h(I_p(t))}-\frac{I_p^*}{I_p(t)}\right)\left(\frac{I_p(t)}{I_p^*}-1\right)h(I_p(t)).$$

The functions $\Psi(X_p(t),I_p(t))$ and $\Lambda(A_p(t),I_p(t))$ are monotonically increasing for all $X_p(t)>0$, and $A_p(t)>0$. Therefore,

$$\begin{aligned}\left(X_p^*-X_p\right)\left(1-\frac{\Psi(X_p^*,I_p^*)}{\Psi(X_p(t),I_p^*)}\right)\le 0,\\ \left(A_p^*-A_p\right)\left(1-\frac{\Lambda(A_p^*,I_p^*)}{\Lambda(A_p(t),I_p^*)}\right)\le 0.\end{aligned}\tag{39}$$

The function $g(y)=1-y+\ln y$, $(y>0)$ has global maximum at $y=1$. Henceforth, for $y>0$, $g(y)\le 0$ and the resulting inequalities are as follows:

$$\begin{aligned}\left(1-\frac{\Psi(X_p^*,I_p^*)}{\Psi(X_p(t),I_p^*)}+\ln\frac{\Psi(X_p^*,I_p^*)}{\Psi(X_p(t),I_p^*)}\right)\le 0,\\ \left(1-\frac{I_p}{I_p^*}+\ln\frac{I_p}{I_p^*}\right)\le 0,\\ \left(1-\frac{\Lambda(A_p^*,I_p^*)}{\Lambda(A_p(t),I_p^*)}+\ln\frac{\Lambda(A_p^*,I_p^*)}{\Lambda(A_p(t),I_p^*)}\right)\le 0,\\ \left(1-\frac{\Psi(X_p(t-\varpi),I_p(t-\varpi))}{\Psi(X_p(t),I_p(t))}\frac{\Psi(X_p(t),I_p^*)}{\Psi(X_p^*,I_p^*)}\frac{I_p^*}{I_p(t)}+\ln\frac{\Psi(X_p(t-\varpi),I_p(t-\varpi))}{\Psi(X_p(t),I_p(t))}\frac{\Psi(X_p(t),I_p^*)}{\Psi(X_p^*,I_p^*)}\frac{I_p^*}{I_p(t)}\right)\le 0,\\ \left(1-\frac{\Lambda(A_p(t-\varpi),I_p(t-\varpi))}{\Lambda(A_p(t),I_p(t))}\frac{\Lambda(A_p(t),I_p^*)}{\Lambda(A_p^*,I_p^*)}\frac{I_p^*}{I_p(t)}+\ln\frac{\Lambda(A_p(t-\varpi),I_p(t-\varpi))}{\Lambda(A_p(t),I_p(t))}\frac{\Lambda(A_p(t),I_p^*)}{\Lambda(A_p^*,I_p^*)}\frac{I_p^*}{I_p(t)}\right)\le 0.\end{aligned}\tag{40}$$

Thus, using (P8.)–(P10.) and the inequalities (39)–(40), it follows that $\frac{\mathrm{d}W(t)}{\mathrm{d}t}\le 0$ for all $X_p(t)\ge 0$, $A_p(t)\ge 0$, $I_p(t)\ge 0$. Thus, it is easy to verify that the largest invariant set in $\left\{(X_p(t),A_p(t),I_p(t))\in R_+^3:\frac{\mathrm{d}W(t)}{\mathrm{d}t}=0\right\}$ is singleton $\{E_e\}$. Hence, by the

Lyapunov-LaSalle asymptotic stability theorem [23, 26, 35], the endemic equilibrium $E_e$ is globally asymptotically stable. □

## References


1. Ahituv, A., Hotz, V.J., Philipson, T.: The responsiveness of the demand for condoms to the local prevalence of AIDS. The Journal of Human Resource **31**(4), 869–897 (1996)
2. Anderson, R.M., May, R.M.: Regulation and stability of host-parasite population interactions: I. regulatory processes. J. Anim. Ecol. **47**(1), 219–267 (1978)
3. Brauer, F., van den Driessche, P.: Jian Hong Wu (Eds.). Mathematical Epidemiology, Mathematical Biosciences Subseries, Springer-Verlag Berlin Heidelberg, 1945 (2008)
4. Brauer, F., Castillo-Chavez, C.: Mathematical models in population biology and epidemiology. Springer, New York (2001)
5. Buonomo, B., Lacitignola, D.: On the backward bifurcation of a vaccination model with nonlinear incidence. Nonlinear Analysis: Modelling and Control **16**(1), 30–46 (2011)
6. Capasso, V., Serio, G.: A generalization of the Kermack-Mckendrick deterministic epidemic model. Math. Biosci. **42**(1–2), 43–61 (1978)
7. Castillo-Chavez, C., Song, B.: Dynamical models of tuberculosis and their applications. Math. Biosci. Eng. **1**(2), 361–404 (2004)
8. Cui, J., Tao, X., Zhu, H.: An SIS infection model incorporating media coverage. Rocky Mountain Journal of Mathematics **38**(5), 1323–1334 (2008)
9. Cui, J., Sun, Y., Zhu, H.: The impact of media on the control of infectious disease. J. Dyn. Diff. Equat. **20**(1), 31–53 (2008)
10. Das, D.K., Khajanchi, S., Kar, T.K.: The impact of the media awareness and optimal strategy on the prevalence of tuberculosis. Appl. Math. Comput. **366**(2), 124732 (2019)
11. Dubey, B., Patra, A., Srivastava, P.K., Dubey, U.S.: Modeling and analysis of an SEIR model with different types of nonlinear treatment rates. Journal of Biological System **21**(03), 1350023 (2013)
12. Dubey, B., Dubey, P., Dubey, U.S.: Role of media and treatment on an SIR model. Nonlinear Analysis: Modelling and Control **21**(2), 185–200 (2015)
13. Dubey, B., Dubey, P., Dubey, U.S.: Dynamics of an SIR Model with Nonlinear Incidence and Treatment Rate. Appl. Appl. Math. **10**(2), 718–737 (2015)
14. Dubey, P., Dubey, B., Dubey, U.S.: An SIR Model with Nonlinear Incidence Rate and Holling Type III Treatment Rate. Applied Analysis in Biological and Physical Sciences **186**, 63–81 (2016)
15. Faiz, M.A., Alamgir, A.: Chikungunya Outbreak in Dhaka: Lessons for Bangladesh. Journal of Bangladesh College of Physicians and Surgeons **35**(3), 108–109 (2017)
16. Funk, S., Gilad, E., Watkins, C., Jansen, V.A.A.: The spread of awareness and its impact on epidemic outbreaks. Proc. Natl. Acad. Sci. U.S.A. **106**(16), 6872–6877 (2009)
17. Goel, K., Kumar, A., Nilam: A deterministic time-delayed SVIRS epidemic model with incidences and saturated treatment. Journal of Engineering Mathematics **121**, 19–38 (2020)
18. Goel, K., Kumar, A., Nilam: Nonlinear dynamics of a time-delayed epidemic model with two explicit aware classes, saturated incidences, and treatment. Nonlinear Dynamics **101**, 1693–1715 (2020)
19. Goel, K., Nilam: A mathematical and numerical study of a SIR epidemic model with time delay, nonlinear incidence and treatment rates. Theory in Biosciences **138**(2), 203–213 (2019)
20. Goel, K., Nilam: Stability behavior of a nonlinear mathematical epidemic transmission model with time delay. Nonlinear Dynamics **98**, 1501–1518 (2019)
21. Guckenheimer, J., Holmes, P.: Nonlinear oscillations, dynamical systems and bifurcations of vector fields. Springer-Verlag, Berlin (1983)
22. Gumel, A.B.: Causes of backward bifurcations in some epidemiological model. J. Math. Anal. Appl. **395**(1), 355–365 (2012)
23. Hale, J.K., Lunel, S.M.V.: Introduction to functional differential equations. Springer, New York (1993)
24. Hattaf, K., Lashari, A.A., Louartassi, Y., Yousfi, N.: A delayed SIR epidemic model with general incidence rate. Electron. J. Qual. Theory Differ. Equ. **3**(3), 1–9 (2013)
25. Hethcote, H.W., van den Driessche, P.: An SIS epidemic model with variable population size and a delay. J. Math. Biol. **34**, 177–194 (1995)

26. Huang, G., Takeuchi, Y., Ma, W.: Lyapunov functionals for delay differential equations model of viral infections. SIAM J. Appl. Math. **70**(7), 2693–2708 (2010)
27. Just, W., Saldana, J., Xin, Y.: Oscillations in epidemic models with spread of awareness. Mathematical Biology **76**, 1027–1057 (2018)
28. Kermack, W.O., McKendrick, A.G.: A contribution to the mathematical theory of epidemics. Proceedings of Royal Society of London **115**(772), 700–721 (1927)
29. Kumar, A., Goel, K., Nilam: A deterministic time-delayed SIR epidemic model: mathematical modeling and analysis. Theory in Biosciences, **139**, 67–76 (2020)
30. Kumar, A., Nilam, Kishor, R.: A short study of an SIR model with inclusion of an alert class, two explicit nonlinear incidence rates and saturated treatment rate. SeMA Journal, **76**, 505–519 (2019)
31. Kumar, A. Nilam.: Stability of a delayed SIR epidemic model by introducing two explicit treatment classes along with nonlinear incidence rate and Holling type treatment. Computational and Applied Mathematics, **38**, 130 (2019)
32. Kumar, A.: Stability of a Fractional-Order Epidemic Model with Nonlinear Incidences and Treatment Rates. Iranian Journal of Science and Technology, Transactions A: Science **44**, 1505–1517 (2020)
33. Lacitignola, D., Diele, F.: Using awareness to Z-control a SEIR model with overexposure: Insights on Covid-19 pandemic. Chaos, Solitons Fractals **150**, 111063 (2021)
34. Lacitignola, D., Saccomand, G.: Managing awareness can avoid hysteresis in disease spread: an application to coronavirus Covid-19. Chaos, Solitons Fractals **144**, 110739 (2021)
35. Li, M., Liu, X.: An SIR epidemic model with time delay and general nonlinear incidence rate. Abstr. Appl. Anal. **2014**, 131257 (2014)
36. Li, X.Z., Li, W.S., Ghosh, M.: Stability and Bifurcation of an SIR Epidemic Model with Nonlinear Incidence and Treatment. Appl. Math. Comput. **210**(1), 141–150 (2009)
37. Liu, Y., Cui, J.: The impact of media convergence on the dynamics of infectious diseases. Int. J. Biomath. **1**(1), 65–74 (2008)
38. Liu, R., Wu, J., Zhu, H.: Media/psychological impact on multiple outbreaks of emerging infectious diseases. Comput. Math. Methods Med. **8**(3), 153–164 (2007)
39. Misra, A.K., Sharma, A., Singh, V.: Effect of awareness programs in controlling the prevalence of an epidemic with time delay. J. Biol. Syst. **19**(2), 389–402 (2011)
40. Pandit, N., Patel, Y., Bhavsar, B.: Awareness and practice about preventive method against mosquito bite in Gujarat. Healthline **1**, 16–20 (2010)
41. Song, X., Cheng, S.: A delay-differential equation model of HIV infection of CD4+ T-cells. Journal of the Korean Mathematical Society **42**(5), 1071–1086 (2005)
42. Tai, Z., Sun, T.: Media dependencies in a changing media environment: The case of the 2003 SARS epidemic in China. New Media Soc. **9**(6), 987–1009 (2007)
43. van den Driessche, P., Watmough, J.: Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. **180**(1–2), 29–48 (2002)
44. Wang, X.: A simple proof of Descartes's rule of signs. Am. Math. Mon. **111**(6), 525–526 (2004)
45. Wang, W.: Backward bifurcation of an epidemic model with treatment. Math. Biosci. **201**(1–2), 58–71 (2006)
46. Wang, W., Ruan, S.: Bifurcations in an epidemic model with constant removal rate of the infectives. J. Math. Anal. Appl. **291**, 775–793 (2004)
47. Wang, W., Ruan, S.: Bifurcation in an epidemic model with constant removal rates of the infectives. J. Math. Anal. Appl. **21**, 775–793 (2004)
48. Wei, C., Chen, L.: A delayed epidemic model with pulse vaccination. Discret. Dyn. Nat. Soc. **2008**, 746951 (2008)
49. World Health Organization: Consensus document on the epidemiology of SARS. World Health Organization, Geneva, Switzerland (2003)
50. Xu, R., Ma, Z.: Stability of a delayed SIRS epidemic model with a nonlinear incidence rate. Chaos, Solitons & Fractals **41**(5), 2319–2325 (2009)
51. Zhang, J.Z., Jin, Z., Liu, Q.X., Zhang, Z.Y.: Analysis of a Delayed SIR Model with Nonlinear Incidence Rate. Discret. Dyn. Nat. Soc. **2008**, 636153 (2008)


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

Defence Science Journal, Vol. 72, No. 3, May 2022, pp. 371-381, DOI : 10.14429/dsj.72.3.17522
© 2022, DESIDOC

# A Novel Traffic Based Framework for Smartphone Security Analysis

Sumit Kumar,[#,*] S. Indu[$] and Gurjit Singh Walia[#]

[#]*DRDO-Scientific Analysis Group (SAG), Delhi - 110 054, India*
[$]*Electronics and Communication, Delhi Technological University, Delhi - 110 042, India*
[*]*E-mail: sumitkr@hotmail.com*

**ABSTRACT**

Android Operating system (OS) has grown into the most predominant smartphone platform due to its flexibility and open source characteristics. Because of its openness, it has become prone to numerous attackers and malware designers who are constantly trying to elicit confidential information by articulating a plethora of attacks through these designed malwares. Detection of these malwares to protect the smartphone is the core function of the smartphone security analysis. This paper proposes a novel traffic-based framework that exploits the network traffic features to detect these malwares. Here, a unified feature (UF) is created by graph-based cross-diffusion of generated order and sparse matrices corresponding to the network traffic features. Generated unified feature is then given to three classifiers to get corresponding classifier scores. The robustness of the suggested framework when evaluated on the standard datasets outperforms contemporary techniques to achieve an average accuracy of 98.74 per cent.

**Keywords:** Smartphone; Malware; Network traffic; Android; Fusion

## 1. INTRODUCTION

Smartphones are replacing conventional mobiles as well as computational devices due to their portability and ease of handling almost everything ranging from storing private data to making banking transactions. Smartphones having Android OS are extensively familiar and have wide usage due to their open architecture and the assortment of apps it affords. As per the recent information by IDC[1] (International Data Corporation), the market share of android smartphones is 83.8 per cent till March 2021 and it will grow to 85 per cent by March'2025. Due to its widespread usage, we are deluged with a variety of smartphone apps that makes our life simple and easier. Flooding of these apps tempts attackers to design a variety of malapp (malicious applications) which are directed toward smartphones to steal vital private information, acquire root privileges, build botnets to attack networks, or extract critical data from smartphones. These malwares after stealing the critical data sent it to the remote servers controlled by hackers. This paper aims at identifying the existence of those malapps which are distantly controlled by hackers via remote servers to collect the critical data. The network traffic-based framework extracts required features that are fused resulting in a unified feature. The unified feature is further given to the optimal classifier for the detection of malwares. This n/w traffic analysis-based framework aids in creating robust solutions for identifying android malwares.

These malapps are the most deterrent to the security of these devices. Smartphone security refers to the measures taken to thwart these malwares to protect sensitive information stored in or communicated by smartphones. The security analysis of smartphones mainly deals with findings of the vulnerabilities and intimidations while using smartphone applications.

Mobile devices including Smartphones generated about 54.8 per cent of worldwide web traffic[2] and analyzing this traffic leads to incredible results in detecting malapps. Analyzing traffic[3,4] is accomplished by studying the patterns in the network traffic for its identification and segregation for further investigation. Numerous traffic features are extracted from the network traffic patterns. Mainly Hypertext Transfer Protocol (HTTP) and Transmission Control Protocol (TCP) are two types of traffic that are prevalent in the smartphone ecosystem. Features extracted from HTTP and TCP are exploited in detecting the malapps. The HTTP header features could not detect the malapps in the encrypted traffic and TCP-based detection models are impervious to encrypted traffic. Therefore, TCP flow-based detection methods are mainly exploited in detecting malapps.

Two widely used malware detection methods employed by researchers pivots around static[5] and dynamic analysis[6-8]. The amalgamation of these widely used detection methods is also exploited by some researcher's resulting in hybrid analysis[9]. Static investigation-based detection techniques failed to detect apps having code obfuscation, and conventional dynamic investigation-based detection needs are quite cumbersome. N/w traffic based dynamic detection excerpts detection attributes from n/w traffic and uses Machine Learning (ML) techniques to categorize mobile apps. Our detection prototype is based on TCP-based features.

In short, the following are the main contributions to the manuscript:

Received : 24 September 2021, Revised : 30 March 2022
Accepted : 11 May 2022, Online published : 01 July 2022

- Proposed a traffic feature-based fusion that comprises of optimal combination of multiple traffic features by cross-diffusion of order and sparse graphs to produce a unified feature.
- The unified feature vector thus generated is given to the three parallel ML classifiers and classifiers scores obtained are fused to enhance the accuracy attained by separate classifiers.
- Presented the performance comparison with existing state-of-the-art methods using standard data sets available.

In Section 2, related work is presented. In Sec 3, we cover the proposed identification framework for malware identification and the assessment of our technique is presented in Sec 4. In Sec 5, we conclude the paper.

## 2. RELATED WORK

Arora[10], *et al.* examine the TCP-based features based on traffic to shape the classifier for Android malware with more than 90 per cent of detection accuracy. Arora[11] *et al.* came up with a hybrid model named NTPDroid (**N**etwork **T**raffic and **P**ermissions based An**droid** malapp detection framework), that uses permissions and traffic features from the apps and exploits a Frequent Pattern Growth algo to generate frequent patterns of permissions and traffic features to achieve detection accuracy of 94 per cent. Wang[12] *et al.* proposed an efficient malware detection technique by using text semantics of n/w traffic by studying each HTTP flow. These HTTP packets were further processed by NLP (Natural Language Processing) to take out text-level features achieving an accuracy of 99.15 per cent but the method achieves 54.81 per cent for unknown apps in the wild.

Liu[13] *et al.* proposed a malware detection technique built on TCP n/w traffic, where network traffic generated by apps gets a greater number of a TCP flow to extract packet sizes as features. Results achieve 97 per cent of detection accuracy. Ding Li[14] *et al.* introduced a framework named DroidClassifier for the identification of HTTP header fields of n/w traffic created by malapps by using a supervised method to train the malware dataset. Moreover, Clustering is also used to increase the classification efficiency. The results achieve 90 per cent of detection accuracy. Li[15] *et al.* proposed a multilevel detection system named MulAV, in which it obtains info from n/w traffic, App's source code, and geospatial info where n/w traffic is collected by TCPdump Tool. The info is further fed to the ML method to train the model which identifies malapps. The result achieves a detection rate of 97.8 per cent. Wang[16] *et al.* discussed a technique to parse the HTTP packets of n/w traffic where features analysed are packet avg. length, number of upload and download packets, distribution of packet size, etc. Features were further extracted to obtain the pure malicious traffic dataset and this is used to detect malwares. Su[17] *et al.* presented an Android detection method that uses TCP-based behavioral characteristics to detect malapps. Here capturing of n/w traffic is done using NTM [network traffic monitor] tool and training are done via n/w traffic classifier. Results achieve 99.2 per cent and 94.2 per cent of detection accuracy by using Random forest and J48 classifier.

Zulkifli[18], *et al.* proposed a detection process based on n/w traffic which registers the app behavior and considered 7 TCP-based n/w traffic features from Contagio dumpset and Drebin dataset in which Drebin dataset achieved 98.4 per cent of detection accuracy on J48 decision tree algo. Malik[19], *et al.* proposed a pattern-based detection method CREDROID which identifies malapps based on the DNS (Domain Name Service) queries, data it transfers to the remote server from n/w traffic logs, and also the protocol used for communication for identifying the credibility of the app. Moreover, the Android app can be checked without rooting the android phone. Wang[20] et.al. proposed a malapp detection framework exploiting the URLs (Uniform Resource Locator) visited by them. Here the malapp detection model is based on a multi-view neural network with a detection accuracy of 98.35. Multiple views maintain copious semantic info from inputs for segregating the apps. Wang[21], *et al.* suggested a framework for android malapp identification leveraging both the TCP and HTTPS features. Here, the app detection was done on the server-side without affecting the user experience. C4.5 ML algo is used to train the model with 8312 benign and 5560 malign apps for identifying unknown apps with an accuracy of 97.89 per cent. Sanz[22], *et al.* offered a lightweight malapp detection framework using TCP-based network features with an accuracy of 90 per cent and a false-positive rate of less than 3 per cent. Here, the total number of 359 malapp and benign apps are used along with two Random forests and AdaBoost ML algorithms.

Upadhayay[23] et.al. suggested a hybrid-based malware detection model using network traffic and permissions with a higher frequency of occurrence to achieve the detection accuracy of 95.96 per cent. Alshehri[24], *et al.* proposed an innovative method to detect the repackaged apps by investigating the network traffic behavior of smartphones. Here authors exploited the request traffic generated by the apps. A total number of 8645 applications were used for experimentation. Here the accuracy of request flows attained is 95.1 per cent an improvement of 18.3 per cent of accuracy when compared with contemporary methods. Sihag[25], *et al.* proposed network packet-based investigations of captured traffic of the smartphone. Here, the authors represent the captured network packet interactions as images. These images were given to CNN (Convolution Neural Network) to achieve a detection accuracy of 99.12 per cent. Norouzian[26], *et al.* offered a hybrid detection method based on network traffic flow and static features. Here network flow features are combined with static graph vectors to detect malapps with 97 per cent accuracy.

## 3. IDENTIFICATION FRAMEWORK PROPOSED

The proposed traffic analysis framework for smartphone security analysis is elucidated in Fig. 1. The framework consists of four blocks viz. traffic feature fusion, classifier score-fusion, decision criteria, and reference apps update to accomplish efficient malapp detection. Extracted traffic features are converted into three traffic feature vectors. All the three traffic feature vectors are used for constructing similarity graphs by the means of cosine similarity. Similarity graphs are

again converted to normalised graphs by using the anchored normalisation technique. Normalised graphs are again used to form the sparse and order graphs. These obtained sparse and order graphs are further cross diffused to generate three fused traffic feature vectors. The three fused feature vectors are further concatenated to form the highly distinct unified feature vector. This distinct unified feature is discriminatory and given to three classifiers. The scores obtained from these classifiers are again optimally combined to classify a given test app.

Because of the numerous attributes encapsulated in the android based smartphone apps, individual ML algo shows its incompetence to categorize these apps accurately. To achieve overall detection accuracy, the framework uses three classifiers viz. Random Forest (RF), k-Nearest Neighbor (KNN), and Support Vector Machine (SVM) for app categorisation. RF performs superbly when the dataset is large and it is not susceptible to outliers. SVM performs better in the limited dataset and it is optimal for binary classification. If there is no training period, then KNN performs best. The framework leverages modified PCR-5 rules for score fusion. Finally, in decision block score $w_{PCR5}$ is matched with the threshold $w_{th}$ and a particular test app is categorised as benign if $w_{PCR5} \geq w_{th}$ or otherwise malign.

### 3.1 Traffic Feature Fusion

It is apparent from Fig. 1 that traffic feature fusion comprises feature extraction followed by similarity, normalised, sparse, and order graph generation. The generated sparse and order graphs are further cross diffused giving unified feature vector U as output.

#### *3.1.1 Traffic Feature Extraction*

The traffic gathering platform is used to collect the malign and benign traffic data produced by malign and benign apps, respectively. A firewall is installed on the platform to guarantee its security. Figure 2 shows our methodology for traffic gathering. The traffic gathering platform comprises four constituents, i.e., the control server, traffic collection module, app repository having downloaded malign and benign apps, and TCP traffic storage module containing only filtered TCP flows. These four components converse with the aid of a LAN switch. The control server is controlling the traffic gathering job in the platform by assigning the job to the different modules. The apps from the apps repository are directed to the traffic collection module, where the android virtual machine (AVM) is used to run the apps and collect the corresponding traffic. The collected traffic is further directed to the TCP traffic storage module, where only TCP flows are stored and the rest of the traffic is filtered out. Here android emulators are used to install and running apps on AVM. AVM comprises of packed android s/w stack and it runs just like a physical smartphone. Apps are run on an emulator. An emulator is restarted to fuel the malign apps to generate malicious behavior in the network traffic. A script in python was written to extract the features from the TCP flows. The extracted TCP features used are tabulated in Table 1.

Fifteen traffic features were extracted for a test app ***t*** along with N apps from the reference dictionary, $d \in \{D^+, D^-\}$, $D^+$ which comprises of benign apps and $D^-$ comprises of the malign app. The traffic feature extraction procedure is shown in Fig. 2. Update in Fig.1 shows that the reference dictionary

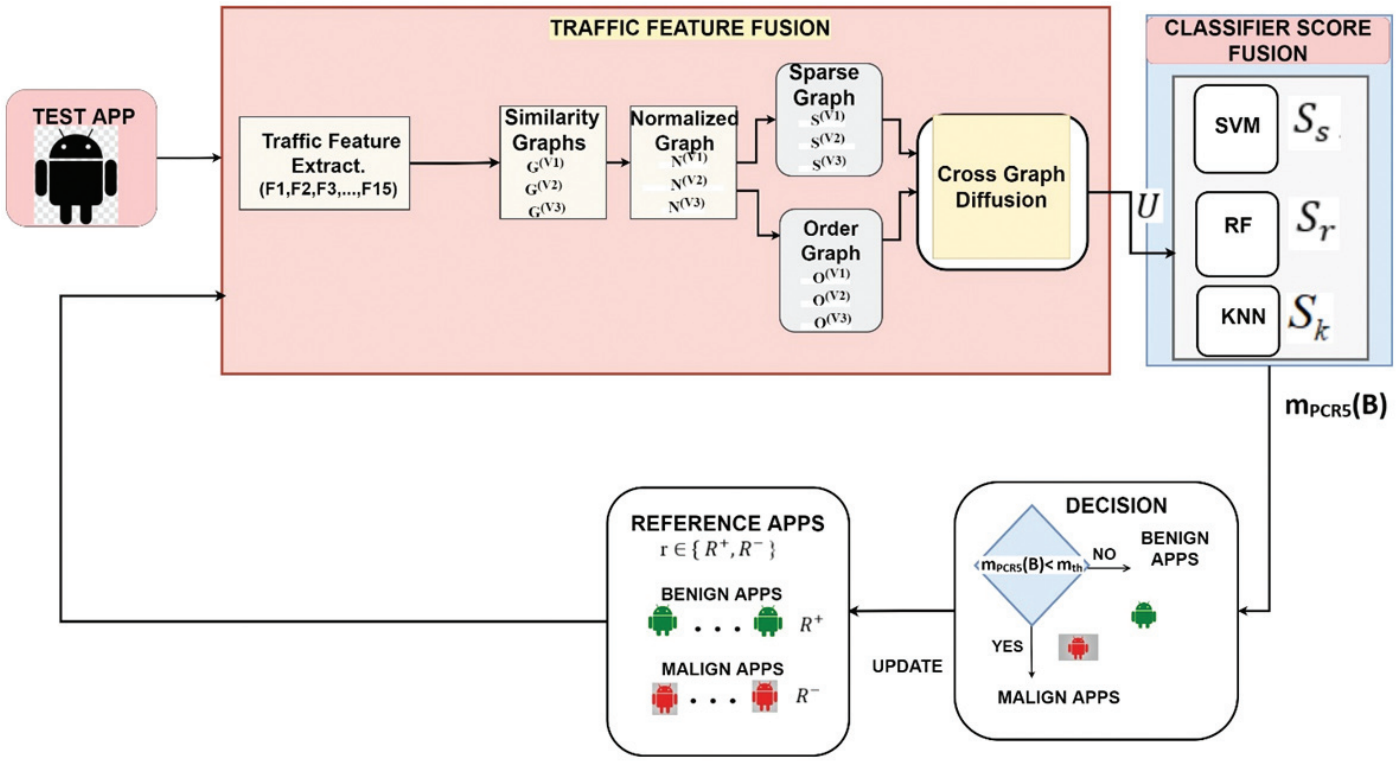



**Figure 1. Proposed traffic-based framework for smartphone security analysis.**

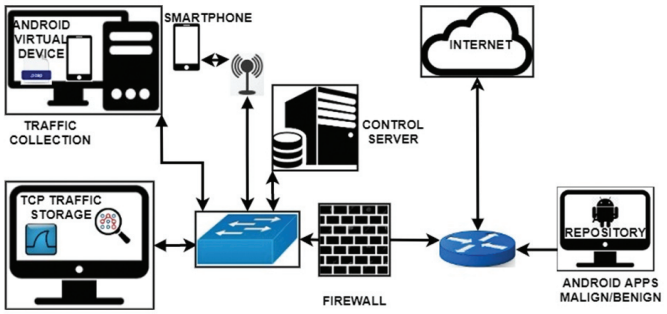



**Figure 2. Traffic gathering platform.**

**Table 1. Extracted TCP based traffic features**

| Feature symbol | Feature description |
|---|---|
| $F1$ | Avg. no. of bytes sent |
| $F2$ | Avg. no. of bytes received |
| $F3$ | Total no. of headers bytes sent |
| $F4$ | Avg. no. of bytes per second |
| $F5$ | The ratio of the no. of incoming to outgoing bytes |
| $F6$ | Avg. no. of the packet sent per second |
| $F7$ | Avg. no. of the packet received per second |
| $F8$ | The ratio of the no. of incoming to no. of outgoing packets |
| $F9$ | Std. deviation of the packet- size sent |
| $F10$ | The standard deviation of the packet- size received |
| $F11$ | Avg. no. of the packet sent per-flow |
| $F12$ | Avg. no. of the packet received per-flow |
| $F13$ | Avg. no. of bytes sent per-flow |
| $F14$ | Avg. no. of bytes received per-flow |
| $F15$ | Std. deviation of the length of the flow |

apps are constantly updated with the most recent malign and benign apps to improve the detection ability of the framework. Here, only TCP packets and TCP-related traffic features are considered. Traffic features exploited in our framework are depicted in Table 1. From the above 15 traffic features, we form three vectors such that each vector complements the other in detecting the malicious app. The vectors formed are as follows:

$$V^1 = \{F1, F2, F3, F4, F5\} \quad (1)$$

$$V^2 = \{F6, F7, F8, F9, F10\} \quad (2)$$

$$V^3 = \{F11, F12, F13, F14, F15\} \quad (3)$$

The proposed solution is realised by the formation of the three complementary traffic feature vectors, namely $V^1$ (byte-based features), $V^2$ (packet-based features), and $V^3$ (flow-based features). We have built three feature vectors as stated in Eqn. (1) - Eqn. (3) for each test app and reference app. In feature-fusion, traffic-features vectors for reference and test apps are utilised for graph formation. The test app's traffic-feature vectors represent one node and the reference app's traffic-feature vector represents other nodes. Consequently, non-linear graphs are formed for all test app ***t*** corresponding to three traffic feature vectors.

### *3.1.2 Graph Generation*

This subsection mainly comprises the generation of similarity, normalised, sparse, and order graphs. For traffic feature vectors, $V_t^1$, $V_t^2$ and $V_t^3$ of test app ***t*** corresponding to three traffic-based features, we construct graphs $G^\phi = (Ver^\phi, E^\phi, w^\phi)$, where $\phi \in \{1,2,3\}$, corresponding to three traffic-based features and $w^\phi$ are edge weights acting as similarity betwixt traffic-feature vectors of apps ***t*** *d* and where $d \in \{D^+, D^-\}$, $Ver^\phi$ correlates to the vertices of the created similarity graphs, $E^\phi$ correlates to the edges of the similarity graphs that characterize the association between test apps and the reference apps. In the suggested framework, similarity matrices $G^\phi \in \mathbb{R}^{n \times n}$ are constructed by calculating the cosine similarity between the three traffic-feature vectors of the test app $t$ and reference apps $d$. For feature set values $\left(V_t^\phi, V_d^\phi\right)$, where $\phi \in \{1,2,3\}$ corresponds to three traffic-feature vectors, the similarity edge weights are symbolised by the vector $w^\phi(t,r)$ and are derived by the cosine similarity between the pair $\left(V_t^\phi, V_d^\phi\right)$ from the following Eqn. (4).

$$w^\phi(t,d) = \frac{V_t^\phi * V_d^\phi}{\left\|V_t^\phi\right\| \left\|V_d^\phi\right\|} \quad (4)$$

Similarity graphs created using Eq.(4) for three traffic feature vectors are further normalised using an anchor $A^\phi = A_{avg}^\phi + A_{std}^\phi$, where $A_{avg}^\phi$ $A_{std}^\phi$ and are average and standard deviation of malicious score distribution for $\phi$ traffic feature vector. We consider only malicious scores for calculating the anchor. The normalised graphs $N^\phi$, whose edge weights matrix $\bar{w}_j^\phi$, where $\forall j \in \{1,2,.....,n\}, \forall \phi \in \{1,2,3\}$ is constructed by Eqn. (5),

$$\bar{w}_j^\phi = \begin{cases} \dfrac{w_j^\phi - \min(w^\phi)}{2(A^\phi - \min(w^\phi))}, w_j^\phi \le A^\phi \\ 0.5 + \left( \dfrac{w_j^\phi - A^\phi}{\max(w_j^\phi) - A^\phi} \right), w_j^\phi > A^\phi \end{cases} \quad (5)$$

The obtained normalised graphs were transformed to obtain the sparse $S^\phi$ and order $O^\phi$ graphs. Sparse graphs guarantee robustness to noise while boosting the strong info and suppressing the weak info corresponding to each traffic feature vector. A sparse graph $S^\phi = \{Ver^\phi, E^\phi, \eta^\phi\}$ is built using KNN by Eqn. (6).

$$\eta^\phi = \begin{cases} \bar{w}_j^\phi, \left(\bar{w}_j^\phi \in KNN_{N^\phi} \mid \nu\right) \\ 0, Otherwise \end{cases} \quad (6)$$

$\nu$ is the parameter controlling the sparseness of the graph. Edges corresponding to reference apps that are similar to the test apps are retained and the rest of the edges are removed to ensure robustness to noise. To discriminate the significant reference apps from the insignificant ones, each of the reference apps is assigned a weight according to its order. Therefore, order(s) is

assigned to each reference app based on its similarity with the test app, which is calculated by the edge weight between the test app and the reference app in the normalised graph. The order graph $O^{\phi} = \{Ver^{\phi}, E^{\phi}, \mu^{\phi}\}$ having the weight matrix is constructed by Eqn.(7) :

$$\mu^{\phi} = order(\overline{w_j}^{\phi}), \forall j \in \{1,2,...n\} \tag{7}$$

Where function *order*, allocates order(s) to each reference app.

Details of feature unification are presented in the subsequent subsection.

### *3.1.3 Cross Diffusion*

Three traffic feature vectors created are fused in a way to extract complementary info in them by optimal non-linear cross-diffusion of generated sparse and order graphs. Features fusion via cross-diffusion technique[27] was offered and its results confirm that the feature fusion by non-linear graphs techniques are significantly more accurate than linear graph-centered approaches. Graph-built unification retains robust features of apps and rejects all the frail features that add to misclassification.

Order graphs averts biasness and sparse graphs guarantees elimination of any outlier behavior. Cross diffusion [Eq. (9)] basically comprised of addition of sparse graphs of two other feature vectors to form $\alpha^{\phi^t}$ [Eq.(8)], where $\alpha^{\phi^t}$ is the $t^{th}$ element of the set $\phi$ and subsequently their multiplication with order graph of the feature vector with $\alpha^{\phi^t}$ .

$$\alpha^{\phi^t} = \sum_{Y} \eta^{Y} \text{ where, } Y \in \{\phi\} - \{\phi^t\} \tag{8}$$

$$\beta^{\phi^t} = \alpha^{\phi^t} \square \mu^{\phi^t} \quad \forall \phi \in \{1,2,3\} \tag{9}$$

Here, we add the sparse graph of two other feature vectors and then multiply it with the order graph of the feature vector turn by turn to generate three fused vectors. Eqn. (9) can be further represented [Eqn. (10,11,12)] in the form of fused vectors ( $F^{V^1}, F^{V^2}, F^{V^3}$ ), sparse vectors ( $S^{V^1}, S^{V^2}, S^{V^3}$ ) and order vectors ( $O^{V^1}, O^{V^2}, O^{V^3}$ ) as follows:

$$F^{V^1} = (S^{V^2} + S^{V^3}) \square O^{V^1} \tag{10}$$

$$F^{V^2} = (S^{V^3} + S^{V^1}) \square O^{V^2} \tag{11}$$

$$F^{V^3} = (S^{V^1} + S^{V^2}) \square O^{V^3} \tag{12}$$

The above-fused feature vectors $F^{V^1}$, $F^{V^2}$ and $F^{V^3}$ are concatenated to form the unified feature vector, U by Eqn. (13).

$$U = F^{V^1} + F^{V^2} + F^{V^3} \tag{13}$$

This unified feature vector is given to three parallel classifier(s) i.e. SVM (Support Vector Machine), RF (Random Forest) & KNN (K Nearest Neighbors). The classifiers scores of these parallel classifiers are further fused to classify apps. Details of Optimal classifier fusion follow in the next sub-section.

## 3.2 Classifier Score-Fusion

Created U is fed to three classifiers connected in parallel. Obtained classification scores [$S_s$ (SVM), $S_r$ (RF), $S_k$ (KNN)] are again fused by the classifier score fusion technique. There are various score fusion techniques[28,29] reported in the literature. Here we have chosen the PCR-5[30] to solve the highly contradictory scores of the three classifiers. PCR-5 is used to solve ambiguous problems in multi-sensor score fusion. Android app detection is a certainly ambiguous problem as we are uncertain whether the app is malicious or not. In the suggested model, three classifiers are selected and the fusion of the output of these classifiers can be modeled as a multi-sensor score fusion problem as their outputs are independent of each other. Therefore, Android app detection satisfies all the conditions of the PCR-5 theory. In our framework, the frame of discernment $\Theta$ has two elements B M and corresponding benign and malign. Classifier scores [$S_s$, $S_r$, $S_k$] are converted individual Basic Belief assignments or belief masses by the Eqn.(14).

$$\begin{aligned} m_i(\mathrm{B}) &= C_i \times S_i(\mathrm{B}) \\ m_i(\mathrm{M}) &= 1 - C_i \times S_i(\mathrm{B}) \end{aligned} \tag{14}$$

Where $i \in (s,r,k)$ & $C_i$ denotes the confidence measure of the single classifier.

These belief masses are combined by PCR-5 rules. The conjunctive consensus among the classifiers is assessed by Eqn. (15, 16)

$$m_{srk}(\mathrm{B}) = m_s(\mathrm{B}) * m_r(\mathrm{B}) * m_k(\mathrm{B}) \tag{15}$$

$$m_{srk}(\mathrm{M}) = m_s(\mathrm{M}) * m_r(\mathrm{M}) * m_k(\mathrm{M}) \tag{16}$$

Overall conflict among the classifiers is estimated by Eqn. (17). It comprises six partial conflicts-masses.

$$\begin{aligned} m_{srk}(\mathrm{B} \cap \mathrm{M}) = {} & m_s(\mathrm{B}) * m_r(\mathrm{M}) * m_k(\mathrm{M}) + m_r(\mathrm{B}) * m_s(\mathrm{M}) * m_k(\mathrm{M}) + \\ & m_k(\mathrm{B}) * m_s(\mathrm{M}) * m_r(\mathrm{M}) + m_s(\mathrm{M}) * m_r(\mathrm{B}) * m_k(\mathrm{B}) + \\ & m_r(\mathrm{M}) * m_s(\mathrm{B}) * m_k(\mathrm{B}) + m_k(\mathrm{M}) * m_s(\mathrm{B}) * m_r(\mathrm{B}) \end{aligned} \tag{17}$$

Six partial conflict masses are further redistributed using PCR-5 rules in ratio to masses assisting these partial conflicts. The values of $p_i$ and $q_i$ are a contribution to Benign and Malign masses following reallocation of partial conflicts, where $i$ = 1,…,6 and are determined by Eqn. (18-23).

$$\frac{p_1}{m_s(\mathrm{B})} = \frac{q_1}{m_r(\mathrm{M}) * m_k(\mathrm{M})} = \frac{m_s(\mathrm{B}) * m_r(\mathrm{M}) * m_k(\mathrm{M})}{m_s(\mathrm{B}) + m_r(\mathrm{M}) * m_k(\mathrm{M})} \tag{18}$$

$$\frac{p_2}{m_r(\mathrm{B})} = \frac{q_2}{m_s(\mathrm{M}) * m_k(\mathrm{M})} = \frac{m_r(\mathrm{B}) * m_s(\mathrm{M}) * m_k(\mathrm{M})}{m_r(\mathrm{B}) + m_s(\mathrm{M}) * m_k(\mathrm{M})} \tag{19}$$

$$\frac{p_3}{m_k(\mathrm{B})} = \frac{q_3}{m_s(\mathrm{M}) * m_r(\mathrm{M})} = \frac{m_k(\mathrm{B}) * m_s(\mathrm{M}) * m_r(\mathrm{M})}{m_k(\mathrm{B}) + m_s(\mathrm{M}) * m_r(\mathrm{M})} \tag{20}$$

$$\frac{p_4}{m_r(\mathrm{B}) * m_k(\mathrm{B})} = \frac{q_4}{m_s(\mathrm{M})} = \frac{m_s(\mathrm{M}) * m_r(\mathrm{B}) * m_k(\mathrm{B})}{m_s(\mathrm{M}) + m_r(\mathrm{B}) * m_k(\mathrm{B})} \tag{21}$$

$$\frac{p_5}{m_s(\mathrm{B}) * m_k(\mathrm{B})} = \frac{q_5}{m_r(\mathrm{M})} = \frac{m_r(\mathrm{M}) * m_s(\mathrm{B}) * m_k(\mathrm{B})}{m_r(\mathrm{M}) + m_s(\mathrm{B}) * m_k(\mathrm{B})} \tag{22}$$

$$\frac{p_6}{m_s(\mathrm{B}) * m_r(\mathrm{B})} = \frac{q_6}{m_k(\mathrm{M})} = \frac{m_k(\mathrm{M}) * m_s(\mathrm{B}) * m_r(\mathrm{B})}{m_k(\mathrm{M}) + m_s(\mathrm{B}) * m_r(\mathrm{B})} \tag{23}$$

Approximated contributions $p_i$ and $q_i$ where $i$ = 1,…,6 are added on to their respective conjunctive consensus. The final

belief of the app being benign $m_{PCR5}$ (B) and that of an app being malign $m_{PCR5}$ (M) are determined by Eq. (24) and Eq. (25) respectively.

$$m_{PCR5}(\mathrm{B}) = m_{srk}(\mathrm{B}) + \sum_{i=1}^{6} p_i \tag{24}$$

$$m_{PCR5}(\mathrm{M}) = m_{srk}(\mathrm{M}) + \sum_{i=1}^{6} q_i \tag{25}$$

The decision about the given test app $\boldsymbol{t}$ is taken after a comparison of the $m_{PCR5}$ (B) with $m_{th}$. If $m_{PCR5}(\mathrm{B}) \geq m_{th}$ then the $\boldsymbol{t}$ is acknowledged as benign else it is taken as malign.

## 4. EVALUATION

Quantitative analysis on various performance matrices like Precision, Accuracy, F1 Score, Specificity, and Sensitivity was performed on the suggested and two other state-of-the-art methods. Their ROC (Receiver Operating Characteristic) plots were also drawn for comparison. The ROC curve is the plot of FAR (False Acceptance Rate) vs 1-FRR (False Rejection Rate) Following are the details of the evaluation process.

### 4.1 Databases

We choose about 3000 samples of malign apps and benign apps each downloaded from the benchmarked datasets i.e. Drebin[31], AMD[32], Androzoo[33], and CICMalDroid2020[34]. Benign apps are taken from Google Playstore. The malign apps generated both the benign and malign traffic as most of these malign apps are formed from the benign apps by their reverse engineering and repackaging after inserting maliciousness into them. Finally, a dataset is constructed by TCP-based network flows to assess the performance of the proposed framework. Network flows corresponding to benign and malign apps are stored in pcap format by running selected apps over and over for a period of 6 hrs. We select only those apps from the datasets that produce the network traffic and filter out the TCP-based flows. Traffic features as tabulated in Table 1 are extracted from these flows using the script written in python.

We select about 500 feature-flow vectors each corresponding to benign and malign flows from the total extracted flows corresponding to DB1, DB2, DB3, and DB4 datasets to train our model. We further integrate these 500 feature flow vectors of benign and malign flows into an integrated set of 2000 benign and malign flows each to form the fifth set of TCP features containing 2000 benign and malign feature vectors each. Further, we apply a ten-fold cross-validation procedure for training and testing are done on five sets of feature sets and finally, average values are taken as the results. All investigations are done on MATLAB R2018a on 16GB RAM, i7, and 2.7 GHz processor.

### 4.2 Performance Assessment

Performance of the suggested framework is realised by the means of TCP features extracted from TCP flows corresponding to malign and benign apps as in Table 2 and calculating evaluation matrices through ten-fold cross-validation. It is compared to training time and mean-time-to-detect (MTTD) against three state-of-the-art approaches. Here, MTTD is the time taken to detect malapps. Evaluation matrices results are also compared with three state-of-the-art techniques FED[12], MMD[20], and LWN[21].

**Table 2. Experimentation dataset**

| App type Dataset | Malign_ Apps(M) | Benign_ Apps(B) | Source |
|---|---|---|---|
| DataBase(DB)1 | 750 | 750 | Androzoo(M) GooglePlay(B) |
| DataBase(DB)2 | 750 | 750 | AMD(M) CICMalDroid2020(B) |
| DataBase(DB)3 | 750 | 750 | CICMalDroid2020(M) GooglePlay(B) |
| DataBase(DB)4 | 750 | 750 | Drebin(M) GooglePlay(B) |

The three state-of-the-art methods viz FED, MMD and LWN are compared with the proposed method. In FED, a malware detection technique is built on TCP n/w traffic, wherein network traffic generated by apps comprises TCP flows. These flows are used to extract packet sizes as features. Standard deviation, Mean, min, and max values of the sizes of the first few packets of TCP flows are taken as features. Here author leverages the RF ML algorithm to train the model to achieve 97 per cent of detection accuracy. In the second state-of-the-art method MMD, the authors proposed a framework for android malapp identification leveraging both the TCP and HTTPS features. Here, the malapp detection is done on the server-side without affecting the user experience. C4.5 Decision tree ML algo is used to train the model with 8312 benign and 5560 malign apps for identifying unknown apps with an accuracy of 97.89 per cent. In the LWN method, a lightweight malapp detection framework using TCP-based network features was given with an accuracy of 90 per cent and a false-positive rate of less than 3 per cent. Here, the total number of 359 malapp and benign apps are used along with two RF and AdaBoost ML algorithms.

Detection Accuracy, F1 Score, Specificity, Sensitivity, and Precision are calculated using equations 26, 27, 28, and 29 respectively.

$$Detection\ Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{26}$$

$$F1\ Score = \frac{2TP}{2TP + FP + FN} \tag{27}$$

$$Specificity = \frac{TN}{TN + FP} \tag{28}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{29}$$

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{30}$$

where, TP, FP, TN, and FN are true positive, false positive, true negative, and false negative respectively. ROC curves for the proposed method and comparative methods FED, MMD, and LWN for flow sets corresponding to datasets DB1, DB2, DB3, and DB4 are depicted in Fig. 3. The experimental results are presented in Table 3. It is observed that the mean value of F

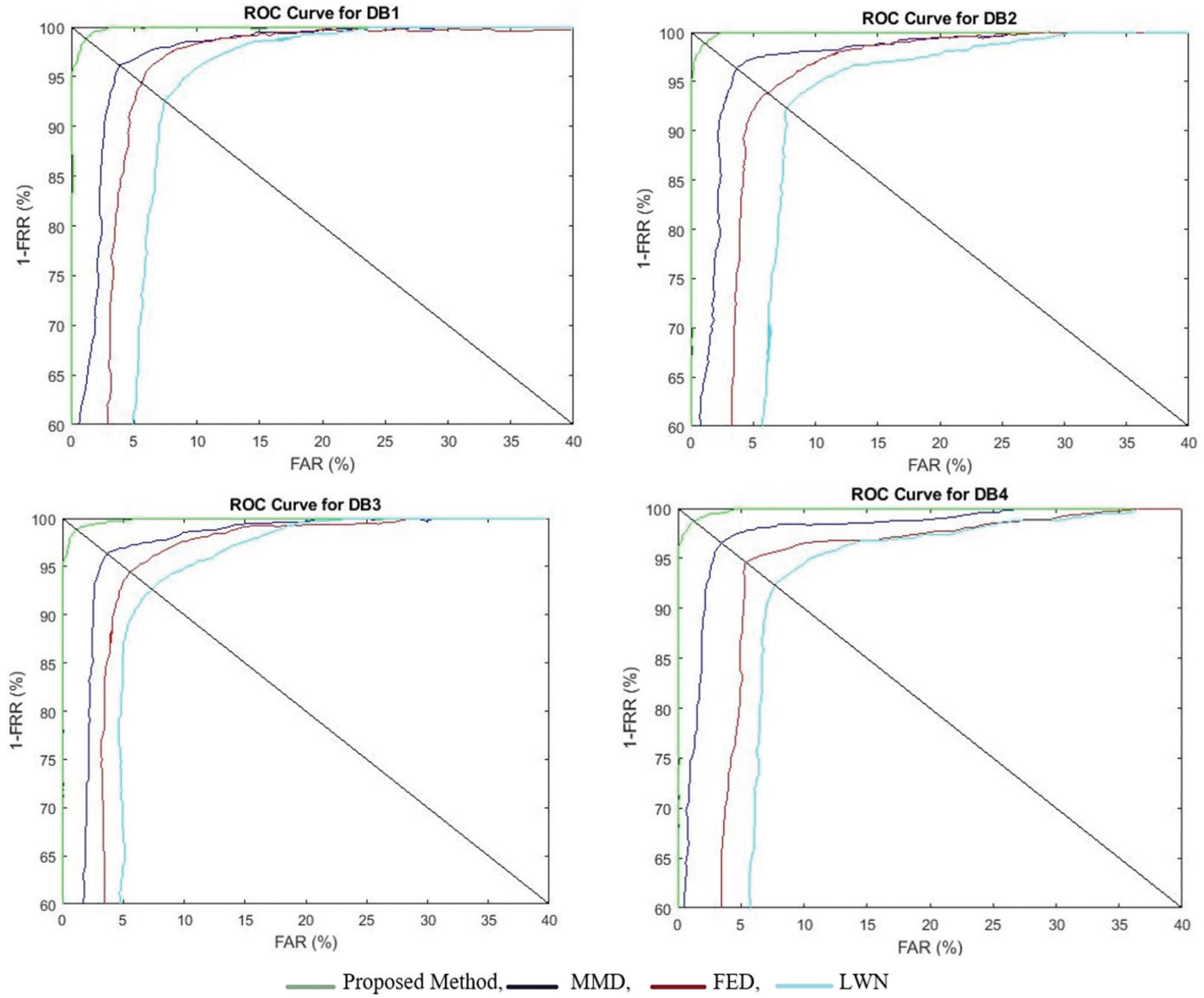



**Figure 3. ROC curves for the proposed method and comparative methods MMD, FED, and LWN for flow sets corresponding to datasets DB1, DB2, DB3, and DB4.**

score, detection accuracy, sensitivity, specificity, and precision for the proposed framework are 0.98742, 98.74 per cent, 99.04 per cent, 98.44 per cent, and 98.456 per cent respectively. The highest value of the F score, detection accuracy, sensitivity, specificity, and precision is 0.9890, 98.90 per cent, 99.80 per cent, 99.60, and 98.61 respectively.

The proposed technique outshines similar state-of-the-art traffic-based methods when evaluated on extracted TCP features as presented in Table 1. Enhancement for mean accuracy of the suggested scheme is achieved by 6.28 per cent, 4.34 per cent, and 2.3 per cent over LWN, FED, and MMD respectively. This enhanced value of accuracy of the suggested framework is attained by the optimal combination of traffic features using a cross-diffusion strategy and fusion of classifier(s) score using the DSmT-based PCR-5 rule to form a clear-cut border for differentiating malign apps from benign apps. Here, three feature vectors comprising the three sets of complementary features are fused in a nonlinear style through cross-diffusion wherein the rugged features are enhanced and the frail features are repressed resulting in the robust UF. Further, in the optimal score level fusion process, UF is fed to the three ML classifiers where concurrent scores are boosted and discordant classifier scores are suppressed. Also, the conflict between the multiple classifiers is resolved to get the final decision score.

The suggested framework is also compared to MTTD of different state-of-the-art techniques. To calculate the MTTD of contemporary methods, learned models are fed with the 250 arbitrary apps for investigation. Our proposed method attains an amazing average analysis performance of 5.8 seconds per app. Similarly, the average analysis performance of LWN, FED, and MMD comes out to be 6.5 seconds, 7.3 sec., and 6.9 seconds respectively. Hence, our proposed framework is outshined other methods in respect of detection time, detection accuracy, and efficacy in real-life app scenarios. MTTD also confirms that the suggested framework detects the apps with high accuracy in a reasonable time.

**Table 3. Performance comparison in terms of accuracy with existing methods using captured data from 20 different apps**

| Apps | UF+RF | LWN | FED | MMD | Proposed method |
|---|---|---|---|---|---|
| | Accuracy | | | | |
| AccuRadio | 0.8900 | 0.8300 | 0.8500 | 0.8900 | 0.9400 |
| Maps | 0.8700 | 0.8100 | 0.8400 | 0.8100 | 0.9100 |
| WhatsApp | 0.8600 | 0.8200 | 0.8100 | 0.8700 | 0.9300 |
| Outlook | 0.8300 | 0.8100 | 0.8300 | 0.8500 | 0.9100 |
| Mail | 0.7800 | 0.8400 | 0.8500 | 0.8400 | 0.9300 |
| Netflix | 0.9000 | 0.7700 | 0.8900 | 0.9100 | 0.9500 |
| Twitter | 0.9300 | 0.7500 | 0.7600 | 0.9200 | 0.9400 |
| Facebook | 0.9100 | 0.9000 | 0.8700 | 0.9300 | 0.9500 |
| Youtube | 0.8800 | 0.8600 | 0.8800 | 0.9200 | 0.9700 |
| Gmail | 0.9200 | 0.9300 | 0.8900 | 0.9500 | 0.9600 |
| DroidDream | 0.8100 | 0.8400 | 0.9000 | 0.9200 | 0.9800 |
| DroidKungFu1 | 0.8700 | 0.8800 | 0.9300 | 0.9400 | 0.9600 |
| Buzz | 0.8200 | 0.8400 | 0.9100 | 0.8900 | 0.9400 |
| BlueScanner | 0.8500 | 0.8400 | 0.9400 | 0.9300 | 0.9800 |
| Plankton | 0.8200 | 0.8500 | 0.9200 | 0.9100 | 0.9600 |
| WallpaperGirls | 0.8100 | 0.8200 | 0.8300 | 0.8600 | 0.9300 |
| StylePhotoCollage | 0.8700 | 0.8400 | 0.9100 | 0.9200 | 0.9400 |
| PrivateSms | 0.8300 | 0.8500 | 0.9400 | 0.8700 | 0.9900 |
| PartMessage | 0.8800 | 0.9000 | 0.9500 | 0.8900 | 0.9700 |
| IdeaSecurity | 0.8200 | 0.8800 | 0.9000 | 0.9600 | 0.9800 |
| Average Accuracy | 0.8575 | 0.8430 | 0.8882 | 0.8990 | 0.9510 |
| Variance | 0.00171 | 0.00182 | 0.00246 | 0.00154 | 0.00054 |

**Table 4. Performance Metrics (PM) for proposed methods and other comparative methods.**

| Dataset | PM | LWN | FED | MMD | Proposed Method |
|---|---|---|---|---|---|
| DB1 FLOWSET | Accuracy | 0.9250 | 0.9440 | 0.9620 | 0.9880 |
| | Specificity | 0.9100 | 0.9280 | 0.9540 | 0.9960 |
| | Sensitivity | 0.9400 | 0.9600 | 0.9700 | 0.9800 |
| | F1 Score | 0.9261 | 0.9449 | 0.9623 | 0.9879 |
| | Precision | 0.9126 | 0.9302 | 0.9547 | 0.9959 |
| DB2 FLOWSET | Accuracy | 0.9230 | 0.9380 | 0.9630 | 0.9890 |
| | Specificity | 0.9060 | 0.9360 | 0.9660 | 0.9860 |
| | Sensitivity | 0.9400 | 0.9400 | 0.9600 | 0.9920 |
| | F1 Score | 0.9243 | 0.9381 | 0.9629 | 0.9890 |
| | Precision | 0.9091 | 0.9363 | 0.9658 | 0.9861 |
| DB3 FLOWSET | Accuracy | 0.9260 | 0.9440 | 0.9630 | 0.9881 |
| | Specificity | 0.9140 | 0.9420 | 0.9460 | 0.9780 |
| | Sensitivity | 0.9380 | 0.9460 | 0.9800 | 0.9980 |
| | F1 Score | 0.9269 | 0.9441 | 0.9636 | 0.9881 |
| | Precision | 0.9160 | 0.9422 | 0.9478 | 0.9784 |
| DB4 FLOWSET | Accuracy | 0.9220 | 0.9460 | 0.9660 | 0.9870 |
| | Specificity | 0.8860 | 0.9260 | 0.9560 | 0.9780 |
| | Sensitivity | 0.9580 | 0.9660 | 0.9760 | 0.9960 |
| | F1 Score | 0.9247 | 0.9471 | 0.9663 | 0.9871 |
| | Precision | 0.8937 | 0.9288 | 0.9569 | 0.9784 |
| INTEGRATED FLOWSET | Accuracy | 0.9270 | 0.9480 | 0.9680 | 0.9850 |
| | Specificity | 0.9405 | 0.9455 | 0.9645 | 0.9840 |
| | Sensitivity | 0.9133 | 0.9505 | 0.9715 | 0.9860 |
| | F1 Score | 0.9260 | 0.9481 | 0.9681 | 0.9850 |
| | Precision | 0.9388 | 0.9458 | 0.9647 | 0.9840 |

Performance comparison in terms of accuracy with existing methods using online captured data from 20 different apps chosen from diverse sources is shown in Table 3. Here, we have taken Smartphone Samsung Galaxy S9 having android 8 OS with 8 GB RAM and 64 GB storage for generating the network traffic. Our framework detects the apps that generate the flows similar to one generated during the training phase with high accuracy. The trained model is also tested on the network flows captured from the 20 different apps under an unconstrained environment using real smartphones instead of emulators and the performance comparison is done with our framework versus three different state-of-the-art methods viz. LWN, FED, and MMD and one self-proposed method RF+UF i.e. unified feature fed to the random forest algorithm. From Table 3, the average accuracy obtained when the random apps are tested on LWN, FED, MMD, RF+UF, and the proposed method is 85.75 per cent, 84.30 per cent, 88.82 per cent, 89.90 per cent, and 95.10 per cent respectively. Our proposed framework is dependent on the selected number of features. Generating the traffic flows in an unconstrained environment is cumbersome as traffic generation gets hindered when the internet connection is interrupted. Sometimes, the malign app took a significant amount of time to generate malicious behavior in traffic flow.

## 5. CONCLUSION

A novel network traffic analysis-based framework has been proposed in the manuscript to detect android malapps by exploiting the traffic features. It consists of four blocks viz. traffic feature fusion, classifier score-fusion, decision criteria, and reference apps update to accomplish efficient malapp detection Fifteen TCP-based features have been used in feature fusion to get a unified feature vector which is given to three Multilayer classifiers and find fusion scores for detecting malwares. The reference dictionary apps are regularly updated with zero-day malign and benign apps to improve the detection ability of the framework. It has been shown that the Framework proposed achieves an average detection accuracy of 98.74 per cent and outperforms other existing methods. A comparison of the evaluation matrices of the suggested framework with other contemporary approaches reveals better detection accuracy of malapps. Our future endeavor is to incorporate other dynamic and static attributes for detecting sophisticated malapps and to overcome all the limitations discussed.

## REFERENCES


1. IDC. https://www.idc.com/promo/smartphone-market-share (Accessed on August 06, 2021).
2. Statistica. https://www.statista.com/statistics/277125/share-of-website-traffic-coming-from-mobile-devices (Accessed on August 06, 2021).
3. Yadav, A. & Ratan, R. Identifying traffic of same keys in cryptographic communications using fuzzy decision criteria and bit-plane measures. *Int. J. Syst. Assur. Eng. Manag*, 2020, **11**(9), 466–480. doi: 11. 10.1007/s13198-019-00878-7.
4. Ratan, R. & Yadav, A. Security analysis of bit plane level image encryption schemes. *Def. Sci. J.,*, 2021, **71**(2), 209-221. doi:10.14429/dsj.71.15643
5. Kumar, S.; Indu, S. & Walia, G.S. An efficient multistage fusion approach for smartphone security analysis *Def. Sci. J.,* 2021, **71**(4),476–490. doi: 10.14429/dsj.71.15077.
6. Kumar S.; Indu S. & Walia G.S. Smartphone traffic analysis: a contemporary survey of the state-of-the-art. *In* Proceedings of the Sixth International Conference on Mathematics and Computing. Advances in Intelligent Systems and Computing, Springer, Singapore, 2020, **1262**, 325-343 doi:10.1007/978-981-15-8061-1_26.
7. Yerima, S.Y. & Alzaylaee, M.K. & Sezer, S. Machine learning-based dynamic analysis of android apps with improved code coverage. 2019, *Eur. J. Inf. Secur.*, 2019, **1**, 1–24. doi: 10.1186/s13635-019-0087-1.
8. Agrawal, A.; Bhatia, A.; Bahuguna, A.; Tiwari, K.; Haribabu, K.; Vishwakarma, D. & Kaushik, R. A survey on analyzing encrypted network traffic of mobile devices. *Int. J. Inf. Secur.*, 2022, **1**. doi: 10.1007/s10207-022-00581-y.
9. Kumar, S.; Indu, S. & Walia, G.S. Optimal unification of static and dynamic features for smartphone security analysis. *Intell. Autom. Soft Co.*, 2022, **33**(4). doi:10.32604/iasc.2022.024469
10. Arora, A. & Garg, S. & Peddoju, S.K. Malware detection using network traffic analysis in android based mobile devices. *In* Eighth International Conference on Next Generation Mobile Apps, Services and Technologies, 2014, 66–71. doi: 10.1109/NGMAST.2014.57.
11. Arora, A. & Peddoju, S.K. NTPDroid: A hybrid android malware detector using network traffic and system permissions. *In* 17th IEEE Eighth International Conference on Trust. Secur. Priv. Comput. Commun., 2020, 808–813. doi: 10.1109/TrustCom/BigDataSE.2018.00115
12. Wang, S.; Yan, Q.; Chen, Z.; Yang, B.; Zhao, C. & Conti, M. Detecting android malware leveraging text semantics of network flows. *IEEE Trans. Inf. Forensics Secur.*, 2018, **13**(5),1096–1109. doi: 10.1109/TIFS.2017.2771228.
13. Liu, A.; Chen, Z.; Wang, S.; Peng, L.; Zhao, C. & Shi, Y. A fast and effective detection of mobile malware behavior using network traffic. *In* Algorithms and Architectures for Parallel Processing, edited by J. Vaidya and JoLi. 11337, Springer, 2018. doi:10.1007/978-3-030-05063-4_10
14. Li, Z.; Sun, L.; Yan, Q.; Srisaan, W. & Chen, Z. DroidClassifier: Efficient adaptive mining of application-layer header for classifying android malware. *In* International Conference on Security and Privacy in Communication Systems. **2017**, 597–616. doi: 10.1007/978-3-319-59608-2_33.
15. Li, Q.; Chen, Z.; Yan, Q.; Wang, S.; Ma, K.; Shi, Y. & Cui, L. MulAV: Multilevel and explainable detection of

android malware with data fusion. **11337,** 2018 In: *LNCS*. Springer International Publishing.
16. Wang, S.; Hou, S.; Lei, Z.; Chen, Z. & Han, H. Android malware network behavior analysis at HTTP protocol packet level. 2015. *In* Proceedings of the ICA3PP International Workshops and Symposiums on Algorithms and Architectures for Parallel Processing*.,* **9532**, 497–507.
doi:10.1007/978-3-319-27161-3_45.
17. Su, X.; Lin, J.; Shen, F. & Zheng, Y. Two-phases detection scheme: Detecting android malware in android markets, *In* ATCI Springer **842**, 389–399, 2019.
doi:10.1007/978-3-319-98776-7_41
18. Zulkifli, A.; Hamid, I.; Shah, W. & Abdullah, Z. Android malware detection based on network traffic using decision tree algorithm. *Adv. Intell. Syst. Comput.*, 2018, **700**, 485–494.
doi: 10.1007/978-3-319-72550-5_46.
19. Malik, J. & Kaushal, R. CREDROID: Android malware detection by network traffic analysis. *In PAMCO*. 2016, 28-36.
doi:10.1145/2940343.2940348.
20. Wang, S.; Chen, Z; Yan, Q.; Ji, K.; Peng, L.; Yang, B. & Conti, M. Deep and broad URL feature mining for android malware detection. *Information Science*, 2000, **513**, 600–613.
doi: 10.1016/j.ins.2019.11.008.
21. Wang, S.; Chen, Z.; Yan, Q.; Ji, K.; Peng, L.; Yang, B. & Jia, Z. A mobile malware detection method using behavior features in network traffic. *J. Netw. Comput. Appl*., 2019, **133**, 15–25, .
doi: 10.1016/j.jnca.2018.12.014.
22. Sanz, I.J.; Lopez, M.A.; Viegas, E.K. & Sanches, V.R. "A lightweight network-based android malware detection system," In: *IFIP Networking Conference (Networking)*, 2020, 695-703.
23. Upadhayay, M.; Sharma, G.; Garg, G. & Arora, A. RPNDroid: Android malware detection using ranked permissions and network traffic, *In* Fifth World Conference on Smart Trends in Systems Security and Sustainability (WorldS4)*,* 2021, 19-24.
doi: 10.1109/WorldS451998.2021.9513992.
24. Alshehri, M. APP-NTS: A network traffic similarity-based framework for repacked Android apps detection. *J. Ambient Intell. Human Comput.*, 2022, **13,** 1537–1546.
doi: 0.1007/s12652-021-03023-0
25. Sihag, V.; Choudhary, G.; Vardhan, M.; Singh, P. & Seo, J. T. PICAndro: Packet inspection-based android malware detection 2021, *Secur. Commun. Netw*., 2021, 9099476, 11.
doi: 10.1155/2021/9099476
26. Norouzian, M.R.; Xu, P.; Eckert, C. & Zarras, A. Hybroid: toward android malware detection and categorisation with program code and network traffic. *In*: Information Security. ISC 2021 Springer, Cham., *Lecture Notes Comput. Sci.*, **13118**, 259-278. .
doi:10.1007/978-3-030-91356-4_14
27. Walia, G.S. *et al.* Unified graph-based multicue feature fusion for robust visual tracking. *IEEE T. Cybernetics*, 2020, **50**(6), 2357-2368.
doi: 10.1109/TCYB.2019.2920289.
28. Xiao, Fuyuan. Multi-sensor data fusion based on the belief divergence measure of evidence and the belief entropy. *Information Fusion*, 2019, **46**, 23-32.
doi: 10.1016/j.inffus.2018.04.003.
29. Xiao, Fuyuan. A new divergence measure for belief functions in D–S evidence theory for multisensor data fusion. *Information Sciences*, 2020, **514**, 462-483.
doi: 10.1016/j.ins.2019.11.022.
30. Smarandache, Florentin & Dezert, J. Advances and applications of DSmT for information fusion, *Am. Res. Press*, 2015, **4**. http://fs.gallup.unm.edu/DSmT-book4.pdf. (Accessed on 10/09/2021).
31. Daniel, A.; Spreitzenbarth, M.; Hubner, M. & Gascon, H. DREBIN: Effective and explainable detection of android malware in your pocket. Symposium on Network and Distributed System Security (NDSS)*,* 2014.
doi:10.14722/ndss.2014.23247.
32. Wei, F.; Li, Y.; Roy, S.; Ou, X.; Zhou, W. Deep ground truth analysis of current android malware. In: Polychronakis, M., Meier, M. (eds) Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA) 2017. Springer, Cham. *Lecture Notes Comput. Sci.*, 2017, **10327**.
doi:10.1007/978-3-319-60876-1_12
33. Mahdavifar, S.; Kadir, A.; Fatemi, R. & Alhadidi, D. Ghorbani. Dynamic android malware category classification using semi-supervised deep learning. *In* 18th IEEE International Conference on Dependable, Autonomic, and Secure Computing (DASC)*,* 2020, 17-24.
34. Allix, K.; Bissyandé, T.F.; Klein, J. & Traon, Y.L. AndroZoo: Collecting millions of android apps for the research community. 2016 IEEE/ACM 13th Working Conference on Mining Software Repositories (MSR), 2016, 468-471.
doi:10.1109/MSR.2016.056

## CONTRIBUTORS

**Mr Sumit Kumar** received his ME in Electronics and Communication Engineering from Delhi College of Engineering, University of Delhi, Delhi. He is working as a Senior Scientist at DRDO-SAG, Delhi, India. His current areas of research interest include: Traffic analysis, smartphone security, cloud computing, and hardware analysis.
In the present study, he has carried out complete design and development scheme. He has also completed the experimental work by evaluating the suggested framework and comparing the results with other state-of-the-art methods.

**Prof S. Indu** received her PhD in the area of visual sensor networks from University of Delhi, Delhi, India. She is working as Dean (Student Welfare) and Professor of ECE Department of Delhi Technological University. Her areas of research interest are: Computer vision, sensor networks and image processing.
In the present study, she has provided expert guidance in problem formulation and offered necessary direction and overall support to carry out this study successfully.

**Dr Gurjit Singh Walia** obtained his PhD in the field of Computer Vision from Delhi Technological University (Formerly Delhi College of Engineering), Delhi. He is working as a Senior Scientist in DRDO, Delhi. His current research interests include: Machine learning, pattern recognition, and information security.

In the present study, he has guided the author and supervised the work.

© 2022. This work is licensed under http://creativecommons.org/licenses/by-nc-nd/2.5/in(the “License”). Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License.

2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) | 978-1-6654-1120-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAECT54875.2022.9807920

# *A Performance Study on Deep Learning Covid-19 Prediction through Chest X-Ray Image with ResNet50 Model*

Dhirendra Kumar
*Department of Applied Mathematics*
*Delhi Technological University*
Delhi, India
dhirendrakumar@dtu.ac.in

Pulkit Sharma
*Department of Electrical Engineering*
*Delhi Technological University*
Delhi, India
pulkitsh08@gmail.com

Anupama
*Department of Electrical Engineering*
*Delhi Technological University*
Delhi, India
anupama@dtu.ac.in

Parth Sharma
*Department of Electrical Engineering*
*Delhi Technological University*
Delhi, India
parthsh15@gmail.com

***Abstract*—The COVID-19 epidemic has claimed many lives throughout the world and constitutes an unprecedented public health concern. The key challenge in early detection of the corona virus is early detection. And the main obstacle was the similarity of COVID-19 symptoms to flu symptoms. With the goal of saving human lives and stemming the spread of a worldwide pandemic, an accurate and speedy analysis of COVID-19-induced pneumonia has now taken centre stage. Responding this urgent concern and to reduce the burden as well as chances of faulty manual diagnosis, several deep learning approaches are developed to conduct early diagnosis. Based on the availability of reliable patient's records, an accepted technique is pre-trained deep learning prediction approach through patient's chest X-Rays. Convenience of this approach led development of a number of novel deep learning-based lung screening technologies. However, little emphasis is placed on ensuring the quality of their output. Pre-trained deep learning systems will be used in this project to evaluate their ability to recognise and diagnose disorders. To categorise COVID and normal pictures, a neural network-based ResNet50 architecture is presented. The implementation is based on the normal, COVID, and lung opacity datasets. For data pre-processing, ImageDataGenerator is used, which rescales, flips, and modifies the range to meet the model. To categorise the x-ray images, the suggested method ResNet50 architecture is used. Performance matrices like precision, accuracy, recall, as well as F1-score are examined to verify the algorithm's usefulness. The suggested technique has an accuracy of 80%, indicating that the proposed model is quite good in classifying COVID and normal x-ray pictures. This research will have a significant influence on real-time since it will accurately diagnose COVID in less time, perhaps lowering the mortality rate.**

***Keywords— ResNet50, deep learning, chest x-ray, COVID-19***

## I. INTRODUCTION

Corona virus disease 2019 has been erupting since December 2019, as one of the deadliest pandemics in decades, and has been rising since then. It has wreaked havoc on the economy and generated unimaginable turmoil [1]. COVID-19 has been found in humans, cats, pigs, dogs, birds, and rodents in various forms in countries such as the United States, Spain, India, Italy, China, the United Kingdom, and Iran, among others. Identifying and isolating ill persons is the WHO's technique for preventing the virus from spreading to healthy people. As PCR, antigen, and antibody testing has numerous drawbacks including the amount of time it takes to receive results, the accuracy of its findings, and cost of testing, it is crucial to remember that these tests have many limits that must be considered. Central laboratories, such as the CDC, are required to perform the PCR test to detect COVID 19 in respiratory swabs [2], are required. X-rays and Computed Tomography (CT) scans were also used by clinicians as a consequence. To diagnose lung disease, an X-ray or CT scan may be performed since COVID 19 affects the cells that line the respiratory system. Due to the limits of early pandemic testing like polymerase chain reaction with reverse transcription (RT-PCR) and the need to create new techniques to reduce COVID-19 spread, a slew of machine learning methods based on lung scans have been proposed.

Several advances in computer vision-related fields, such as image segmentation, identification, and object recognition, This has been made possible because to the adoption of deep learning techniques in convolutional neural networks. (CNN) [3]. These properties were highly helpful in recognising COVID-19 when multi-class and binary classification tasks VGG19, MobileNet, ResNet, as well as other pre-trained CNN models were used in the experiments.

- **Global Impact of Covid-19 Infection**

Within two months following the first confirmed Covid-19 case in Wuhan, China, there were roughly 1000 confirmed cases and over 5000 suspected cases. Concerning A Public Health Emergency of International Concern, classified as COVID-19 by the WHO on January 30, 2020, has occurred as a result of a recent epidemic of the Coronavirus [4]. On March 11, 2020, a

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:40:24 UTC from IEEE Xplore. Restrictions apply.

global pandemic was declared due to the disease's rapid global spread. The new coronavirus pneumonia epidemic had spread over the world by September 2020, with the number of confirmed cases increasing every day. Phylogenetically, COVID-19 is linked to SARS-like bat viruses, according to a genomic analysis. Thus, bats might be a reservoir for viral replication. COVID-19's intermediate host, the pangolin, has also been discovered [5]. For the first time, this newly discovered virus is distinct from SARS-CoV, influenza, MERS-CoV and avian influenza. A COVID-19 detection system based on X-ray data extraction as well as communication mechanism are the most challenging aspects of the system to implement. This would take a long time and need the assistance of medical professionals. In order to identify COVID-19 participants from X-ray images of the chest, computer-aided approaches are necessary.

COVID-19 incidents have been discovered using machine learning techniques as the outbreak has spread. COVID-19 was discovered by Wang et al. using a CT-based Inception network [6]. Using the transfers learning system and an inception V3 model of transfer, The X-ray images of the COVID-19 chest were correctly identified by Asif et al. ResNet50 pre-trained network was used by Song et al. to distinguish between healthy persons, COVID-19 and bacterial pneumonia. COVID-19, as well as bacterial and viral pneumonias, were successfully detected by Loey et al. using the GooLeNet pre-training model [7]. Rahimzadeh and Attar developed a pair of CNNs using Xception and ResNet50V2 that properly identified 99.56 percent of COVID-19 chest X-ray images.

There are several benefits of using deep learning algorithms, such as: Making the greatest use of unstructured information, minimising superfluous expenses, decreasing feature engineering, and eliminating selective data labelling. Photographs may be categorised using deep learning algorithms because of their self-aware nature [8]. The analysis of medical pictures and the creation of exceptional classifier performance with much less time-consuming simulation workload have been made possible thanks to deep learning algorithms.
The proposed research discussed in this paper is planned and designed to justify the following large scale medical diagnosis concerns:

- Identification of the significant aspects of deep learning pre-trained technique that can be utilized for disease identification and diagnosis.
- Efficacy of ResNet architecture as Using Chest X-Ray pictures, automatic data categorization was performed for Covid-19 prediction.

The study experimentally analyses fundamental areas that impact the model efficacy, such as, feature extraction, classification and dataset choices and others. Alongside it measures the model performance as an automated diagnosis tool.

The paper is organized in such a way that chapter-2 deals with literature review, chapter-3 explains the proposed methodology, chapter-4 displays the results and finally chapter-5 concludes the work.

## II. RELATED WORKS

Covid-19 pandemic during 2020-21 witnessed a rapid and unpredictable rise of patient count all over the world that resulted in urgent need of fast and reliable disease diagnosing procedure. In several medical academic literatures, deep learning-based diagnosis schemes are suggested as the solution to reduce the laborious, time consuming and fault prone manual clinical processes. Deep Learning-based Covid-19 prediction from X-ray pictures of a patient's chest is one acceptable way. that is thoroughly studied in this research to fulfil its goal. Authors [9] conducted an experimental assessment of Deep Learning Disease Prediction model supported with ResNet utility to justify its relevance in large scale predictive diagnosis system. There were two dimensions to this research. The first component showed how Covid-19 registration slips were used to explain an innovative method to patient diagnosis.

Using ResNet-101, the data set of COVID-19 patient registration slips was analysed. This data set included 5003 E-registration slips with precise time. In terms of time, 82 percent of the predictions were correct. COVID-19 positive cases might be expected to occur on the next day.

The juxtaposition of MOE and MAE was also incorporated into the error framework. In the second dimension, A chest X-ray indicated COVID-19. This double chest X-ray was used to predict the presence of COVID-19. 8009 chest X-rays from the local population were acquired. Faster R-CNN, faster mask-CNN, as well as faster ResNet-50 are all examples of neuronal networks that have been suggested. R-CNN with a faster processing time has the best accuracy of 87 percent. ResNet-50 accuracy was 72 percent, while Mask RCNN accuracy was 83 percent. The ACC, PRC, and RCL performance parameters were used. To increase SVM performance, the batch normalising technique was added. The authors [10] built a multi-channel which was before ResNet architecture to assist in the diagnosis for COVID-19 chest X-rays. Normal, sick, and COVID-19 as well as non-COVID19 patients were categorized using three ResNet models. Pneumonia, COVID-19, and healthy X-rays were all utilised to build a one-to-one classification framework for all three instances.

Experts determined that the ensembles model outperformed the single model in this investigation. For each class, the model might extract more semantic information. The approach had a 94 percent precision and a 100 percent recall rate. It has the potential to assist doctors in screening patients for COVID-19, allowing for faster triage and treatment for improved outcomes. The authors of [11] aimed to propose a practical and quick solution to the diagnosing challenges that physicians experience when treating Covid-19 patients. In the Deep Transfer Learning (DTL) technique, they employed both the Apache Spark data platform as well as Convolutional Neural Networks (CNNs). COVID-19 chest X-ray pictures were used to evaluate Inception V3, ResNet 50, as well as VGG 19. The investigation employed both

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:40:24 UTC from IEEE Xplore. Restrictions apply.

Covid-19 as well as conventional X-ray images. which allowed for perfect precision. For COVID/Normal/Pneumonia, the inception V3 model, the ResNet 50 model, and the VGG19 model all had 97% detection accuracy.

## III. METHODOLOGY

This research employs Deep Learning Covid-19 prediction system through patient's X-Ray images by using ResNet 50 Model to validate the reliability and performance efficacy of the approach in terms of its automated prediction accuracy and sustenance. Using two different types of patient datasets—Normal as well as Covid-19 infected—the suggested model is trained and evaluated. Performance metrics used for model evaluation are, namely, Precision, Recall, F1-Score and Support.

- **Dataset**

Public Corona virus patient's repository named COVID-19_Radiography_Dataset available is used for the purpose of experimental analysis. The dataset is available in - https://www.kaggle.com/newsknowledge/covid-prediction-using-resnetmodel.
The dataset consists of 3616 files of COVID x-rays, 6012 files of lung opacity x-rays, 10,200 files of normal lung x-rays, 1345 files of viral Pneumonia x-rays.

- **Data Pre-processing**

Make directory- This creates a blank directory for the file system. This will store the data outside the program which can be referred in future. In this, the separate folders are created to store the data of COVID, normal and lung opacity.

Image Data Generator- the Image Data Generator class in Keras allows you to augment photos quickly and easily [12]. It includes a variety of augmentation methods like as standardisation, rotation, shifts, flips, and brightness changes, among others. The augmentation techniques used in this research are-

Random rotations- Giving the Image Data Generator class an integer value for the rotation range parameter allows you to flexibly rotate pictures between 0 and 360 degrees. There will be an empty area while rotating the picture since some pixels will travel outside of it [13]. There are several methods to fill this, such utilising a constant value, the closest pixel values, and so on. As a result of this default setting, the pixel values that are closest to each other are used to fill up any gaps. This research used a rotational range of 30 degrees.

Random flips- It's possible to flip an image either horizontally or vertically using the Image Data Generator class. As long as the object in the photograph matches what is shown in this technique, it is OK. The horizontal flip is used in this investigation.
Random zoom- The zoom range parameter of the Image Data Generator class allows a float value. Alternatively, you might list the lower and upper bounds in a two-valued list. Unless you specify a floating-point value, zooming will be restricted to the range. There's a maximum zoom level of 0.3.

Rescaling- For neural network models, the Image Data Generator class may be used to scale pixel values from 0-255 to 0-1. The technique of reducing data to the a range of 0-1 values is known as normalisation. [14]. Setting the rescale parameter to a ratio that allows each pixel to be multiplied to obtain the specified range does this.

Shear range- Shear' refers to a distortion of an image along an axis, usually to produce or correct perception angles. It's often used to enhance photographs so computers can see how people perceive things from various perspectives.

- **Data Analysis**

The data is analysed using the resnet50 architecture.

- **ResNet50 Architecture**

One Max Pool layer and one Average Pool layer make up the ResNet50 ResNet variation. It is capable of doing 3.8 x 109 floating-point calculations at once. The ResNets architecture was originally designed for image recognition tasks, but as noted in the article, it may also be used to boost accuracy in non-computer vision applications.

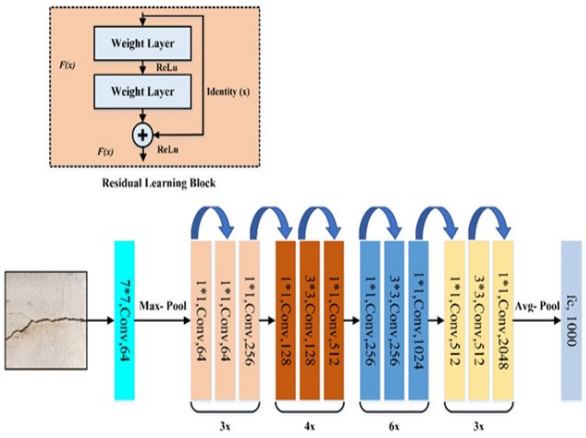



Figure 1: ResNet50 Architecture (https://www.researchgate.net/figure/The-architecture-of-ResNet-50-model_fig4_349717475)

The following elements are included in the Resnet 50 architecture [15]:

- One layer is produced using a convolution with a stride of size 2 and 64 unique kernels with a kernel size of 7 × 7.
- Next, a stride length of 2 for maximum pooling.
- • A 1 * 1, 64 kernel is preceded by a 3 * 3, 64 kernel, as well as finally a 1 * 1,256 kernel in the following convolution. For a total of nine layers in this phase, these three levels are repeated three times.
- 3 * 3,128, and then a kernel of 1 * 1,512 are the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:40:24 UTC from IEEE Xplore. Restrictions apply.

next three kernels. For a total of 12 layers, this procedure was performed four times.

- This is followed by six further iterations of the 1 * 1, 256, 3 * 3,256, and 1 * 1, 1024 kernels, for a total of 18 levels in the design.
- For a total of nine layers, further 3 x 3,512 and 1 x 1,2048 Kernels were added.
- This is followed by averaging the results of each pool, and then a softmax function and a 1000 node layer are added on top of that.

This totally gives 50-layer deep convolutional networks.

The proposed algorithm is implemented as follows:

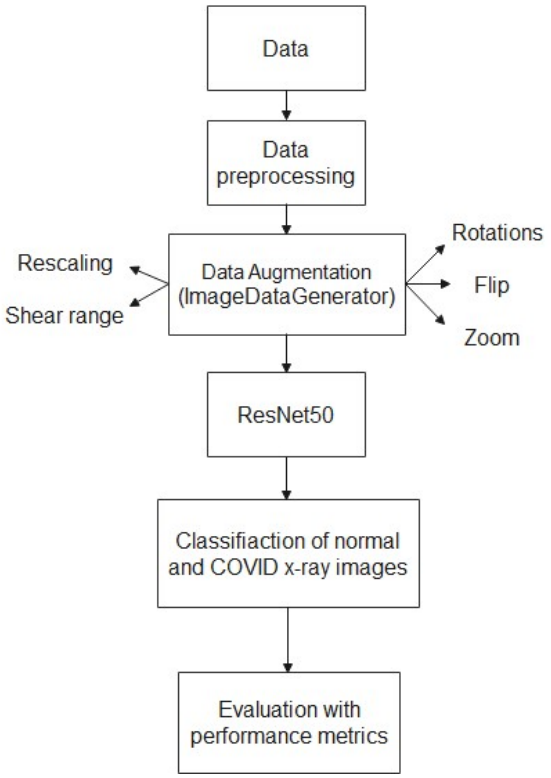



Figure 2: Flow chart of proposed algorithm

The following performance measures are used to assess the effectiveness of the suggested algorithm.

- **Performance metrics**

In order to assess an algorithm's performance in terms of the confusion matrix's performance measures, such accuracy, sensitivity, precision, as well as F1-score, are used.

Accuracy: It's the proportion of subjects successfully identified to a total number of topics.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Sensitivity: Recall, also known as sensitivity, is the proportion of correctly positive labels recognized by our computer.

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

Precision: It is possible to calculate the precision of an outlook by taking into account the total number of accurate forecasts. Predictive value is another title for this concept.

$$\Pr ecision = \frac{TP}{TP + FP} \quad (3)$$

F1-Score: The F1-score is a statistic that considers both accuracy and recall.

$$F1 - score = 2 * \frac{\Pr ecisison * \mathrm{Re}\, call}{\Pr ecision + \mathrm{Re}\, call} \quad (4)$$

Where,

TP= True Positive

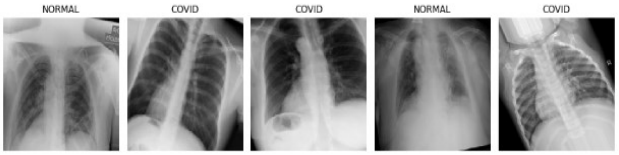
TN= True Negative

FP= False Positive

FN= False Negative

## IV. EXPERIMENTAL RESULTS

A graphic representation of the x-ray categorization findings may be seen in this part of the report. As part of this evaluation, the suggested model's performance indicators are also provided.



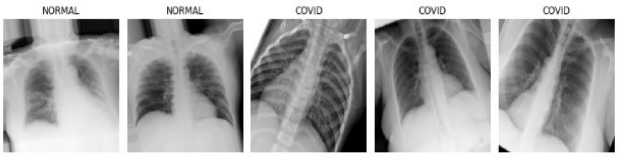




Figure 3: Sample dataset of X-ray images

The above figure is the sample pictures of the data and the labels. It is observable that the dataset consists of both normal and the COVID x-rays.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:40:24 UTC from IEEE Xplore. Restrictions apply.

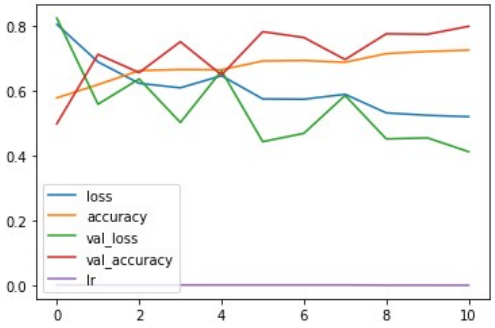



Figure 4: Performance metrics

The model's performance metrics were analysed and shown graphically in the picture above. The accuracy, loss, val_loss, val_accuracy and lr is been displayed. It is observable that the accuracy is higher than 0.8 which indicates the model is effective in categorizing the COVID patients and normal people's x-rays.

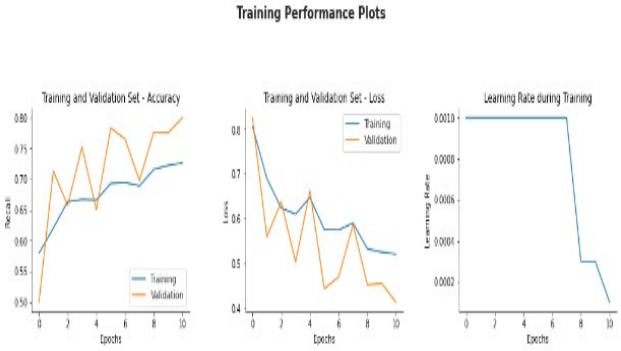



Figure 5: Training performance plots

In the graphs above, comparisons of training and validation are shown. The accuracy of the training and validation sets is shown in the first graph. It can be seen that, as compared to training, validation has a greater accuracy with the number of epochs increases. The loss is seen in the second graph. The validation sets have a lower loss than the training sets, as seen in the graph. As a consequence, validation is more effective in loss than training sets. The learning rate during training is seen in the third graph. The learning rate has been steadily falling throughout the epochs, and it has now hit its lowest point with epoch 10.

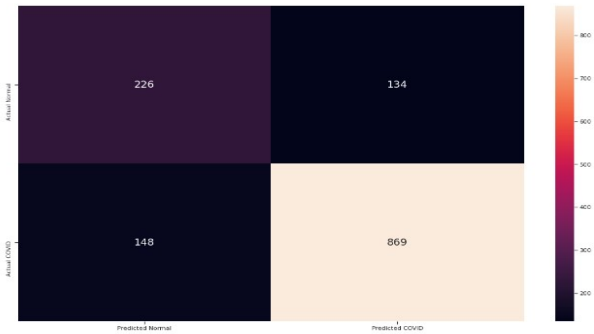



Figure 6: Confusion matrix

The suggested model's confusion matrix for the categorization of expected and actual COVID x-rays is shown above. The model correctly predicted the real normal x-rays 226 times in the dataset. The model correctly predicted the real sick x-rays 869 times. There are also some mis-interpretations where one class is been mistaken as other class.

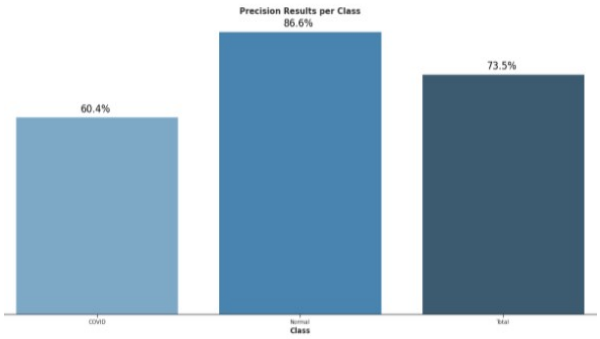



Figure 7: Precision result per class

The graph displays the precision results according to the class. From the figure, it is observable that Normal class is with high precision and COVID is comparatively with low precision value.

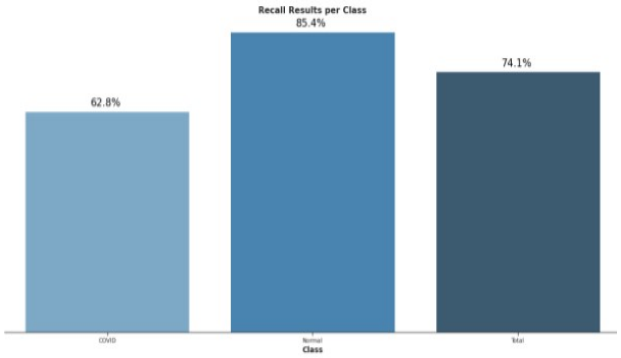



Figure 8: Recall results per class

The graph displays the results of recall according to the class. It is observable that normal class which is with 85.4% is with high recall value than COVID class which is with 62.8%.

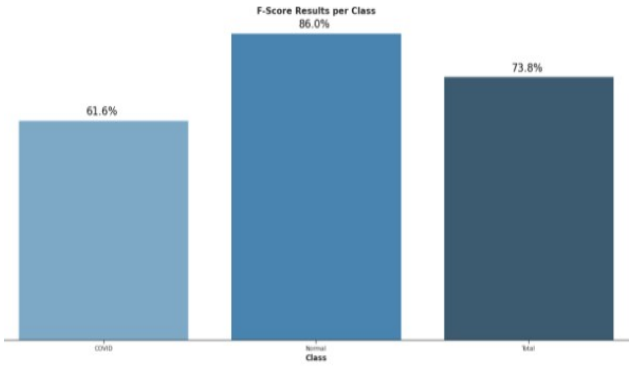



Figure 9: F-score results per class

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:40:24 UTC from IEEE Xplore. Restrictions apply.

The result of F-score per class is displayed in the above figure. The class normal is with 86% and this is the high in F-score when compared to COVID class which is with 61.6%.

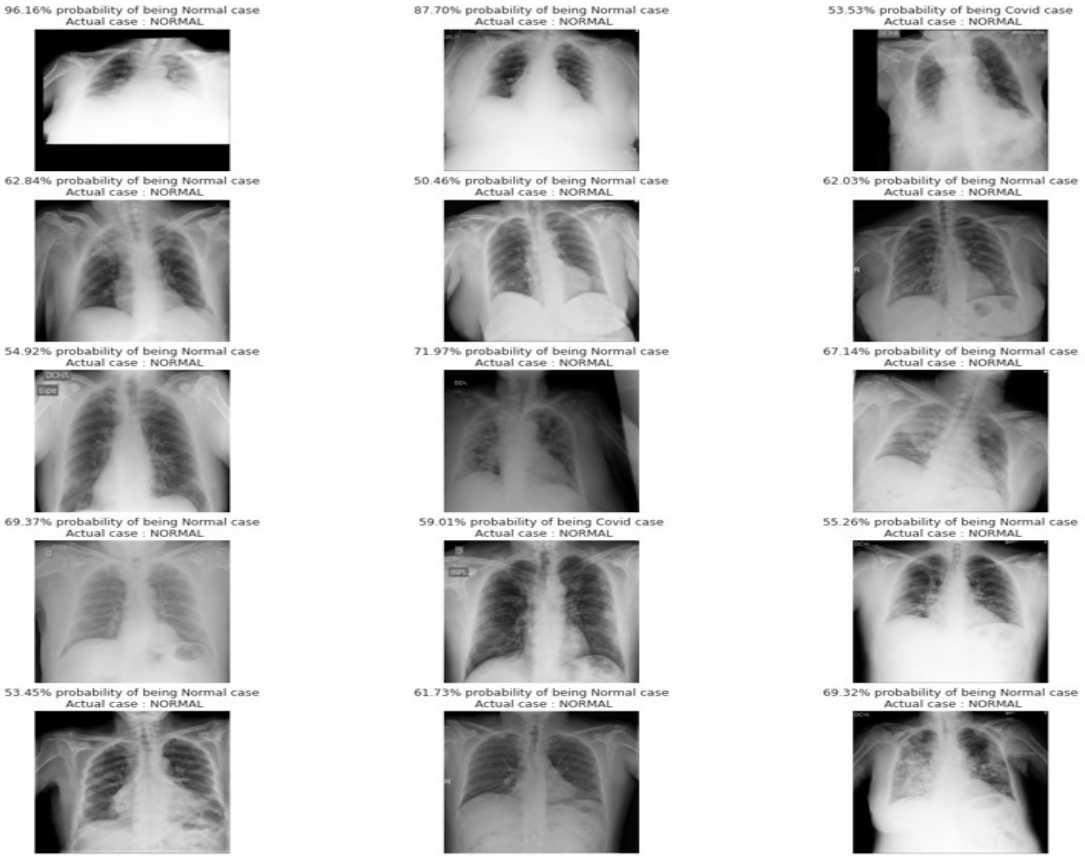



Figure 10: Probability of images

The above grid of images depicts the probability value of image being a normal case. The percentage of probability of being a normal case is been displayed above along with the actual case.

Here all the actual cases are normal. But the probability values of the model are differing from image to image. The maximum probability of model that had detected is 96.16% and minimum probability is 50.46%.

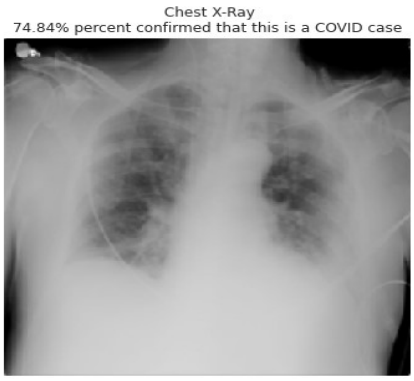



Figure 11: Input image of COVID x-ray

The model uses a COVID case X-ray as an input picture to identify the case. The model had 74.84% accurately predicted that the given input image is of a COVID case.

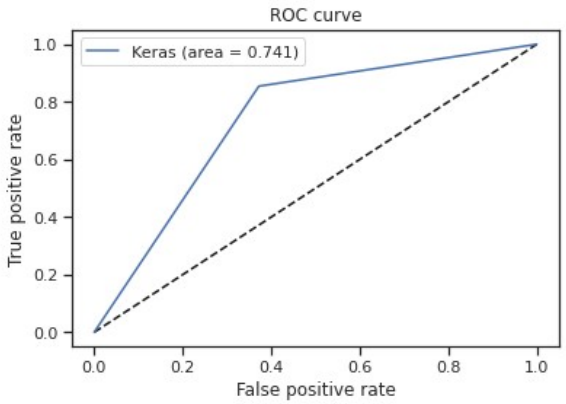



Figure 12: ROC curve

The following graph displays the true positive as well as false positive rate. The ROC curve is often used in binary classification to investigate the output of a classifier. The model's performance is determined by the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:40:24 UTC from IEEE Xplore. Restrictions apply.

area of the curve. The area value 1 is determined to be the ideal and 0.5 is the worst case of the model. The area of the model is 0.741, as shown in the above image, implying that the proposed model is suitable for the categorization procedure.

## V. CONCLUSION

Neural network-based algorithms are utilised to categorise normal and COVID patients using x-ray image data in this study. In the proposed model, a convolutional neural network approach called ResNet 50 architecture is employed for classification. The dataset consisting of normal, COVID and lung_opacity is used for the implementation. ImageDataGenerator is utilized for data pre-processing which rescales, flips, and adjusts the range to fit to the model. Further, the proposed algorithm ResNet50 architecture is imported to classify the x-ray images. Performance matrices like as precision, accuracy, recall, as well as F1-score are examined to demonstrate the algorithm's effectiveness. The suggested approach has an accuracy of 80%, indicating that it is extremely efficient in identifying COVID as well as normal x-ray pictures. This study will have a huge impact on real-time by accurately identifying the COVID in less time which possibly minimizes the fatality rate.

## REFERENCES


[1] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," *Appl. Intell.*, vol. 51, no. 2, pp. 854–864, 2021, doi: 10.1007/s10489-020-01829-7.

[2] S. Asif, Y. Wenhui, H. Jin, and S. Jinhai, "Classification of COVID-19 from Chest X-ray images using Deep Convolutional Neural Network," *2020 IEEE 6th Int. Conf. Comput. Commun. ICCC 2020*, pp. 426–433, 2020, doi: 10.1109/ICCC51575.2020.9344870.

[3] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *Pattern Anal. Appl.*, vol. 24, no. 3, pp. 1207–1220, 2021, doi: 10.1007/s10044-021-00984-y.

[4] R. M. Pereira, D. Bertolini, L. O. Teixeira, C. N. Silla, and Y. M. G. Costa, "COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios," *Comput. Methods Programs Biomed.*, vol. 194, 2020, doi: 10.1016/j.cmpb.2020.105532.

[5] J. I.-Z. Chen, "Design of Accurate Classification of COVID-19 Disease in X-Ray Images Using Deep Learning Approach," *J. ISMAC*, vol. 2, no. 2, pp. 132–148, 2021, doi: 10.36548/jismac.2021.2.006.

[6] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks," *Phys. Eng. Sci. Med.*, vol. 43, no. 2, pp. 635–640, 2020, doi: 10.1007/s13246-020-00865-4.

[7] A. Sharma, S. Rani, and D. Gupta, "Artificial Intelligence-Based Classification of Chest X-Ray Images into COVID-19 and Other Infectious Diseases," *Int. J. Biomed. Imaging*, vol. 2020, 2020, doi: 10.1155/2020/8889023.

[8] A. M. Ismael and A. Şengür, "Deep learning approaches for COVID-19 detection based on chest X-ray images," *Expert Syst. Appl.*, vol. 164, p. 114054, 2021, doi: 10.1016/j.eswa.2020.114054.

[9] H. Tahir, A. Iftikhar, and M. Mumraiz, "Forecasting COVID-19 via registration slips of patients using ResNet-101 and performance analysis and comparison of prediction for COVID-19 using faster R-CNN, mask R-CNN, and ResNet-50," *Proc. 2021 1st Int. Conf. Adv. Electr. Comput. Commun. Sustain. Technol. ICAECT 2021*, 2021, doi: 10.1109/ICAECT49130.2021.9392487.

[10] S. Misra, S. Jeon, S. Lee, R. Managuli, I. S. Jang, and C. Kim, "Multi-channel transfer learning of chest x-ray images for screening of covid-19," *Electron.*, vol. 9, no. 9, pp. 1–12, 2020, doi: 10.3390/electronics9091388.

[11] M. Elpeltagy and H. Sallam, "Automatic prediction of COVID− 19 from chest images using modified ResNet50," *Multimed. Tools Appl.*, vol. 80, no. 17, pp. 26451–26463, 2021, doi: 10.1007/s11042-021-10783-6.

[12] A. Munshi, A. Mehra, A. Choudhury, Y. Sahu, and A. Gupta, "Real-Time Face Mask Detection to Prevent COVID-19 in Confined Spaces," *Int. J. New Technol. Res.*, vol. 7, no. 11, 2021, doi: 10.31871/ijntr.7.11.18.

[13] S. Kawakura and R. Shibasaki, "Distinction of Edible and Inedible Harvests Using a Fine-Tuning-Based Deep Learning System," *J. Adv. Agric. Technol.*, vol. 6, no. 4, pp. 236–240, 2019, doi: 10.18178/joaat.6.4.236-240.

[14] B. Zoph, E. D. Cubuk, G. Ghiasi, T. Y. Lin, J. Shlens, and Q. V. Le, "Learning Data Augmentation Strategies for Object Detection," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12372 LNCS, pp. 566–583, 2020, doi: 10.1007/978-3-030-58583-9_34.

[15] D. Theckedath and R. R. Sedamkar, "Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks," *SN Comput. Sci.*, vol. 1, no. 2, pp. 1–7, 2020, doi: 10.1007/s42979-020-0114-9.


Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:40:24 UTC from IEEE Xplore. Restrictions apply.

*2022 3rd International Conference for Emerging Technology (INCET)*
*Belgaum, India. May 27-29, 2022*

# Accessible Polycystic Ovarian Syndrome Diagnosis Using Machine Learning

Akanksha Tanwar
*Dept. of Information Technology*
*Delhi Technological University*
New Delhi, India
akankshatanwar73@gmail.com

Anima Jain
*Dept. of Information Technology*
*Delhi Technological University*
New Delhi, India
anima.jain18@gmail.com

Anamika Chauhan
*Dept. of Information Technology*
*Delhi Technological University*
New Delhi, India
anamika@dtu.ac.in

2022 3rd International Conference for Emerging Technology (INCET) | 978-1-6654-9499-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/INCET54531.2022.9824049

***Abstract*—Polycystic Ovarian Syndrome also known as PCOS is an endocrine syndrome affecting women of reproductive age worldwide. It is one of the leading causes of infertility in women. PCOS is a complex syndrome that also causes insulin resistance in women - when the body can make insulin but can't use it effectively; which multiplies the risk of developing diabetes. Furthermore, women with PCOS often have higher levels of androgen (male hormones present in females), which can subsequently lead to anovulation (absence or lack of ovulation) and cause irregular periods, acne, thinning scalp hair, and excess hair growth on the face and body. As per the results of a large scale survey conducted across India in 2020, about 16 per cent of women respondents between the ages of 20 and 29 years suffered from polycystic ovary syndrome. PCOS, if left untreated can lead to serious health issues like heart diseases, acne scars, cancer and infertility. Despite the long term effects and fatal implications, the majority of women with PCOS remain undiagnosed. The taboo attached to the diagnosis of the syndrome clubbed with a lengthy initial diagnostic phase can be one of the many contributing factors to the alarming undiagnosed cases. To help bridge the gap of accessibility of PCOS diagnosis, in this paper we demonstrate a web app for women to easily predict their chance of suffering from the syndrome from the comfort of their homes, till help is available to them. We use minimal but optimal clinical features for the prediction of the same, by using feature selection techniques to cherry-pick the parameters which are easily measured at home and are non-invasive. We use the random-forest classifier which has shown promising results in the past for the diagnosis of PCOS using Machine Learning techniques. This paper uses the Kaggle dataset of 541 data points with 41 clinical attributes like weight gain, follicles in the left/right ovary, LHS levels, hair loss, BMI, pimples, etc.**

***Index Terms*—polycystic ovarian syndrome, diagnosis, machine learning, classifier**

## I. Introduction

With growing technology, we can shape up future of healthcare and make medical diagnosis tools readily available to everyone. Machine Learning can learn from the existing dataset and help predict the disease using some generic and easily measurable attributes. One such area where Machine Learning is being widely used in clinical decision support.

PCOS is a hormonal disorder causing enlarged ovaries with small cysts on the outer edges. Despite being a pervasive disorder, it often remains undiagnosed and unmanaged in most people who have it. Its most common symptoms include acne, hair growth on the face and body, infertility, weight gain and irregular periods. Lifestyle changes have been shown to be highly effective in managing PCOS. Management of PCOS is imperative to eliminate the chances of any future health risks.

Being a highly complex syndrome, it is still not understood what causes PCOS. PCOS is often marked by inflammation and insulin resistance and may involve environmental, behavioural, lifestyle and genetic factors. About 8 in 10 people who have PCOS have irregular cycles [1]. About 7 in 10 people with PCOS have high testosterone [1]. High testosterone causes symptoms including facial and body hair (hirsutism), hair loss on the head, and acne. Not everyone with PCOS develops small cysts in their ovaries—cysts are a symptom of PCOS, rather than a cause. PCOS is very common. It may affect roughly 1 in 12 women and people with cycles of reproductive age (8%, or 6-13%), but this likely varies across populations [2]–[5].

PCOS diagnosis requires an analysis of a combination of blood tests, ultrasound reports and physical factors. Gynaecologists diagnose a patient to be having PCOS if any two of the following three factors are satisfied:

Irregular cycles, especially ones that are long or absent, Evidence of androgen excess (hyperandrogenemia) such as excess body hair or hair thinning on the scalp and/or high levels of androgenic hormones (like testosterone) in the blood and/or having a number of cyst-like sacs in the ovaries [2]–[4]

Some of the symptoms such as absent or infrequent menstrual periods can be present due to health conditions such as a thyroid disorder (an over-or underactive thyroid gland). Thus, many tests including an assessment of thyroid hormone levels, as well as levels of the hormone prolactin (which is high in people with hyperprolactinemia—another cause of infrequent or absent periods), might be recommended by many healthcare providers during PCOS diagnosis. These kinds of tests may be imperative for people who otherwise show no visible symptoms of PCOS. Issues with prolactin or thyroid hormones can also overlap in some people, and one condition can contribute to another, which is why tests may be important in getting a diagnosis.

PCOS diagnosis can be expensive and cause mental agony to patients. In developing countries like India, where sex education and female health issues are viewed as a taboo subject, it is common for women suffering from PCOS to remain



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:27:23 UTC from IEEE Xplore. Restrictions apply.

undiagnosed for long periods of time. Our web application attempts to make the initial stage of diagnosis available to all, in an easy and non-invasive manner. It aims to promote early diagnosis in women who are at a greater risk of being diagnosed with PCOS.

The paper is structured into the following sections: Previous/Related Works, Methodology of Proposed Approach, Results, Web Application for PCOS Detection, Conclusion and References.

## II. PREVIOUS/RELATED WORKS

Machine Learning techniques have be widely used to predict and aid clinical decision making, one such great example has been the amount of work during the ongoing pandemic which helped in early detection of the disease and facilitated treatment [17].

One of the most crucial processes is data collection and formation. It is available for download via well-known repository hubs such as Kaggle and UCI datasets. The model's accuracy, and hence its performance, will be determined during this phase, and as the number of data collected increases, so will the model's accuracy. The category of different kinds of data used in the past for conduction PCOS research can be divided into two categories: Clinical Dataset and Ultrasound Image Dataset.

The PCOS clinical features dataset available on Kaggle has been widely used. This dataset has 43 attributes of 541 women, among which 177 are patients of PCOS disease [8], [10], [13]. On the other hand, ultrasound images are also used for the classification/prediction of PCOS. Bedy Purnama et al. used two datasets: dataset A with 40 images that consist of 26 normal images and 14 PCOS-indicated images and dataset B with 40 images consisting of 34 normal images and 6 PCOS-indicated images [12]. Vikas B, Radhika Y, and Vineesha K used a dataset consisting of 100 images i.e., 50 images of PCOS and non-PCOS each is trained over a total of 25 epochs and validated simultaneously on the validation set which consists of 50 images i.e., 25 images of PCOS and non-PCOS each [15]. Prema T. Akkasaligar and Girijamma V. Malagavi used the data set consisting of 26 normal ovary images and 28 cystic ovary images [16]. Out of 54 images, 44 images are used for the training phase and 10 images are used for the testing phase [16].

In the existing work in the field of Polycystic Ovary Syndrome prediction, there is a mix of simple linear and non-linear algorithms in the algorithms chosen. Aroni Saha Prapty and Tanzim Tamanna Shitu had a look at a variety of statistical models [10]. Amsy Denny, Anita Raj, Ashi Ashok, Maneesh Ram C. and Remya George estimated accuracy scores of each model were obtained using hold-out validation, and Cheng and Mahalingaiah used two approaches: rules-based classifier (RBC) and gradient boosted tree method (GBT) on ultra-sound images [8], [9]. We note that the Random Forest Classifier model delivered the best results on the clinical dataset in two of the papers, with an accuracy of 89% after data optimization in [8] and 93.5% accuracy in [10]. Cheng and Mahalingaiaha's work employs RBC on using data from ultrasound images with a 97% accuracy [9]. However, the ultrasound pictures' limitations and highly precise dataset needs must not be neglected. Lele, Priyanka R., and Anuradha D. Thakare used statistical methods analysis, in which the K-star algorithm performed the best [11].

The following are some of the algorithms that have been employed. The Random Forest Classifier is found to be superior in clinical aspects [8], [10]. This automated technique, together with ultrasound report analysis, can help doctors save time when assessing patients and diagnosing PCOS risk more quickly. In the future, studies on the influence of Vitamin D on PCOS, studies on the impact of PCOS on preterm labour/abortions, and attempts to uncover the number of normal-weight PCOS patients, among other topics, could be held. The summary of previous work is presented in Table I.

## III. METHODOLOGY

In this paper, we present a diagnosis system that will aid in determining women at a greater risk of being diagnosed with PCOS. This system comprises the following steps as shown in Fig. 1.

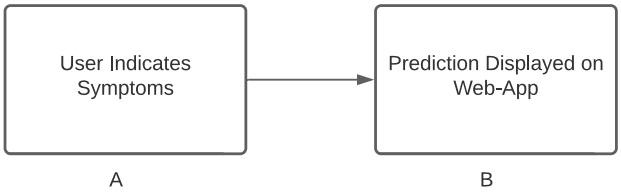



Fig. 1. Stages of the Proposed System

### A. Understanding the Features

The dataset is highly domain-specific in terms of medical and biological terminologies. Hence, it becomes imperative to first study the meaning of these terms. These features once understood can then be used to determine the feature selection process in terms of accessibility. For instance, features such as follicle numbers require an ultrasound and LH, FSH, HCG hormones require a blood test whereas weight gain, acne, menstrual cycle regularity can be checked without any external help.

Combined with accessibility in terms of the selected features being easily measurable at home it then becomes extremely important to consider the diagnostic accuracy of the selected features in the prediction of PCOS.

### B. Data Collection

Data collection of relevant features can be done by accumulating various patients' clinical information, both suffering from the syndrome and not suffering from it. For our study, we used a Kaggle dataset containing data from 10 different hospitals in Kerala, India. The dataset has 44 different features



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:27:23 UTC from IEEE Xplore. Restrictions apply.

TABLE I
RESULT OF SOME PREVIOUS WORK IN THE FIELD OF POLYCYSTIC OVARIAN SYNDROME PREDICTION

| S.No. | Research Paper | Models Used | Accuracy |
|---|---|---|---|
| 1. | A Classification of Polycystic Ovary Syndrome Based on Follicle Detection of Ultrasound Images | Neural Network – LVQ test<br>KNN - Euclidean Distance Test<br>SVM Test | 74.63%<br>80.73%<br>82.55% |
| 2. | Diagnosis of Polycystic Ovary Syndrome Using Machine Learning Algorithms | Gradient Boosting<br>Random Forest<br>Logistic Regression<br>RFLR | 88.92%<br>88.24%<br>88.53%<br>91.01% |
| 3. | An Efficient Decision Tree Establishment and Performance Analysis with Different Machine Learning Approaches on Polycystic Ovary Syndrome | K- nearest neighbors<br>Support Vector Machine<br>Naive<br>Random Forest | 75%<br>90%<br>93%<br>93.5% |
| 4. | i-HOPE: Detection And Prediction System For Polycystic Ovary Syndrome (PCOS) Using Machine Learning Techniques | Logistic Regression<br>K- nearest neighbors<br>Classification and Regression Tree<br>Random Forest Classifier<br>Gaussian Naive Bayes<br>Support Vector Machine | 85.63%<br>86.58%<br>82.92%<br>89.02%<br>84.14%<br>82.92% |

ranging from clinical measurements such as cycle regularity, follicle numbers in the left and right ovaries, weight gain, acne issues, etc. Overall the dataset has patient information of 541 different patients. It is important to note the rather scarce availability of datasets for PCOS research.

### C. Data Preparation and Pre-Processing

The following tasks were performed to pre-process the available dataset:

- Removed unnecessary columns such as Serial Number and Patient File Number
- Checked for any duplicate rows; none of the rows was duplicate in nature
- Filled the 'NAN' values with mean value of the respective feature
- Division of the dataset into features and target label: the target label being 0 or 1 indicating whether PCOS has been detected or not

### D. Feature Selection

In our study, feature selection plays the most important role in picking up the best features that are easy to measure as well. Feature selection is a widely used technique that reduces the dimension of a dataset with respect to a given value [6]. Selecting a subset of potential inputs from the total feature set can help in compacting relevant information, and removing the noise in prediction [7]. This helps the machine learning model to improve the learning process and its generality as well.

Thus, it is always suggested to implement feature selection to reduce the number of features the model has to deal with. For building a quick and easy to use application we will be selecting the top five easy to measure features at home. Our study implements and analyses three of the well-known feature selection techniques:

- Chi-Square Method
- Extra Tree Classifier
- Correlation Matrix

### E. Chi-Square Method

The Chi-Square test or the Chi-Square method is implemented to gauge the independence of two features in terms of relationships and see if the features in question are related to each other. The formula for the Chi-Square value is defined as follows:

$$\tilde{\chi}^2 = \sum_{k=1}^{n} \frac{(O_k - E_k)^2}{E_k} \tag{1}$$

In (1), O is the observed value(s) and E is the expected value(s).

While using the Chi-Square method for feature selection we are looking for features with high dependence upon the target variable. Thus, we are looking for features that produce a lower Chi-Square value, which indicates lower independence, or in other words higher dependence of the target variable on the feature it is testing against. Table II highlights the top twenty features obtained from the Chi-Square feature selection method.

We note that the five top features that can also be measured at home are:

- Skin Darkening
- Hair Growth
- Weight gain
- Weight
- Fast Food

Applying the Random Forest Classifier on only these five features from the dataset gives an accuracy of 85.18%.

### F. Extra Tree Classifier

Also known as Extremely Randomised Tree Classifier aggregates results from multiple de-correlated decision trees. It implements a meta estimator that fits a number of randomized decision trees (a.k.a. extra-trees) on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting [7].



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:27:23 UTC from IEEE Xplore. Restrictions apply.

TABLE II
TOP-20 FEATURES BASED ON CHI-SQUARE METHOD

| S.No. | Features |
|---|---|
| 1 | Vit D3 |
| 2 | I beta-HCG |
| 3 | LH |
| 4 | FSH |
| 5 | II beta-HCG |
| 6 | Follicle No. (R) |
| 7 | Follicle No. (L) |
| 8 | AMH |
| 9 | FSH/LH |
| 10 | Skin Darkening |
| 11 | Hair Growth |
| 12 | Weight Gain |
| 13 | Weight |
| 14 | Fast Food |
| 15 | Cycle(R/I) |
| 16 | PRG |
| 17 | Pimples |
| 18 | Marriage Status |
| 19 | BMI |
| 20 | Age |

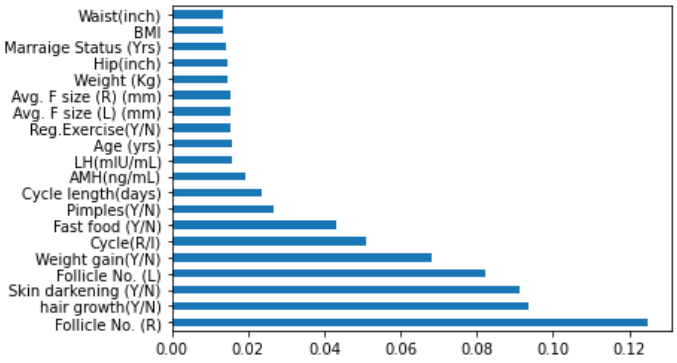



Fig. 2. Top-20 Features Based on Extra Tree Classifier Method

In Fig. 2 the top twenty features obtained from the Extra Tree Classifier method are depicted. Where X-axis depicts the feature importance scores and the Y-axis lists down the features from the dataset.

We can observe features such as hair growth, skin darkening, weight gain and fast-food are overlapping with our result from the previous run of the Chi-Square method.

We note that the five top features that can also be measured at home are:

- Skin Darkening
- Hair Growth
- Weight gain
- Cycle
- Fast Food

Applying the Random Forest Classifier on the dataset containing only these five features gives an accuracy of 92.59%.

### G. Correlation Matrix

It is a matrix used for calculating correlation amongst every feature against every other feature using statistical features. Every cell in Fig. 3 represents the correlation between two variables. A correlation matrix is used to summarize data, as an input into a more advanced analysis, and as a diagnostic for advanced analyses. Thus, the higher the magnitude of the correlation between a feature and the target variable the better is the feature suited for accurate diagnosis. The ten most relevant features are then listed in Table III in descending order of importance.

TABLE III
TOP-10 FEATURES BASED ON EXTRA TREE CLASSIFIER

| S.No. | Features |
|---|---|
| 1 | Follicle No. (R) |
| 2 | Hair Growth |
| 3 | Skin Darkening |
| 4 | Follicle No. (L) |
| 5 | Weight Gain |
| 6 | Cycle(R/I) |
| 7 | Fast Food |
| 8 | Pimples |
| 9 | C Length |
| 10 | AMH |

We note that the five top features that can also be measured at home are:

- Skin Darkening
- Hair Growth
- Weight gain
- Cycle
- Fast Food

These features are overlapping with Extra-Tree Classifier Method and gives the same accuracy of 92.59% when Random Forest Classifier is implemented.

## IV. RESULTS

TABLE IV
RESULT COMPARISON OF FEATURE EXTRACTION TECHNIQUES

| S.No. | Feature Extraction Technique | Machine Learning Model | Accuracy |
|---|---|---|---|
| 1 | Chi-Square Method | Random Forest Classifier | 85.15% |
| 2 | Correlation Matrix | Random Forest Classifier | 92.59% |
| 3 | Extra Tree Classifier | Random Forest Classifier | 92.59% |

- As shown in Table IV, we observed the best results for Random Forest Classifier with the features extracted from both Correlation Matrix and Extra Tree Classifier.
- After comparing three different statistical methods to pick out the best features that can be used for the detection of PCOS, and that are easily measurable at home, are the ones listed below:
  - Skin Darkening
  - Weight Gain
  - Hair Growth
  - Cycle (R/I)
  - Fast Food
- After implementing the Random Forest Classifier to diagnose Polycystic Ovary syndrome using the above-mentioned features shows an accuracy of **92.59%**.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:27:23 UTC from IEEE Xplore. Restrictions apply.

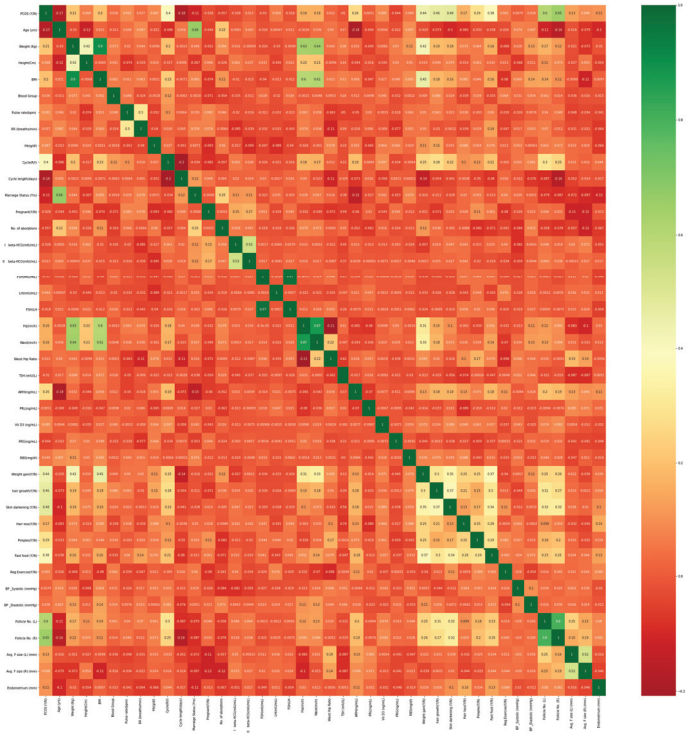


Fig. 3. Correlation Matrix of the PCOS Dataset

## V. Web Application

Our aim is to make PCOS pre-diagnosis easily accessible to every woman via a web application. The features extracted through feature selection techniques are easily measurable at home and can predict the presence of PCOS at high accuracy. The app uses five very easily measurable parameters, filtered using feature extraction to predict PCOS so that women don't require to undergo any clinical tests initially. Many women are very hesitant to go through PCOS diagnosis because of the invasive procedures and the requirement for many clinical tests.

As depicted in Fig. 4, The web app takes the input for the following five features: Skin Darkening(Y/N), Hair Growth(Y/N), Weight Gain(Y/N), Period Cycle (R/I), and Fast Food(Y/N). All of these fields require boolean values with input as Yes/No. 1 stands for "YES" and 0 stands for "NO". The user needs to input whether they are experiencing the corresponding symptom or not and click on the predict button.

### A. Working and Architecture

As we can observe in Fig. 5, working of web app can be divided into three parts:

- *Python model* - We are using the Random Forest Classifier model to predict Polycystic Ovary Syndrome. The pickle module in Python helps you to load the trained model without having to rewrite everything or train the model all over again.
- *Flask* - This is used for integrating machine learning models with the user interface. Here we load the previously created pickle file, and then use it for prediction. We take input from the user through the UI and then fetch it using REST API. These values are sent to the pickle model which finally returns the prediction as 1/0. 1 signifies that the user has a significant probability of having PCOS and 0 signifies that the user has a lesser probability of having PCOS.
- *User Interface* - The app uses HTML to design a very user-friendly interface. The user can simply give input as 1/0 for all the five above mentioned parameters and click on predict. The result will be displayed at the top of the page.
  If the user has significant chances of having PCOS, the app will display "You have significant chances of having PCOS. Please check with your doctor as soon as possible." And if the user has comparatively lesser chances of suffering from PCOS, it will show- "You do not have significant chances of having PCOS. However, if you are experiencing some other symptoms or feeling that you need a diagnosis, it is good to bring it up at your next doctor's appointment."

## VI. Conclusion

This study aims to make Polycystic Ovarian Syndrome diagnosis as simple as possible for all women. The five most important and easily measurable features for PCOS prediction are picked to make this aim achievable. Furthermore, the Random Forest Classifier used in the study predicts PCOS with **92.59%** accuracy. The free web application developed



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:27:23 UTC from IEEE Xplore. Restrictions apply.

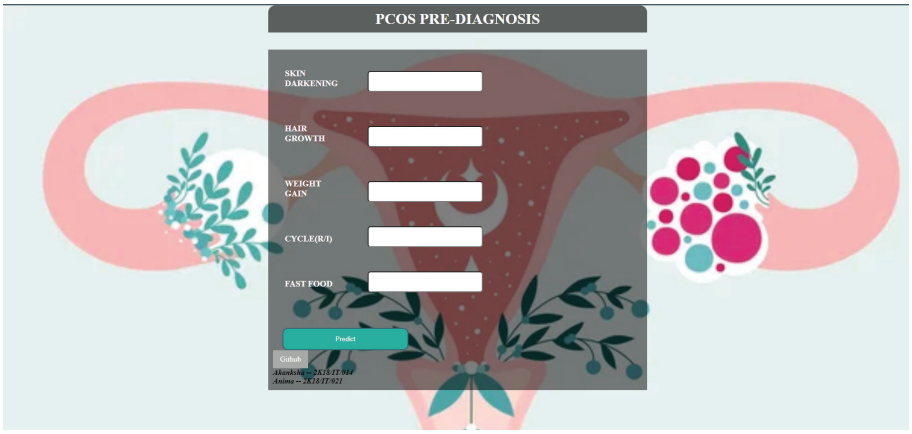



Fig. 4. User Interface of Web Application

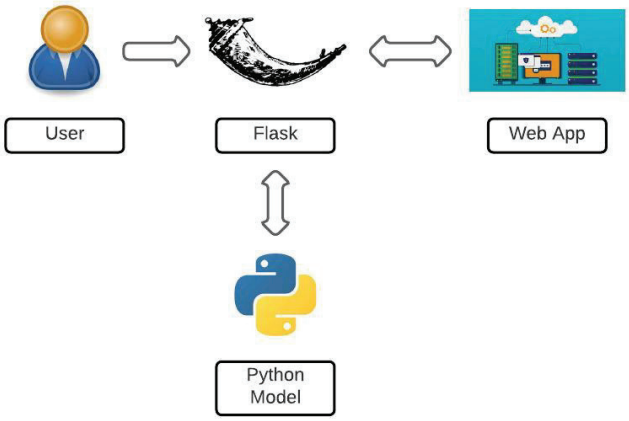



Fig. 5. Architecture of Web Application

for PCOS diagnosis can serve as a pre-diagnostic tool and in the future can be extended to even direct women in need to different resources such as nearest gynaecologists, best practices, online consultations, government health schemes, etc.

## REFERENCES


[1] Dumesic, Daniel A., and Rogerio A. Lobo. "Cancer risk and PCOS." Steroids 78, no. 8 (2013): 782-785.

[2] Bozdag G, Mumusoglu S, Zengin D, Karabulut E, Yildiz BO. The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis. Hum Reprod. 2016 Dec;31(12):2841-2855. doi: 10.1093/humrep/dew218. Epub 2016 Sep 22. PMID: 27664216.

[3] Groot, P Dekkers, Olaf Romijn, Johannes Dieben, Sandra Helmerhorst, F.M.. (2011). PCOS, coronary heart disease, stroke and the influence of obesity: A systematic review and meta-analysis. Human reproduction update. 17. 495-500. 10.1093/humupd/dmr001.

[4] Christensen SB, Black MH, Smith N, Martinez MM, Jacobsen SJ, Porter AH, Koebnick C. Prevalence of polycystic ovary syndrome in adolescents. Fertil Steril. 2013 Aug;100(2):470-7. doi: 10.1016/j.fertnstert.2013.04.001. Epub 2013 Jun 5. PMID: 23756098; PMCID: PMC3813299.

[5] March WA, Moore VM, Willson KJ, Phillips DI, Norman RJ, Davies MJ. The prevalence of polycystic ovary syndrome in a community sample assessed under contrasting diagnostic criteria. Hum Reprod. 2010 Feb;25(2):544-51. doi: 10.1093/humrep/dep399. Epub 2009 Nov 12. PMID: 19910321.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8] Denny, Amsy, Anita Raj, Ashi Ashok, C. Maneesh Ram, and Remya George. "I-HOPE: detection and prediction system for polycystic ovary syndrome (PCOS) using machine learning techniques." In TENCON 2019-2019 IEEE Region 10 Conference (TENCON), pp. 673-678. IEEE, 2019.

[9] Cheng, J.J. and Mahalingaiah, S., 2018. Data mining and classification of polycystic ovaries in pelvic ultrasound reports. bioRxiv, p.254870.

[10] Prapty, Aroni Saha, and Tanzim Tamanna Shitu. "An Efficient Decision Tree Establishment and Performance Analysis with Different Machine Learning Approaches on Polycystic Ovary Syndrome." In 2020 23rd International Conference on Computer and Information Technology (ICCIT), pp. 1-5. IEEE, 2020.

[11] Lele, Priyanka R., and Anuradha D. Thakare. "Comparative Analysis of Classifiers for Polycystic Ovary Syndrome Detection using Various Statistical Measures."

[12] B. Purnama, U. N. Wisesti, Adiwijaya, F. Nhita, A. Gayatri and T. Mutiah, "A classification of polycystic Ovary Syndrome based on follicle detection of ultrasound images," 2015 3rd International Conference on Information and Communication Technology (ICoICT), 2015, pp. 396-401, doi: 10.1109/ICoICT.2015.7231458.

[13] S. Bharati, P. Podder and M. R. Hossain Mondal, "Diagnosis of Polycystic Ovary Syndrome Using Machine Learning Algorithms," 2020 IEEE Region 10 Symposium (TENSYMP), 2020, pp. 1486-1489, doi: 10.1109/TENSYMP50017.2020.9230932.

[14] Cheng, J. Jojo, and Shruthi Mahalingaiah. "Data mining and classification of polycystic ovaries in pelvic ultrasound reports." bioRxiv (2018): 254870.

[15] Vikas, B., Y. Radhika, and K. Vineesha. "Detection of Polycystic Ovarian Syndrome using Convolutional Neural Networks." Int J Cur Res Rev— Vol 13, no. 06 (2021): 156.

[16] Akkasaligar, Prema T., and Girijamma V. Malagavi. "Detection of cysts in medical ultrasound images of the ovary." In Proc. SARC-IRF 5th International Conference Bangalore, India. 20

[17] Chaturvedi, Kunal, Tanmay Kansal, Saksham Gupta, Dinesh Kumar Vishwakarma, and Naokant Deo. "COVID-19 Severity Assessment from Chest X-rays using Attention-based Weakly-Supervised Learning." In 2021 2nd International Conference for Emerging Technology (INCET), pp. 1-5. IEEE, 2021.




Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:27:23 UTC from IEEE Xplore. Restrictions apply.

**International Research Journal of Engineering, IT & Scientific Research**
Available online at https://sloap.org/journals/index.php/irjeis/
Vol. 7 No. 3, May 2021, pages: 76-86
ISSN: 2454-2261
https://doi.org/10.21744/irjeis.v7n3.1003

# Advances in Applications of Non-Destructive Testing (NDT): A Review

**Mridul Gupta** [a]
**Muhsin Ahmad Khan** [b]

***Article history:***

Submitted: 27 January 2021
Revised: 18 February 2021
Accepted: 9 March 2021

**Keywords:**

*acoustic emission;*
*eddy current testing;*
*manufacturing defects;*
*non-destructive testing;*
*ultrasonic testing;*

**Abstract**

Manufacturing defects and discontinuities in a product are anomalies which can lead to severe damages which may sometimes involve loss of life. These defects must be examined and corrected before the product goes into service. There are two methods of testing a product for defects and discontinuities viz. Destructive Testing and Non-Destructive Testing. Destructive Testing entails subjecting the product to conditions that leads to failure of the product whereas Non-Destructive Testing (NDT) is the process of examining the products for defects in a way in which it retains its usefulness for future service. This paper discusses various methods involved in NDT such as Visual Testing, Magnetic Particle Inspection, Penetrant Testing, Ultrasonic Testing, Radiographic Testing, Acoustic Emission and Eddy Current Testing with a major focus on advances in the applications of NDT.

*International research journal of engineering, IT & scientific research © 2021.*
*This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/).*

***Corresponding author:***
Gupta, M.
Department of Mechanical Engineering, Delhi Technological University, Bawana Rd, Shahbad Daulatpur Village, Rohini, Delhi, 110042
*Email address: mridulgupta_2k18me131@dtu.ac.in*

[a] Department of Mechanical Engineering, Delhi Technological University, Bawana Rd, Shahbad Daulatpur Village, Rohini, Delhi, 110042
[b] Department of Mechanical Engineering, Delhi Technological University, Bawana Rd, Shahbad Daulatpur Village, Rohini, Delhi, 110042 Email: muhsinahmadkhan_2k18me132@dtu.ac.in

## 1 Introduction

One of the key objectives of the manufacturing industry is to ensure product quality i.e. to make sure the product is capable of performing its intended function for a long period of time. For a product to achieve this criteria, it is of utmost importance that the product be free of any defect and discontinuity. A defect is an imperfection in a product which can restrict its life by impeding the product's performance eventually causing it to fail, even during operation under ordinary conditions. A discontinuity is similar to a defect, the only difference being that a discontinuity can be rectified whereas a defect cannot. These anomalies can arise during manufacturing stages or even during the operation of the product. It is critical for these anomalies to be detected, studied and if possible be rectified. Non-Destructive Testing or NDT consists of a range of techniques which enable the detection and evaluation of said anomalies in such a way that the product's condition is not exacerbated and can be rectified if possible. One of the greatest advantages of NDT is that they can be performed at all stages of product life, even during their operation. Not only does this provide an accurate idea about the location but also the properties (like dimensions, rate of propagation etc.) and the severity of the defect or discontinuity (Jolly et al., 2015; Gholizadeh, 2016).

The principles employed in various techniques of NDT are also employed in different fields. It is an implication that NDT is not only limited to manufacturing. For example, principles of NDT techniques are prevalent in medical diagnostics. X-Radiography used for detecting tumours, fractures etc and Remote Visual Testing used in procedures like endoscopy are some of the examples of NDT in the medical field. However, today major applications of NDT in manufacturing lie in automobiles, oil and gas, aerospace etc. There are other fields that use the principles of NDT like geological exploration, architecture etc. This paper aims at co-relating principles of various NDT techniques with advances in applications in fields which may or may not be manufacturing oriented.

## 2 Literature Review

According to P Cawley (2001) eddy current, ultrasonic, magnetic particle, radiography and penetrant testing are the 'big five' techniques that dominate the NDT market. Thus, research is being done continuously to reduce the preparation required, increase the speed of inspection and to inspect without shutting down the operation. How the process of using previous cases of data-interpretation to interpret the new data can be helpful in solving the issue of NDT data interpretation is shown by Jacek Jarmulak and others Jacek Jarmulak, Eugene J.H. Kerckhoffs, Peter Paul van't Veen (2001). The article shows how Case Based Reasoning (CBR) can be a successful alternative to expert systems and statistical classifiers.

Li Zhijun (2001) states that common defects and damages in composites can be effectively checked by the ultrasonic inspection method as compared to other methods of NDT. The author also introduces the development of studying and the applications of the ultrasonic method. According to G.S. Park and others G. S. Park, P. W. Jang, Y. W. Rho (2001), in a Magnetic Flux Leakage (MFL) type NDT system, the defect signals depend on the modification of the magnetic flux leakage in the area of defect. Thus, the authors describe a method to design a magnetic system which maximises the MFL in a NDT system.

P. Tschelisnig (2001) presents the principles of hyperbolic triangulation and the basis of propagation of waves. The author also explains the AE inspection systems by taking an example from TUV Vienna's 32 channel software and equipment. The results gained after the tests performed on the TUV Vienna are discussed in the integrity analysis and leakage tests. H. A. Crostack, W. Reimers ( 2001) state that the ability to measure internal and induced stresses or to differentiate between structures of materials can be provided by the Non-destructive testing. This inturn provides the data for the calculation of potential lifetime and reliability. They also state that the materials may be better understood by observing fast moving processes like crack propagation when they are under impact load.

Giovanni M. Carlomagnoet, Carosena Meola (2002) examines the application of infrared thermography in the field of architectural restoration by taking 3 samples made of a support of tuff, marble or brick and covering them with a layer of plaster. This layer of plaster is given inclusions to simulate cracks and detachments in frescoes. To detect the artificially induced flaws various techniques such as lateral heating thermography, pulse thermography, pulse phase thermography and modulated thermography are used. K. Bollas, D. Papasalouros, D. Kourousis & A. Anastasopoulos (2011) discuss the monitoring of defects in the wheel sets of the trains using the Acoustic Emission (AE) monitoring method of NDT. The authors present the results of the AE measurements performed on the wheel sets of the trains with and without the defects, with varying speed and directions. The conclusions drwan show that

the AE monitoring method can be an efficient method for monitoring the wheel sets of the trains. Opportunities for the inspection of rock bolt in application such as reinforcement of coal mine roof, are limited to destructive techniques like pull-out test only. Thus, M.D. Beard and others Beard, M.J.S. Lowe (2003) try to develop a NDT technique for monitoring the condition of rock bolts.

Bruce W. Drinkwater, Paul D. Wilcox (2006) reviews the most relevant works on the use of ultrasonic arrays for non destructive evaluation applications in addition to which the author also discusses its future direction. The author also reviews the most relevant papers on the application of ultrasonic arrays in medical and sonar fields. On the other hand Malcolm K. Lim , Honggang Cao explain how one NDT technique is not sufficient for evaluation. Every technique has certain limitations which restrict their efficacy under different situations. The authors do so by differentiating the results between UT and ultra response method for evaluation of defects in a concrete structure.

A disadvantage of NDT is that most of the evaluations are heavily dependent upon the inspector's knowledge and experience. This is most relevant to magnetic particle inspection and penetrant testing, which also happen to be most used inspection techniques. Luk and Alan, H. S. Chan (2007) research on some of the human factors which may affect the accuracy of an evaluation and provide some suggestions in order to tackle these issues.

Gabriel Dan Taşcă, Gheorghe Amza explain how evaluation of complex parts can be done using NDT. Although x-ray provides accurate data about a defect and is mostly preferred over other techniques, sometimes UT can provide better results. The authors also provide better insight as to how UT can provide more information other than just the nature of the defect or discontinuity but also how it occurred. Francesca Cau, Alessandra Fanni Augusto Montisci, Pietro Testoni, Mariangela Usai (2006) worked on detection of defects in pipes using UT. They evaluated the problem with the help of appropriate Artificial Neural Network models. Numerical Techniques like FEM have been used to evaluate problems in pipes by studying the echoes produced by ultrasonic waves.

Eddy Current testing has many variables that affect evaluations made by this method. Javier García-Martín, Jaime Gómez-Gil and Ernesto Vázquez-Sánchez (2011) provide a summary on these variables as well as recent advancements in this mode of NDT. The authors also explain how development in electronics and data evaluation systems have made Eddy current testing more effective. One of the advantages of NDT is that it can be done simultaneously while operating a machine. For example, Rainer Pohl , A Erhard, H.J Montag , H.-M Thomas, H Wüstenberg (2004) illustrated that Eddy current testing can be done on rail tracks simultaneously to trains travelling on them. The authors also emphasized other applications of NDT such as use of ultrasonic testing on railroad gauges and wheels. Umesh Singh, Mohan Singh and M. K. Singh (2012) illustrate the application of magnetic particle separation techniques used in the Indian mining industry. The authors provide an insight as to how this process has made inspection simple and cost effective as well as discuss the disadvantages of this technique.

Advancements in NDT have resulted in better evaluation of defects and discontinuities. D. Bracun, I. Polajnar, J. Diaci (2006) proposed a laser based technique for 3-D modelling of defects which use advanced algorithms to produce 3-D images with the help of electronic visual aids such as cameras. Most of the principles employed in NDT techniques are employed in other fields as well. In fact some of these techniques can be incorporated in improvising existing technology. One such example has been illustrated by Rodrigo Benenson, Stephane Petti, Thierry Fraichard, Michel Parent (2008) where they emphasize on the issues with self navigating cars. A combination of techniques such as ultrasonic waves and thermography can assess surrounding environments more accurately thus increasing reliability.

Another advantage of NDT is that it can be conducted at any stage of production. Lattice defects which are microscopic in nature, arise during material manufacture and affect overall physical integrity of any component. Gerd Dobmann, Iris Altpeter, Bernd Wolter (2008) show that NDT if incorporated at every stage of material production helps in making sure that these properties are uniform throughout.

## 3 Results and Discussions

### *3.1 Visual Testing (VT)*

VT is the oldest mode of NDT. This method involves surface examination by visual aids, mainly the eye. The human eye is the most important visual aid that is required for conducting VT inspection. Light is an important requirement for conducting VT. The amount of light entering the eye determines the quality of the image. Several factors such as surface condition, brightness, temperature, size and shape of the object can affect the amount of light that is being reflected into the eye. Human factors such as physiological and environmental factors also affect the results of VT

inspection. There are two types of VT methods, Direct VT (or DVT) and Remote VT (or RVT). Direct VT is carried out when the eye can be placed within 25 inches and at a minimum angle of 30 degrees to the surface of the specimen that is to be evaluated. Remote VT is done for sites which do not have proper accessibility and no illumination. In this mode, light is supplied by an external source via cable along with a camera which transmits images or a live feed on to a monitor .

*Advantages*- VT has major applications in building construction, aerospace, power plants etc. Some of the major advantages of VT are its simplicity, cost effectiveness and the ability to perform the test at any time. *Disadvantages*- However, its reliability is highly dependent upon a range of technological, physiological and environmental conditions as well as personnel knowledge. Disregarding any criteria of inspection can lead to inaccurate evaluations.

*3.2 Penetrant Testing (PT)*

PT is an application of capillary action. When a special liquid, called a penetrant, is applied on the surface of a component, usually by an aerosol, it fills any discontinuity that may be present by the virtue of capillary action. There are mainly two types of penetrants, fluorescent and visible dye penetrants. However due to better contrast ratio of fluorescent dyes, they are typically used for PT. The component to be inspected is immersed in the penetrant for some time and is then washed. It is then dried and observed under black light (fluorescent dyes). The set up for PT can be manual and or even automatic. However for simplicity in field inspection, dye penetrants (which do not require complicated setups) are used. PT done with dye requires an additional liquid called developer. Developer draws out the dye which may be entrapped in a discontinuity, making it visible. Factors such as wettability, capillary action, dry concentrate and viscosity determine the choice of penetrant to be used. Before any PT is done, there are certain prerequisites that should be noted: 1) temperature, if the surface is too cold a longer period of immersion is required and if it is too high the penetrant may evaporate, 2) environmental considerations which include flammability, fumes etc. 3) lighting and 4) surface condition such as coatings or rust should be addressed as it can affect the action of the penetrant.

*Advantages*- it is the most common mode of NDT because of its cost effectiveness, portability and versatility. *Disadvantages***-** however like VT, PT is limited to surface defects. There are also many prerequisites and factors which affect the effectiveness of the test. A major concern with PT is that it is highly dependent upon the action of the penetrant, cleaning can remove the penetrant from discontinuities thereby evading the purpose of the test. This is known as overwashing.

*3.3 Magnetic Particle Inspection (MPI)*

MPI is a test based upon the behaviour of ferromagnetic materials when exposed to a magnetic field. Due to their high magnetic permeability, the domains align themselves in the direction of the applied magnetic force creating continuous lines. The underlying principle of MPI is studying the continuity of these lines of force, if the material is uniform then the lines are continuous. Whereas the lines distort in the case of a discontinuity because they tend to follow a path where they experience the least reluctance (ability of a domain to resist alignment in the direction of applied magnetic force). In order to do so these lines bend around the discontinuity creating a “distorted” field. Sometimes if discontinuities are large enough, these lines tend to exit the surface and form through air over the disconituity. This is known as leakage of magnetic fields. This leakage can be made visible by applying a ferromagnetic powder. The powder gets accumulated and a visible indication is observed. Usually iron filings are used but for improved visibility ferric oxide powder is used. The detection is however affected by physical factors such as thickness (thicker the material stronger the field required), material (different components affect permeability of the material) and internal stresses (effect on magnetic conductivity) of the component. Some other testing factors which include distancing between the magnetic poles and speed of detector also affect the results of MPI. Test results of MPI are instantaneous and easy to interpret.

*Advantages*-Virtually components of any size and shape can be tested by this technique and can be cheaper to perform than other methods. Like VT, it can be performed at any stage of manufacturing. *Disadvantages***-** However, MPI is only restricted to testing of ferromagnetic materials and detection is limited to discontinuities on or near the surface. Sometimes discontinuities are hard to detect by this method due to the complex shape of a component. Sometimes due to formation of eddy currents, heating takes place which can cause damage to the component.

*3.4 Radiographic Testing (RT)*

RT is based upon the characteristics of radiation. Radiation is nothing but waves or particles emitted from a source passing through a medium. Radiography is the practice of capturing images by employing the characteristics of radiation. One of the key characteristics of radiation is that it can be absorbed by materials and the amount absorbed varies with thickness of the material. When radiation is made to pass through a component it is attenuated by the material. In regions of discontinuities, these rays experience lesser attenuation and travel faster in comparison to rays in uniform regions. The area of the film in the vicinity of the disconituity experiences more spot light and therefore a unique image is obtained making the discontinuity visible. There are mainly two types of radiations used in radiography, x-radiation and gamma radiation. While gamma rays are produced from a radioactive source and x-rays in a vacuum tube, characteristics of radiation apply to both.

*Advantages*- it is highly accurate and can be used on components of different shapes and sizes. Modern equipment has enabled the use of portable equipment which are just as effective as immobile ones. The major advantage lies in its ability to detect subsurface defects as opposed to the methods discussed previously. *Disadvantages*- subsurface defects can be observed only with proper orientation which can make the process time consuming and less reliable. Along with the health hazards, RT has material limitations. It can also be very expensive due to the high cost of machines and depletable sources of radiation. Gamma rays specifically are a major health hazard because of which x-rays are preferred over them.

*3.5 Ultrasonic Testing (UT)*

UT is based on the behaviour of sound waves. Sound is nothing but disturbance of particles in solid, liquid or gas. This disturbance is propagated at different frequencies in different media. In UT, a pulse generated by a piezoelectric material is propagated through a component. The frequency of this pulse is usually in the order of $10^6$Hz to facilitate longer transmission distances. As long as the pulse travels through uniform regions, its properties will not be altered. Changes in properties of the pulse (like velocity) are indicative of a present discontinuity. The changes are detected by a transducer which converts sound energy into visual signals that can be displayed on LCD or CRT. The transducer can also detect sounds of inaudible frequencies, making UT accurate and reliable. The accuracy of UT depends upon variables like temperature which can affect the attenuation of waves and even the dimensions of the component. Surface condition is an important variable as tighter bound coatings facilitate better pulse transmission. Orientation of the discontinuity is however the most important factor in detection by UT.

*Advantages*- UT can also detect surface and subsurface defects. However unlike RT, the technology requires minimum consumables making it cheaper to operate. The equipment is also portable and easy to use. Another major advantage lies within its ability to detect smaller discontinuities. There are also techniques available to measure these defects as precisely as possible. *Disadvantages*- discontinuities of different orientations and of very small size may go undetected. Despite the visual aids and available techniques, improper evaluations can be made due to which UT requires highly skilled and experienced personnel.

*3.6 Eddy Current Testing (ECT)*

ECT is based on the fluctuation of the magnetic field induced due to eddy currents. When a metal is placed in a fluctuating magnetic field produced by alternating current in a loop, current is generated in the metal which is swirly in nature. These currents are called eddy currents. A typical ECT set up consists of two circuits, a primary and secondary circuit. The primary circuit is connected to an AC supply which sets up a primary fluctuating magnetic field which induces eddy currents in the test specimen. These eddy currents set up a secondary magnetic field which interferes with the primary magnetic field thereby affecting the current flowing through the coil. Therefore if eddy currents are altered then the current set up due to the secondary magnetic field is also affected thus changing the primary current. This change will produce a change in the impedance reading indicating a discontinuity. There are two types of ECT, pulsed ECT and Remote ECT. Pulsed ECT is used to detect defects in conductive materials and the latter is used for ECT in remote regions.

*Advantages*- equipment required for ECT is very portable and results are quick making it suitable for mass manufacture. The testing procedure is very safe and non-destructive. *Disadvantages*- ECT is limited to conducting materials. Evaluation of subsurface defects is difficult. In addition to this, eddy current theory is complicated.

### 3.7 Acoustic Emissions (AE)

AE is based on the principle that when a material undergoes deformation, energy is released in the form of elastic waves of high frequency. These waves travel outwards to the surface. Deformations continue to release energy when they are subjected to further loading. These waves are called acoustic emissions. These acoustic emissions are then detected by sensors and the data is transmitted to a transducer which amplifies and converts the signals into interpretable data. One major difference between AE and other modes of NDT as discussed above is that in AE, disturbance is originating from the material and not from an external source. Normally, in other modes a disturbance is generated and transmitted through the specimen and the response is studied. AE can provide information about the origin of the discontinuity and propagation of the same. Attenuation and wave velocity are two important factors which affect the wave propagation process which ultimately influences detection.

*Advantages*- A major advantage of AE lies in the fact that it can be implemented at any stage, even during operation. It provides real time evaluation and has good economic viability. *Disadvantages*- different stresses produce different emissions and therefore different models of analysis are required. Another disadvantage is that observations can be influenced by unnecessary disturbances which can lead to inaccurate evaluations. Sometimes evaluation of a single defect can be disturbed by emissions of other discontinuities.

## 4. Advances In Applications of NDT

### *4.1 Diagnostics and Protection of Cultural Heritage*

Requirement for the preservation of historic buildings involves a great knowledge of materials and structures, possible state of damage and its causes. In order to overcome the gap between the complexity of these structures and the initial low knowledge, experimental investigations which are on-site are recommended and required by Codes of Standards in many countries. The most arduous problems of unseen situations or detection of the mechanical behaviour of various structural elements like iron tie rods, timber roofs and floors, load bearing walls etc., can be solved using the Non- Destructive Testing techniques. In the instance of masonry the most often utilised NDTs are: thermovision, radar and sonic tests and as Minor Destructive techniques, the flat jack tests. Binda, L., Saisi, A., Tiraboschi, C. (2000) and Suprenant, B. & Schuller, M. (1994) outline the main investigation techniques and their physical principles.

A significant information about the glass manufacturing technique employed can be extracted by using the Non-Destructive Technique of Environmental Scanning Electron Microscopy and Fiber Optics Microscopy with an Energy Dispersive X-Ray analysis on glass fragments obtained from archaeological excavations E. Cheilakou, N. Liarokapi & M. Koui (2011). A new Non-Destructive Technique which is Adaptable Neural Network Architecture was introduced by A.D. Doulamis & T.A. Varvarigou (2011) in 2012 is capable of recognising the artistic styles incorporated in paintings. This technique can be very useful in preserving and restoring cultural paintings.

### *4.2 On-site Measurement and Monitoring of Concrete by NDT*

Concrete is extensively used in various structures and defects in concrete structures can lead to severe damages and loss of life. These concrete structures often deteriorate unnoticed and in many cases the initial damage is not visible until there is a huge damage which is beyond repair. These problems occur due to the absence of quality check, no detection of fault during the initial construction stages, shrinkage cracking in the structures during the initial stages etc., all these problems and thus the structural failure can be prevented by using Non-Destructive Tests to check for defects during the initial stages only.

To improve the brittle behaviour, steel fibers are introduced into the concrete after which chemical coating of the fibers is done in order to attain maximum mechanical properties mainly toughness. For monitoring of structures, Acoustic Emission(AE) provides a way to identify between various fracture stages which are pre-peak and post-peak. The Acoustic Emission test also reveals the efficiency of chemical coating D.G. Aggelis, D.V. Soulioti, N.M. Barkoula, A.S. Paipetis & T.E. Matikas (2011). NDT can act as a powerful tool providing valuable information about the current condition of a structure. Using NDT during the initial stages of building construction can help in identifying errors beforehand leading to easy repair as the error will be limited to a specific area, as proved by J.S.

Clausen, N. Zoidis & A. Knudsen (2011) This helps in averting grave situations like loss of life and prevent them from happening in future. From a research done by E. Tatsis, N. Zoidis & C. Vlachopoulos, T.E. Matikas and A. Gotzamanis, Ultrasonic method of NDT can also be used to estimate the condition of the structural elements and the extent of damages caused by a fire in a building E. Tatsis, N. Zoidis, C. Vlachopoulos, T.E. Matikas & A. Gotzamanis (2011).

Shrinkage cracking is one of the major problems during the early stages in the concrete industry. It occurs primarily due to the usage of low quality aggregate which results in volumetric contraction of coarse aggregate. The detection of shrinkage cracking is very important and can be done by using the Acoustic Emission method of NDT which shows some promising results in identification of shrinkage cracking according to M. Ohtsu, T. Matsuo, S. Nozaki & Y. Kawasaki (2011).

*4.3 Application of Ultrasonics*

As discussed above waves can be used to detect defects in a specimen be it a sound wave, a surface wave etc. They are made to propagate through the test specimen and the transmitted signals are monitored or interpreted. Surface waves as described by Rayleigh are the waves which move along the solid surface and their motion is restricted to a region close to the surface whose depth is akin to the wavelength. The surface waves have distinctive characteristics which helps in characterization and detection of surface-breaking defects in materials. Surface waves are sensitive to small defects because the energy of the surface waves is concentrated on the small layer in the neighborhood of the surface. Depending upon the defect size relative to the wavelength of the surface wave, they are diffracted, scattered and reflected by the surface defects. The surface waves have been used to characterize cracks in aluminum (Al) bars (Edwards et al. 2004, Palmer & Dixon 2003, Edwards et al. 2006).

After a specific period of time fouling deposits start developing on the industrial pipes. These deposits are hard to prevent and pose a serious problem to the plant operators. There are a lot of varieties of fouling that may occur, entailing different chemical structures and elemental composition. The presence of fouling in the pipe results in corrosion, blockage, rupture and deformation which leads to substantial increase in operation cost and contamination of products. Pipe inspection can be done by involving ultrasonic guided waves Withers, 1994; Mc Cements, 1995; Hay & Rose, 2003; Lohr & Rose, 2003; da Silva et al, 2005; da Silva et al, 2007b). This technique which is capable of long range identification with great reliability, has been shown to identify not only fouling but also particular defects in pipes. As experimented by Ignacio Garcia de Carellan[35], due to the potential of conventional ultrasonic cleaning baths to eliminate hard deposits there is a possibility to break the bond between the fouling deposits and the steel pipe by using guided waves of sufficient magnitude.

*4.4 Application of Acoustic Emission (AE)*

As discussed above, every material produces high frequency sound known as the acoustic emission when they are stressed, deformed or are going through structural change such formation of crack. These emissions are inaudible to the human ear but can be detected by special devices which helps in knowing about the cracks that may lead to some serious damage in the future. Acoustic Emission monitoring is very useful in evaluating the quality of laser cutting. Laser cutting is a process of cutting a material using a laser. Quality of the laser cut is defined mainly by the state of the laser cut surface and the existence of the re-solidified material occurring as dross at the lower cut edge. Changes in the quality of beam, surface of materials and gas flow can lead to reduction in the quality of laser cut. While the laser cuts the material AE signals are emitted due to the interaction among the cutting gas, laser light and plate material. These AE signals can be captured during or immediately after the cutting process and the dross formation and the degradation of the laser cut surface can be detected thus helping in controlling the cutting process and ensuring high laser cut quality. T. Kek & J. Grum (2011) proved by experimenting that this method of capturing AE signals to control the laser cutting process is applicable to unalloyed as well as alloyed steel(austenitic stainless steel) with different thickness.

Surface rail transportation is one of the most significant and important discoveries of modern technology. With time trains have been improving even if we exclude magnetic levitation technology, these improvements have increased the cruising quality as well as the speed. Still, due to the dynamic nature of loading and frequent working of the trains, deformations develop in the wheels and the axles. One of the major factors that leads to the deterioration of the wheel is the emergency braking. Emergency braking produces a lot of heat due to the friction which leads to small or large surfaces of the tyre to be flattened. These flattened surfaces can lead to misalignment,

bad rolling quality, fatigue, lower speed etc. This defect can be detected by using AE monitoring as the wheels produce transient elastic waves to which AE monitoring is sensitive thus detecting the defects when the train passes. Wind turbine blades are one of the most vulnerable and exposed parts of the turbine. They are continuously in contact with the gushes of wind and sometimes exposed to severe weather conditions. This leads to deterioration of the blades. If the blades are deteriorated the turbine will not work thus, there is a need for monitoring these blades which can be done by using AE monitoring. N. Tsopelas, D. Kourousis, I. Ladis, A. Anastasopoulos, D.J. Lekou & F. Mouzakis (2011) summarizes the first successful long-term AE monitoring of the full-scale in service Wind Turbine.

*4.5 Other Applications*

The ultrasonic method of NDT which is discussed above can be used in Quality Control of the Ground Improvement for setting up Crude oil Storage tanks. This method involves production of vertical impulses by the surface waves on the ground surface. The dispersive characteristic of the incoming wave after being reflected by the ground is calculated which gives the layer stiffness of the layered ground. This procedure permits measuring of shear modulus for profiles of the ground based on depth, with very small deformations before and after the improvement of the ground. The ratio of shear modulus before and after improvement represents an extent of ground improvement, which is then used as an estimate of successfulness for the conducted ground works. Using the ultrasonic method of NDT ensures the safety and stability of the oil storage tank M.S. Kovačević, A. Mirčeta & D. Marčić (2011).

The Radiographic Inspection method of NDT that is discussed above can be used to scan the porous construction materials for the moisture content. If there is excessive moisture content in the construction material, it affects the physical properties of building structures. This leads to dampness and thus weakening the structure. The Radiographic Inspection method of NDT provides continuous and fast measurement and increased metering on a single sample without influence of human factor J. Skramlik, M. Novotny & K. Suhajda (2011).

Aircrafts construction involves usage of carbon fiber laminates which are usually long fiber-enforced composite materials. Quality assurance activities are considered very important because separated fibers are regarded as weak areas. Thus, Structural Health Monitoring is becoming increasingly important nowadays. This structural health monitoring is done using Guided Waves i.e. Ultrasonic method of NDT and also by Acoustic Emission Monitoring technique of NDT.[40]

Ground transportation vehicles can also be monitored using Structural Health Monitoring. The wheel set of these vehicles produce sound signals which can be detected by using the Acoustic Emission Monitoring technique of NDT. If there is any damage in the wheel tread, such as pitting (production of cavities or holes), the sound signal produced changes and these changes in the signal can be detected by integrating AE monitoring system in the wheel set hollow shaft N. Meyendorf, B. Frankenstein & L. Schubert (2011).

## 5 Conclusion

Principle of Non-Destructive testing is simple, to make sure a product conforms to its requirements physically and operationally by conducting a series of tests which do not tamper with the overall viability of the final product. One of the bigger disadvantages of NDT is that a single technique is not sufficient for total satisfaction. However, combining two or more techniques can ensure a greater sense of satisfaction. The underlying working principles of various NDT techniques have proved to be effective in every situation making NDT a relatively mature field. Scope of development in NDT lies in its application and technology. Development in material technology has broadened the field of application of NDT along with the development of electronics over the last few decades which has provided the field with enhanced and sophisticated equipment (with underlying working principles remaining unchanged) making NDT more reliable and cost effective. This paper described the various methods involved in NDT and identifies as well as compiles the advances in the applications of NDT, showing that the field is not only limited to the identification of an issue, but also its classification and rectification followed by preservation.

*Conflict of interest statement*
The authors declared that they have no competing interests.

*Statement of authorship*
The authors have a responsibility for the conception and design of the study. The authors have approved the final article.

*Acknowledgments*
We are grateful to two anonymous reviewers for their valuable comments on the earlier version of this paper.

## References

A.D. Doulamis & T.A. Varvarigou (2011); Image analysis for artistic style identification: A powerful tool

Amenabar, I., Mendikute, A., López-Arraiza, A., Lizaranzu, M., & Aurrekoetxea, J. (2011). Comparison and analysis of non-destructive testing techniques suitable for delamination inspection in wind turbine blades. *Composites Part B: Engineering*, *42*(5), 1298-1305. https://doi.org/10.1016/j.compositesb.2011.01.025

B. L. Luk and Alan, H. S. Chan (2007); Human Factors and Ergonomics in Dye Penetrant and Magnetic Particles Nondestructive Inspection Methods.

Binda, L., Saisi, A., Tiraboschi, C. (2000); Investigation procedures for the diagnosis of historic masonries, Construction and Building Materials, Elsevier, 14(4), 199–233.

Bruce W. Drinkwater, Paul D. Wilcox (2006); Ultrasonic arrays for non-destructive evaluation: A review, NDT & E International, 39(7), 525–541.

Construction and Building Materials, ISSN. 2003:0950-618

D. Bracun, I. Polajnar, J. Diaci (2006); Indentation shape parameters as indicators of spot weld quality; International Journal of Materials and Product Technology, 27(3/4), 247 – 257.

D.G. Aggelis, D.V. Soulioti, N.M. Barkoula, A.S. Paipetis & T.E. Matikas (2011); Influence of the fiber chemical coating on the fracture behavior of steel fiber concrete measured by acoustic emission, Proceedings of the fifth conference on Emerging Technologies in NDT, Ioannina, Greece, 111-116.

da Silva, J., Jaidilson J, Lima, Marcus N.A., da Rocha Neto & Jose Sergio (2007b); Fouling detection using ultrasonic guided waves and parameter estimation, Third IFAC Symposium on System Structure and Control.

da Silva, J., Wanzeller, M.G., de Almeida Farias, P., da Rocha Neto & Jose Sergio. (2005); Development of Circuits for Excitation and Reception in Ultrasonic Transducers for Generation of Guided Waves in Hollow Cylinders for Fouling Detection, Instrumentation and Measurement Technology Conference, 467–471.

E. Cheilakou, N. Liarokapi & M. Koui (2011); NDT characterization of ancient glass objects from the Aegean with an approach of the manufacturing technique, Proceedings of the fifth conference on Emerging Technologies in NDT, Ioannina, Greece, 63-70.

E. Tatsis, N. Zoidis, C. Vlachopoulos, T.E. Matikas & A. Gotzamanis (2011); Visual inspection and evaluation using NDT testing methods of industrial building after fire, Proceedings of the fifth conference on Emerging Technologies in NDT, Ioannina, Greece, 131-136.

El Masri, Y., & Rakha, T. (2020). A scoping review of non-destructive testing (NDT) techniques in building performance diagnostic inspections. *Construction and Building Materials*, *265*, 120542. https://doi.org/10.1016/j.conbuildmat.2020.120542

for preserving cultural heritage, Proceedings of the fifth conference on Emerging Technologies in NDT, Ioannina, Greece, 71-76.

Francesca Cau, Alessandra Fanni Augusto Montisci, Pietro Testoni, Mariangela Usai (2006); A signal processing tool for non-destructive testing of inaccessible pipes; Engineering Applications of Artificial Intelligence, 19(7), 753–760.

G. S. Park, P. W. Jang, Y. W. Rho (2001); Optimum Design of a Non Destructive Testing System to Maximize Magnetic Flux Leakage; Journal of Magnetics, 6(1), 31-35.

Gabriel Dan Taşcă, Gheorghe Amza; Research regarding ultrasonic examination of complex parts.

Gerd Dobmann, Iris Altpeter, Bernd Wolter (2008); Materials characterization a challenge in NDT for quality management; International Non-Destructive Testing Symposium and Exhibition, Istanbul Turkey.

Gholizadeh, S. (2016). A review of non-destructive testing methods of composite materials. *Procedia Structural Integrity*, *1*, 50-57. https://doi.org/10.1016/j.prostr.2016.02.008

Giovanni M. Carlomagnoet, Carosena Meola (2002); Comparison between thermographic techniques for frescoes NDT, NDT & E International, 35(8), 559–565.

H. A. Crostack, W. Reimers ( 2001); Evaluation of component integrity by non-destructive testing; International Journal of Materials and Product Technology, 3(2), 147 - 162.

Hay, T.R & Rose, J.L. (2003); Fouling detection in the food industry using ultrasonic guided waves, Food-Control 14(7), 481–488.

Ignacio Garcia de Carellan, Phil Catton, Cem Selcuk & Tat-Hean Gan (2011); Methods for detection and cleaning of fouling in pipelines, Proceedings of the fifth conference on Emerging Technologies in NDT, Ioannina, Greece, 231-236.

J. Skramlik, M. Novotny & K. Suhajda (2011); Apparatus for measuring of liquid conductivity, Proceedings of the fifth conference on Emerging Technologies in NDT, Ioannina, Greece, 489-494.

J.S. Clausen, N. Zoidis & A. Knudsen (2011); Onsite measurements of concrete structures using Impact-Echo and Impulse Response, Proceedings of the fifth conference on Emerging Technologies in NDT, Ioannina, Greece, 117-122.

Jacek Jarmulak, Eugene J.H. Kerckhoffs, Peter Paul van't Veen (2001); Case-based reasoning for interpretation of data from non destructive testing; Engineering Applications of Artificial Intelligence; 14(4), 401–417.

Javier García-Martín, Jaime Gómez-Gil and Ernesto Vázquez-Sánchez (2011); Non-Destructive Techniques Based on Eddy Current Testing, Sensors, 11(3), 2525-2565.

Jolly, M. R., Prabhakar, A., Sturzu, B., Hollstein, K., Singh, R., Thomas, S., ... & Shaw, A. (2015). Review of non-destructive testing (NDT) techniques and their applicability to thick walled composites. *Procedia CIRP*, *38*, 129-136. https://doi.org/10.1016/j.procir.2015.07.043

K. Bollas, D. Papasalouros, D. Kourousis & A. Anastasopoulos (2011); Acoustic emission monitoring of wheel sets on moving trains, Proceedings of the fifth conference on Emerging Technologies in NDT, Ioannina, Greece, 341-346.

Li Zhijun (2001); Non-Destructive Testing of Advanced Composites; Aerospace Materials & Technology.

Lohr, K.R. & Rose, J.L. (2003); Ultrasonic guided wave and acoustic impact methods for pi1pe fouling detection, Journal-of-Food-Engineering 56(4), 315–324.

M. Ohtsu, T. Matsuo, S. Nozaki & Y. Kawasaki (2011); AE monitoring of shrinkage process in concrete, Proceedings of the fifth conference on Emerging Technologies in NDT, Ioannina, Greece, 143-148.

M.D. Beard, M.J.S. Lowe (2003); Non-destructive testing of rock bolts using guided ultrasonic waves, International Journal of Rock Mechanics and Mining Sciences, 40(4), 527–536.

M.S. Kovačević, A. Mirčeta & D. Marčić (2011); NDT and safety of crude oil storage tanks, Proceedings of the fifth conference on Emerging Technologies in NDT, Ioannina, Greece, 477-482.

Mc Clements, D.J. (1995); Advances in the application of ultrasound in food analysis and processing, Trends in Food Science and Technology, 6, 293–299.

N. Meyendorf, B. Frankenstein & L. Schubert (2011); Structural health monitoring for aircraft, ground transportation vehicles, wind turbines and pipes—prognosis, Proceedings of the fifth conference on Emerging Technologies in NDT, Ioannina, Greece, 15-22.

N. Tsopelas, D. Kourousis, I. Ladis, A. Anastasopoulos, D.J. Lekou & F. Mouzakis (2011); Health monitoring of operating wind turbine blades with acoustic emission, Proceedings of the fifth conference on Emerging Technologies in NDT, Ioannina, Greece, 347-352.

P Cawley (2001); Non-destructive testing—current capabilities and future directions; Journal of Materials: Design and Applications; 215, 213-223.

P. Tschelisnig (2001); Acoustic emission testing (AET) an integral non-destructive testing method; International Journal of Materials and Product Technology, 3(3/4), 267 – 275.

Rainer Pohl , A Erhard, H.J Montag , H.-M Thomas, H Wüstenberg (2004); NDT techniques for railroad wheel and gauge corner inspection; NDT & E International, 37(2), 89–94.

Rodrigo Benenson, Stephane Petti, Thierry Fraichard, Michel Parent (2008); Towards urban driverless vehicles; International Journal of Vehicle Autonomous Systems, 6(1/2), 4 – 23.

Suprenant, B. & Schuller, M. (1994); Non destructive evaluation & testing of masonry structures, The Aberdeen Group, USA.

T. Kek & J. Grum (2011); AE monitoring at laser cutting, Proceedings of the fifth conference on Emerging Technologies in NDT, Ioannina, Greece, 321-326.

Tsioulou, O., Lampropoulos, A., & Paschalis, S. (2017). Combined non-destructive testing (NDT) method for the evaluation of the mechanical characteristics of ultra high performance fibre reinforced concrete (UHPFRC). *Construction and Building Materials*, *131*, 66-77. https://doi.org/10.1016/j.conbuildmat.2016.11.068

Umesh Singh, Mohan Singh and M. K. Singh (2012); Analysis study on surface and sub surface imperfections through magnetic particle crack detection for nonlinear dynamic models of some mining components; Journal of Mechanical Engineering Research, 4(5), 185-191.

Withers, P.M. (1994); Ultrasonics for the detection of fouling. Food Technology International Europe 1994, 96–99 (Reprint 768).

**How to Cite**:
Srivastava, K. N., & Raj, R. (2022). Aerodynamic study of tall buildings under wind load. *International Journal of Health Sciences*, *6*(S3), 2620–2641.
https://doi.org/10.53730/ijhs.v6nS3.6142

# Aerodynamic study of tall buildings under wind load

***Kunwar Navodit Srivastava***
Department of Civil Engineering, Delhi Technological University, Delhi, 110042, India

***Ritu Raj***
Department of Civil Engineering, Delhi Technological University, Delhi, 110042, India

***Abstract*---**In Design of high-rise structures, Wind is considered as one of the important horizontal forces that have significant impact on the response of building. Due to rise in population, the demand for tall buildings is increasing day by day. Wind Load on such structures are calculated using pressure coefficients and force coefficients which are available in various international codes and standards. However, these international codes and standards give information about regular shape buildings such as square, rectangle, circular or octagonal. With the technological advancement, composite plan shape buildings such as square and circular, circular and hexagonal and square and octagonal etc. have been considered by many architects keeping in view of aesthetics of Building. The wind flow around a tall building with composite plan shapes having different height ratio variation differs from what we get in regular shape analysis. Since, data is not available regarding such buildings, the need to carry wind tunnel testing or CFD become important for analyzing wind effect on such buildings. CFD (Computation Fluid Dynamics) for determining wind responses is becoming immensely popular. It has expanded as a tool to replace wind tunnel testing as it is quicker, less expansive and give more information and control to designers.In the present research study, two building plan shape triangle and circle has been considered with different height ratio. The building selected to carry out the numerical analysis are uniform and composite plan tall buildings. The length of equilateral triangle is taken as 40m and height being 200m. For composite plan shape building only height ratio is changed keeping the same dimensions. The prototype building is considered to be located in terrain category 2 as per IS-875 Part 3. Assumption is made that the variation for wind speed with height follows the power law with power law coefficient taken as .177. CFD analysis is carried

International Journal of Health Sciences ISSN 2550-6978 E-ISSN 2550-696X © 2022.
***Corresponding author**: Raj, R.;*
*Manuscript submitted: 18 Dec 2021, Manuscript revised: 27 March 2022, Accepted for publication: 09 April 2022*

out in ANSYS CFX taking 1:200 scaled down model of tall buildings. In total, 5 models have been considered and pressure is evaluated taking various points on the building. Each model is analyzed for different angle of wind incidence namely 0, 60, and 120 degrees. The result obtained for the uniform shape cross section building has been compared with the wind tunnel testing results available in various literatures.

***Keywords*---**tall building, CFD, wind effect, wind angle, force coefficient, pressure coefficient.

## Introduction

Tall Buildings have always fascinated humans having unique appeal and pride associated with them. Due to increasing population and scarcity of land the demand for Tall Buildings have increased day by day. With the development of new better construction techniques, better materials and structural systems, these buildings have proven to be the safe and economical structural solution to problem of spacious designs. In design of tall structures, wind loading is one of the prime important lateral loads that needs to be considered while designing. In general, occupant comfort needs to be considered along with structural safety. Wind is a time varying force having two components, a mean and a fluctuating component. Wind is a complicated phenomenon having eddies of varied size add rotating properties. Due to these eddies wind is turbulent and gusty in nature. Wind Loading initially was viewed as estimating the dynamic pressure of wind at the structure and then just multiply it by some Shape factor and area of a structure to obtain the wind force.

However later it was realized that wind dynamic in nature and dynamic responses such as galloping, flutter, vortex excitement, ovalling etc. need to be examined too. The shape of buildings is a well-known subject in aerodynamics optimization that has a significant impact on the behavior of high structures under wind loads. Wind response can be reduced by optimizing the geometry of supertall structures for aerodynamics during the design stage.Wind is a phenomenon in which the motion of individual particles is so unpredictable that statistical distributions of velocity rather than simple averages must be considered. Although each has its own local impact, the total wind force is equal to the sum of windward pressure and leeward suction. In order to get complete wind analysis, one need to determine the wind climate, influence of terrain and topography, aerodynamic shape of a structure and dynamic effects. Various international codes and standards such as ASCE 7-10: Minimum design loads for buildings and other structures, IS-875 Part 3 : Code of practice for design loads (other than earthquake) for buildings and structures), BS 6399-2:1997 loading for buildings code of practice for Wind loads British standards, EN 1991-1-4: Eurocode : Actions on Structure Part 1-4: General Actions-Wind Actions,2010 estimates the pressure coefficients and force coefficients which are used for computing wind loads on buildings subjected to windloads.

Very Limited Information for uniform plan shaped buildings with different aspect ratio are available in these standards. The Indian code suggests pressure coefficients for different plan shape like square, rectangle etc. at 0-degree and 90-degree angle of wind incidence. The wind pressure of tall building not only influenced by building geometry and wind incidence but also depend upon height ratio between the plan shapes However, these codes do not provide information regarding wind load acting on composite plan shaped buildings. Also, available information does not include wind pressure coefficient (Cp) or wind force coefficient (Cf) for the buildings where cross-sectional shape change with height. For such buildings wind tunnel testing have been proven to be the efficient and practical approach to study the response of buildings and structures under wind load. Recent Studies have also shown the use of CFD to carry out investigation for wind load pressure distribution and computation of pressure and force coefficients.

**CFD (Computational fluid dynamics)**

Computational fluid dynamics involves numerical approach and algorithms to solve and analyze problems that involves fluid flow. CFD analysis is widely employed in aerodynamics and hydrodynamics where pressure and velocities are the parameters. CFD analysis can save alot of time and are cheaper as compared to conventional testing. All of the relevant parameters may be analyzed and monitored at the same time in CFD, with great time and spatial resolution. Because CFD analysis approximates a genuine physical solution, it cannot totally replace actual testing. The steps of CFD analysis include the following:

- Pre-processing
- Solving
- Post processing

Compared to wind tunnel testing, CFD has the following advantages:

- Comprehensive domain analysis
- Simple alternative analysis
- Improved visualisation of outcomes
- Cost-effective

**CFD Validation**

The validity of the ANSYS CFX software is validated before beginning the numerical studyof the building. For this purpose, a square plan shaped building with dimensions of 150 mm ×150 mm and a height of 500 mm (i.e., aspect ratio 1:5) is investigated using the k-modelwith ANSYS CFX in the domain under uniform wind flow.

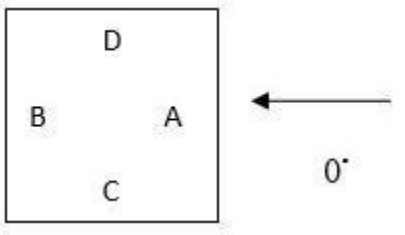


Figure 1. Different faces of the model with direction of wind

At the inlet, a uniform wind flow of 10 m/s is given. As previously stated, the domain is built according to Revuz et al (2010). The ANSYS CFX programme determines the face average values of coefficient of pressure, which are then compared to wind action codes from various regions.

Table 1
Comparison of face average values of coefficients of pressure

| Wind loading code | Face- A | Face-B | Face-C | Face-D |
|---|---|---|---|---|
| By ANSYS CFX | 0.9 | -0.46 | -0.67 | -0.68 |
| ASCE 7-10 | 0.8 | -0.5 | -0.7 | -0.7 |
| AS/NZS-1170.2(2002) | 0.8 | -0.5 | -0.65 | -0.65 |
| IS: 875 (part3) (2015) | 0.8 | -0.25 | -0.8 | -0.8 |

**Model scale**

In this study, two different composite building plan shapes, triangle and circle are consideredwith different height ratios. The sides of triangle and total height of buildings are 40m and 100m respectively. The following dimensions are considered for all models with only height ratio changed. The prototype building has been considered to be situated at in Terrain Category- The free mean wind velocity is taken as 10m/s and each model is studied at various wind incidence angle of 0°, 60°, 120°, 180°. The wind profile boundary layer is governed by power law equation with power law index coefficient as 0.147.

| Parameter | Prototype Dimension(m) | ANSYS Model Dimension(mm) | of Model1:200 |
|---|---|---|---|
| Length | 40 | 200 | |
| Height | 100 | 500 | |

The following geometrical parameters are considered in this study,

- Wind Incidence Angle (0°, 60°,120°, 180°)
- Cross section combination of Triangle and Circle along height.
- Height of Building (50% Triangle & 50% Circle)

This study also has been validated with wind pressure of an isolated square plan shape, tall building using CFD simulation and compared it with wind effects of conventional plan shape building given in Indian code IS: 875 (Part- 3), 2015.

**Methodology**

**Design wind pressure**

As per IS-875 Part 3,
[Design Velocity] = Vb*K1*K2*K3*K4 Where, Vb=Basic Wind velocity
K1= Probability Factor/Risk Coefficient K2=Terrain and Height Factor
K3=Topography Factor

K4=Importance factor of the cyclonic region Design wind pressure is given as:

$$Pd=0.6\,[Vz]^2$$

The mean pressure coefficient 'Cp mean' is calculated from the equation given below:

$$C_{p\,mean} = \frac{(p-p_o)}{\frac{1}{2}\rho U_H^2}$$

where p is the pressure at point on surface,
$p_o$ is the reference height static pressure,
ρ is the air density
$U_H$ is mean wind velocity at the building reference height.

**Models**

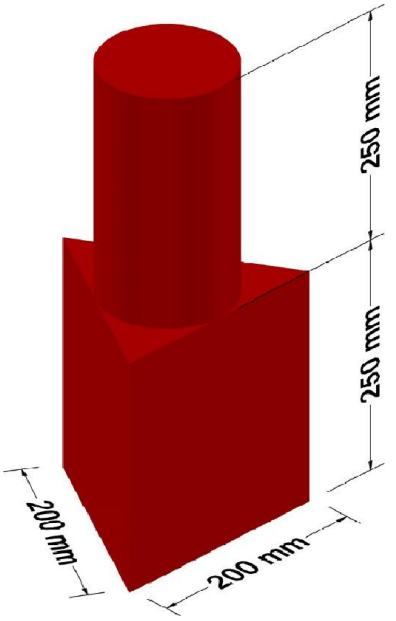



Figure 2. Model 2 50% Triangle 50% Circle

**Domain**

In case of high-rise buildings, domain size is mainly governed by height of the building such that a large number of cell count could be formed and out of them, many being used up in theregion far away from wake region. Domain size selected in modeling is defined as per frank et al (2004), The inlet and outlet distance of the domain from the building position is taken as 5H and 15H, respectively. The

side aspect and top clearance are also taken as 5H, where H is the height of the building. The domain configurations are depicted in Fig.

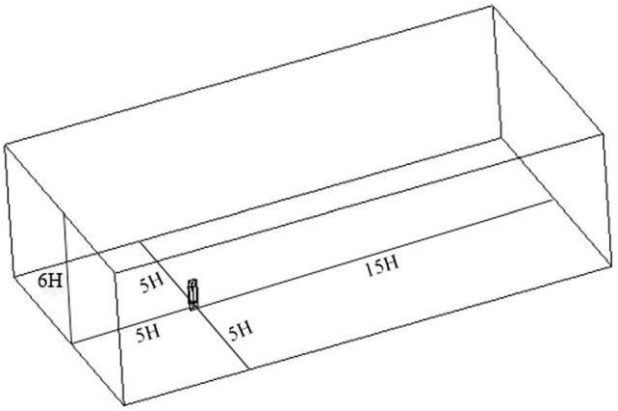



Fig 2. Domain used for the study

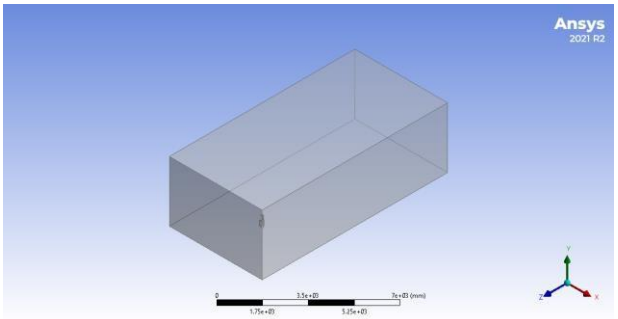



Fig 3. Domain

**Meshing**

Meshing is an integral part of engineering simulation process where complex geometries aredivided to simple elements that can be used as discrete local approximations of larger domain. Meshing influences the accuracy, convergence and speed of the simulation. Finer the mesh,better the accuracy. Types of Mesh:

- Tetrahedron Meshing
- Pyramid Meshing
- Hexahedron Meshing
- Polyhedron Meshing
- Prism Meshing

The meshing in domain is done by tetrahedral mesh elements. Meshing near the buildings aremade comparatively finer for enhancing the accuracy of results. The velocity at the inlet is taken as 10 m/s. No slip condition is defined for side walls and ground.

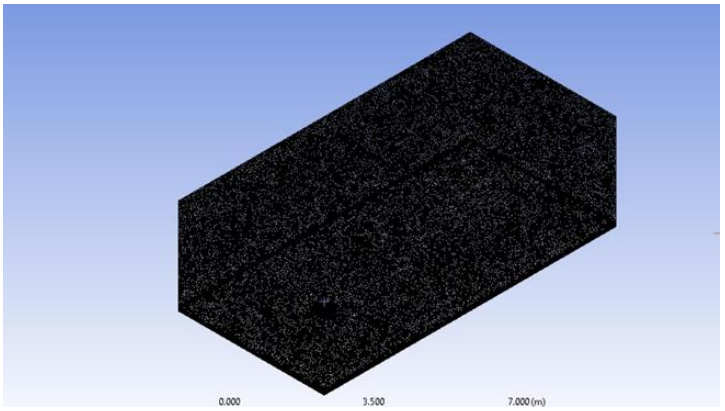



Fig 4. Mesh pattern

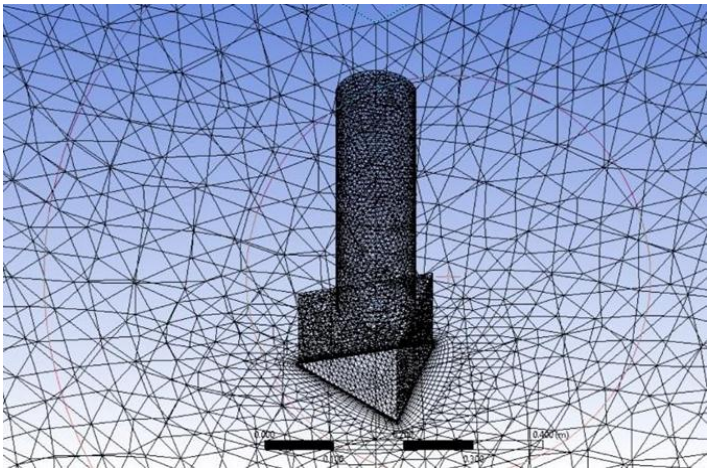


Fig 5. Mesh pattern around the building

**Mean wind profile with height**

Due to the roughness of earth surface, there acts a drag force on wind flow near the ground. This effect gradually decreases as the height increases and at a certain gradient level (around 400m), this drag-force becomes negligible. The degree of surface roughness and drag caused by surrounding projections that oppose wind flow determine the vertical profile of wind speed. Gradient height is the height at which the drag effects disappear, while gradient velocity is the corresponding velocity. The height up-to which wind speed is influenced by topography is called atmospheric boundary layer.

**Power law**

As per Power Law, the wind speed profile within the atmospheric boundary layer is given by:

$$\frac{}{} = [(\quad)^{\wedge}(\quad \frac{1}{\quad})]$$

$Vo$
$Zo$
$alpha$

Where, V = velocity of wind at height Z

Vo = gradient velocity of wind at reference height ZoZ = height above ground
Zo = Nominal height of Boundary layer (also called gradient height)alpha = power law coefficient.

**Logarithmic law**

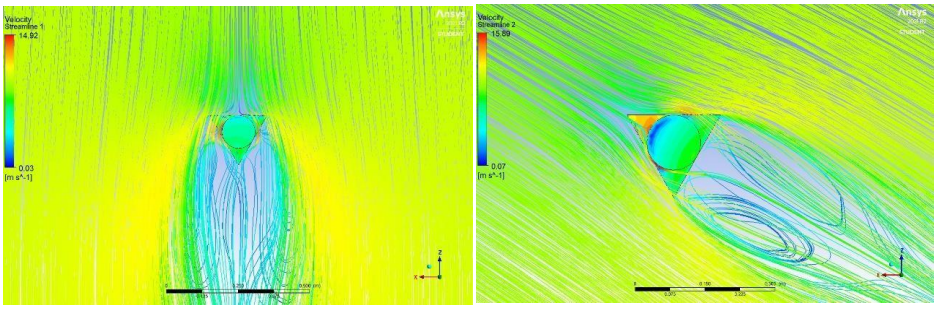
1

$$u = \frac{u_*}{k} \ln \frac{Z}{Z_0}$$

where u is the wind speed at height Z above ground, k is the von Karman constant equal to 0.4 (approximately)and Zo is the ground roughness.
$u_*$ is shear velocity which is defined as:

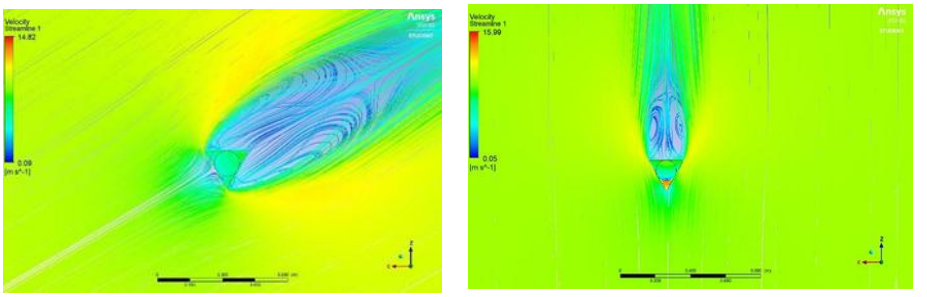
$$u_* = \sqrt{\frac{r_0}{\rho}}$$

where $r_0$ is the stress of wind at ground level and $\rho$ is the air density.



0° Wind Angle 60° Wind Angle

Fig 6. Wind flow pattern around the building for various wind angles



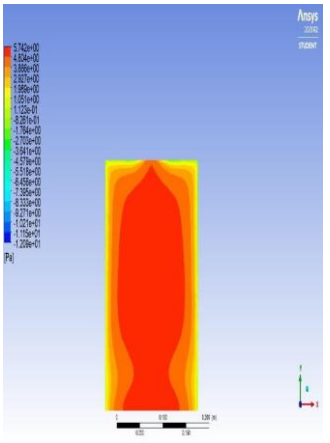
120° Wind Angle 180°Wind Angle

**Pressure distribution**

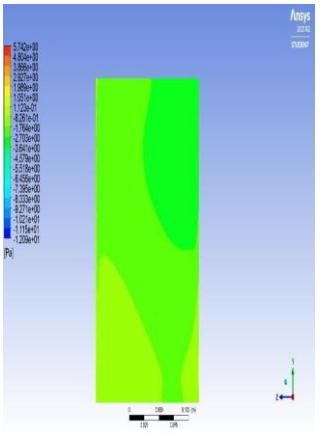
The pressure variation at various faces of Buildings are as shown using contour plots.

**Model- 1 (250 triangle 250 circle)**

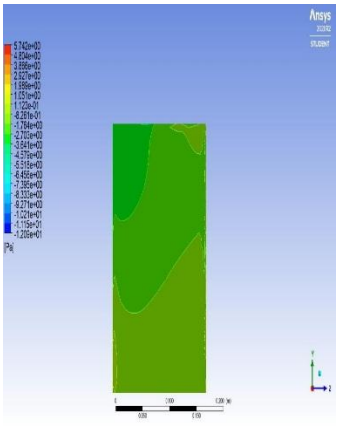
- Figures 5-1 to 5-28 exhibit the pressure distribution on different faces at varying anglesof incidence.
- Initially, at 0 degree of incidence face A shows a positive pressure being the windward face of the triangular building whereas face B and C depicts a negative pressure distribution being the leeward and side wall face of the building.
- Similarly, at 0 degree of incidence face A shows a positive pressure being the windward face of the circular building whereas face B, C And D depicts a negative pressure distribution being the leeward and side wall face of the building.
- As the angle of incidence changes to the pressure distribution changes and so the pressure coefficient.
- For 60° and 120° angle of incidence, positive pressure distribution at face A become slightly less compared to what in case of 0° along with suction pressure increase at face C in triangular building and face D in circular and a comparable change can be viewed from the pressure coefficient data so obtained as a result.
- For 180° angle of incidence, face C in triangular building and face D in circular building become windward and similar contour plot pattern as of face A when angle of attack was 0° can be seen.



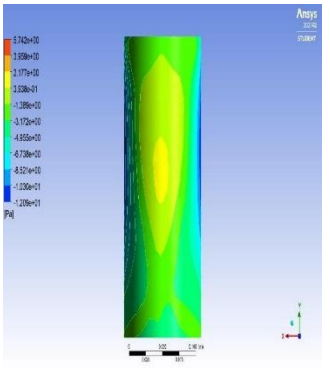
Triangle Face A



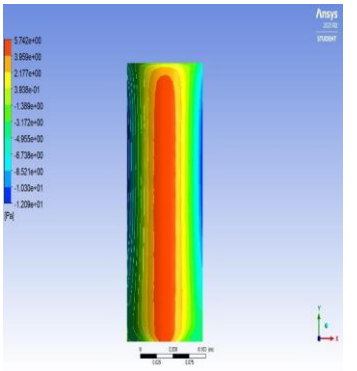
Triangle Face B



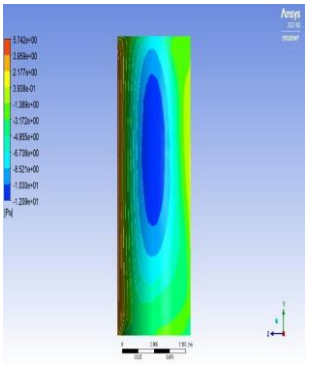
Triangle Face C



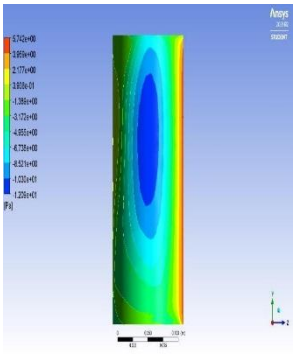
Circle Face A



Circle Face B

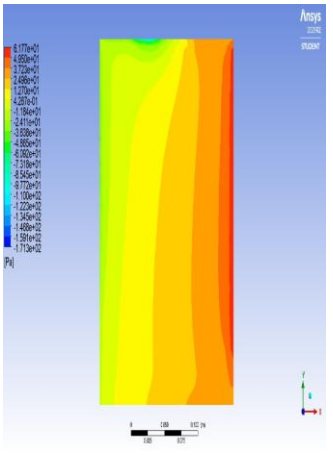


Circle Face C

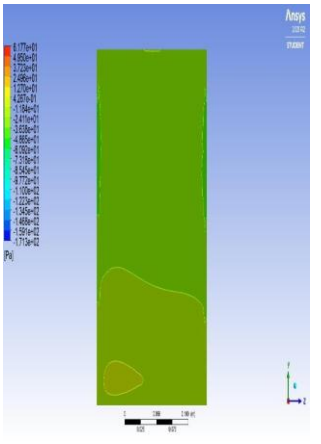


Circle Face D

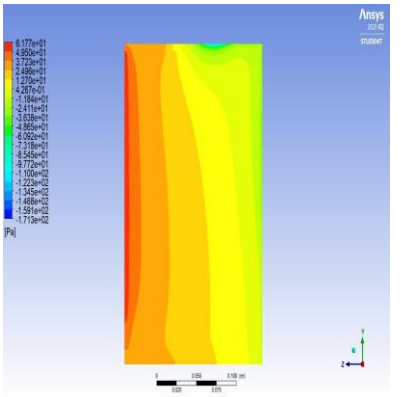
**Sixty-degree contour plot**



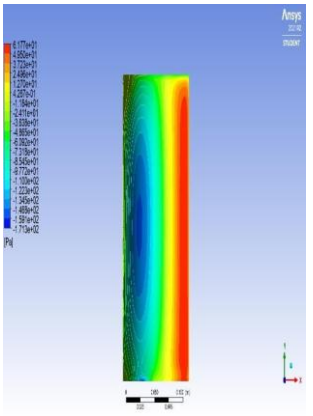
Triangle Face A



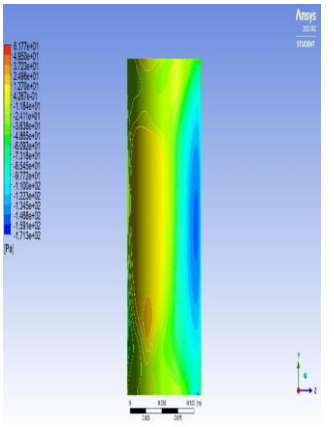
Triangle Face B



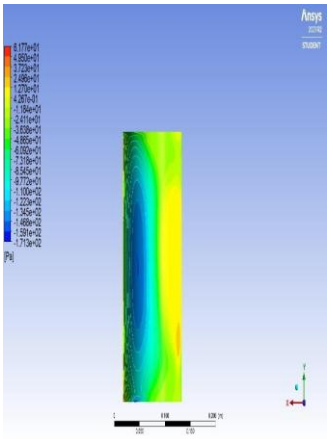
Triangle Face C



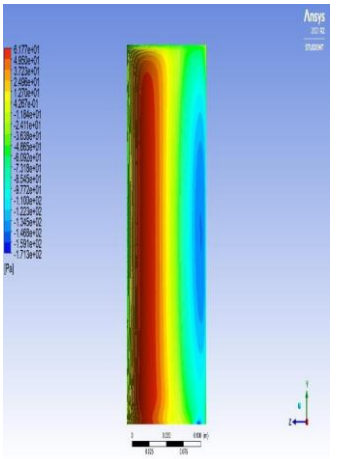
Circle Face A



Circle Face B

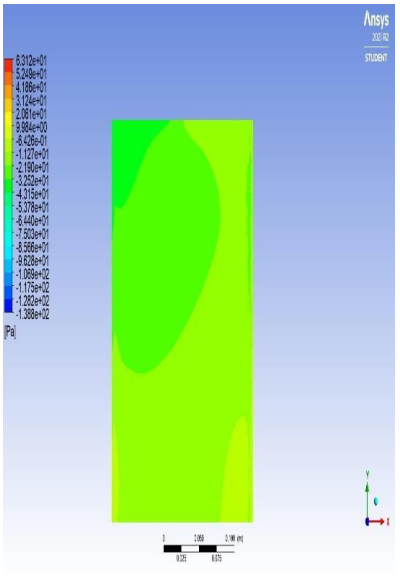


Circle Face C

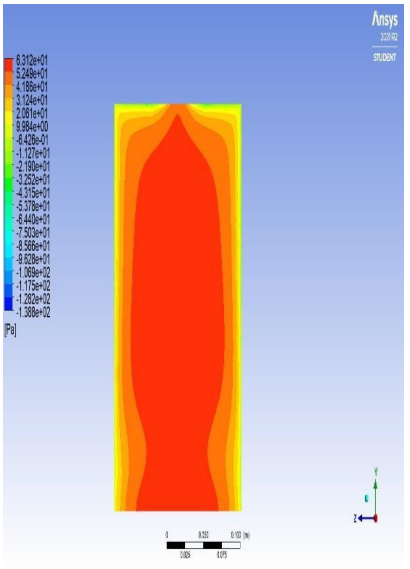


Circle Face D

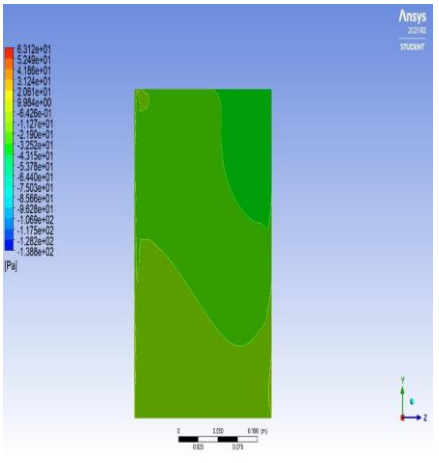
**One twenty contour plot**



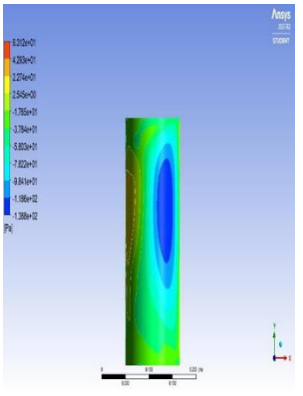

Triangle Face A



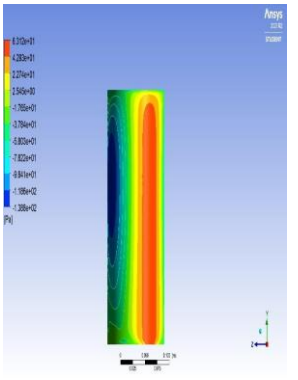

Triangle Face B



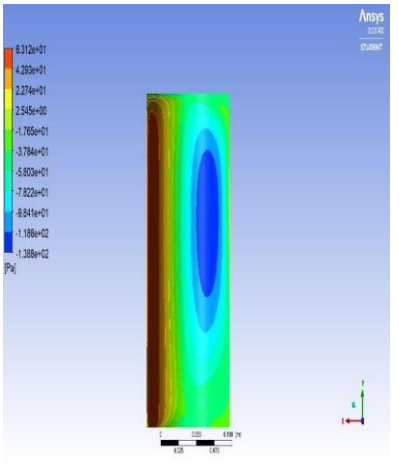

Triangle Face C



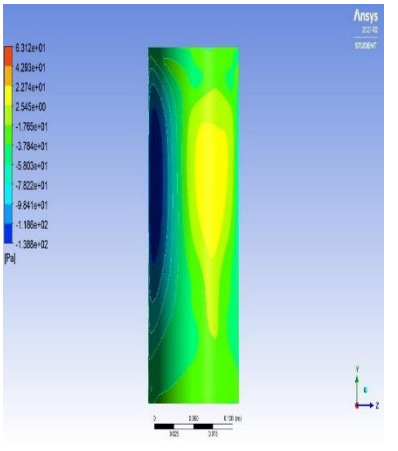
Circle Face A



Circle Face B

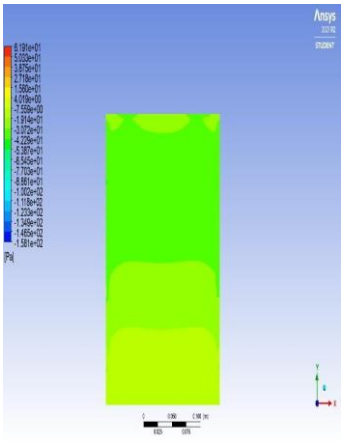


Circle Face C

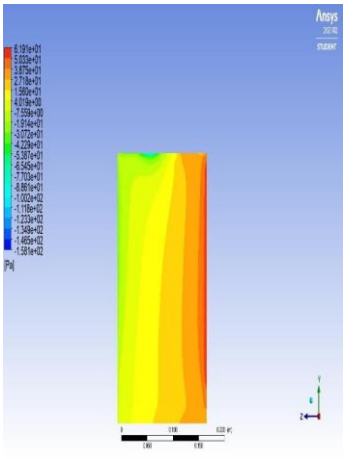


Circle Face D

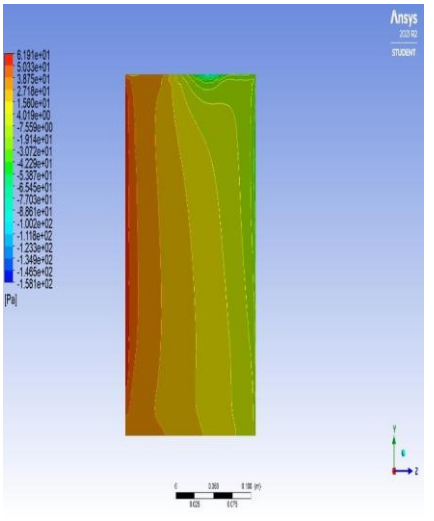
**One eighty contour plot**



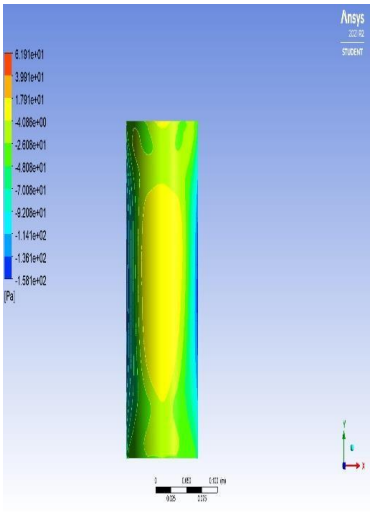

Triangle Face A



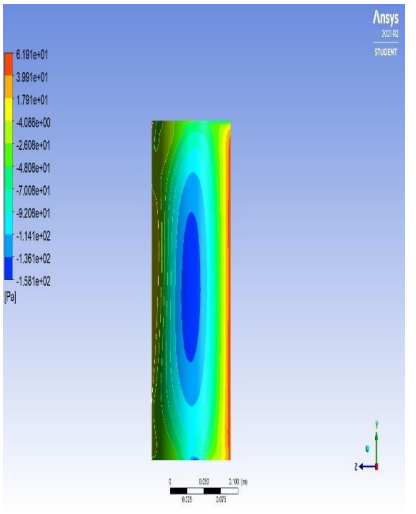

Triangle Face B



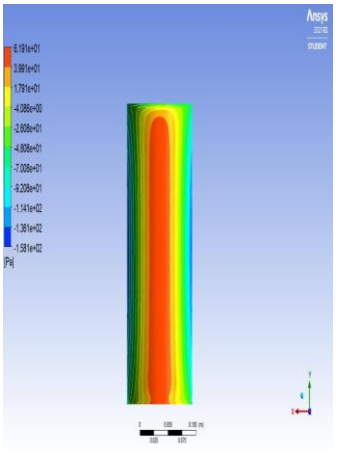

Triangle Face C



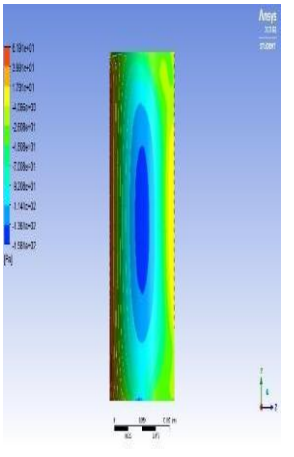

Circle Face A




Circle Face B




Circle Face C




Circle Face D

**Model 2 (250 Triangle 250 Circle)**

- The Cp Variation along the centerline for all the faces at different angle of incidences aredepicted in fig 5-29 to 5-35.
- For faces A of the triangular building the mean face average values of Cp at 0°,60°,120°and 180°angle of attack are +.78, -0.1, -.0.2, and -0.4 respectively.
- For faces B of the triangular building the mean face average values of Cp at 0°,60°,120°and 180°angle of attack are -.78, -.74, -.80, and -.72 respectively.
- For faces C of the triangular building the mean face average values of Cp at 0°,60°,120°and 180°angle of attack are -.48, +0.9, +.82, and -0.49

respectively.

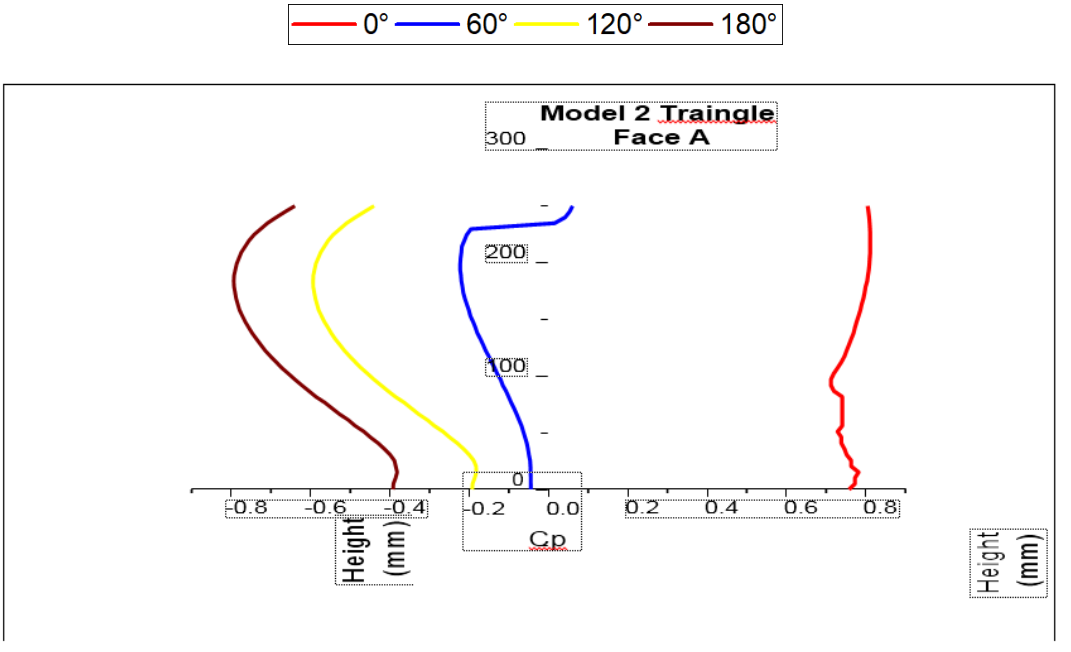



Figure 0-29 Pressure Variation along Centerline for face A for all degrees AOA of model 2

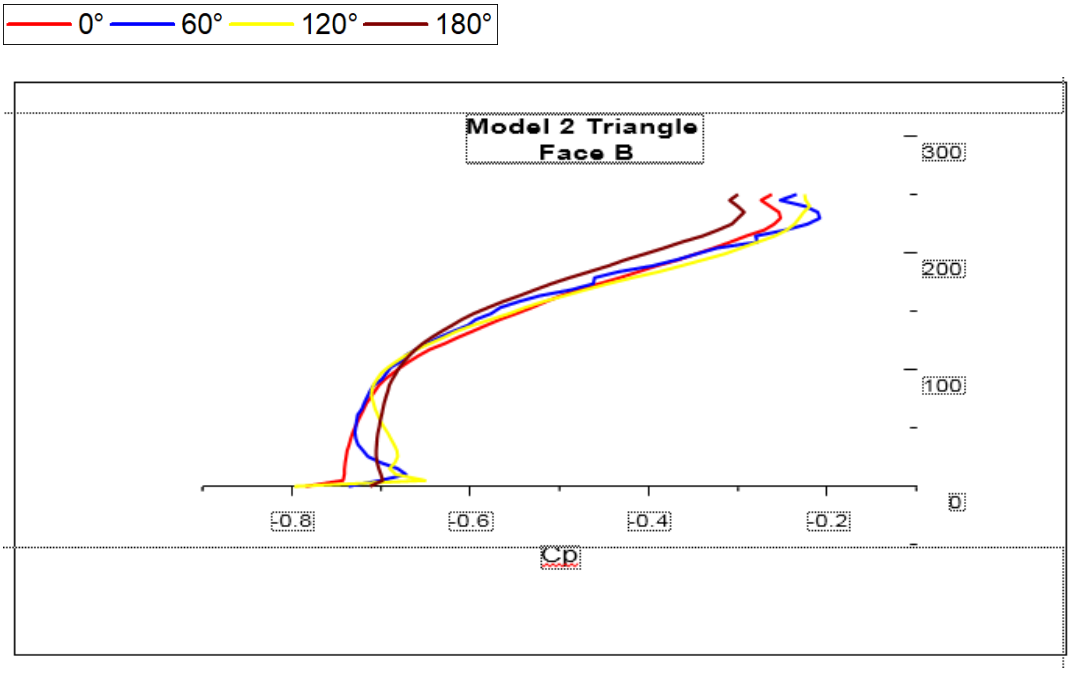



Figure 0-30 Pressure Variation along Centerline for face B for all degrees AOA of model 2

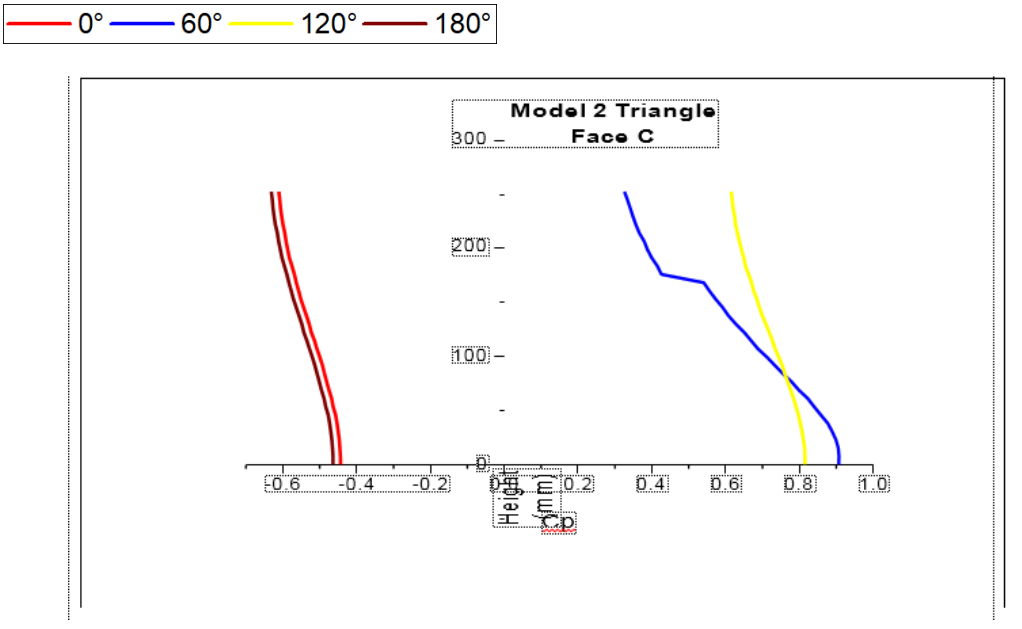
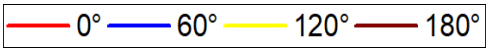



Figure 0-31 Pressure Variation along Centerline for face C for all degrees AOA of model 2

- For faces A of the circular building the mean face average values of Cp at 0°,60°,120°and 180°angle of attack are +1.2, -2.3, -0.3, and -0.2 respectively.
- For faces B of the circular building the mean face average values of Cp at 0°,60°,120°and 180°angle of attack are -2.3, -0.2, -2.2, and -2.4 respectively.
- For faces C of the circular building the mean face average values of Cp at 0°,60°,120° and 180°angle of attack are -.22, -2.25, +1.2, and +.80 respectively.
- For faces D of the circular building the mean face average values of Cp at 0°,60°,120° and 180°angle of attack are -2.5, +1.0, -2.0, and -2.25 respectively.



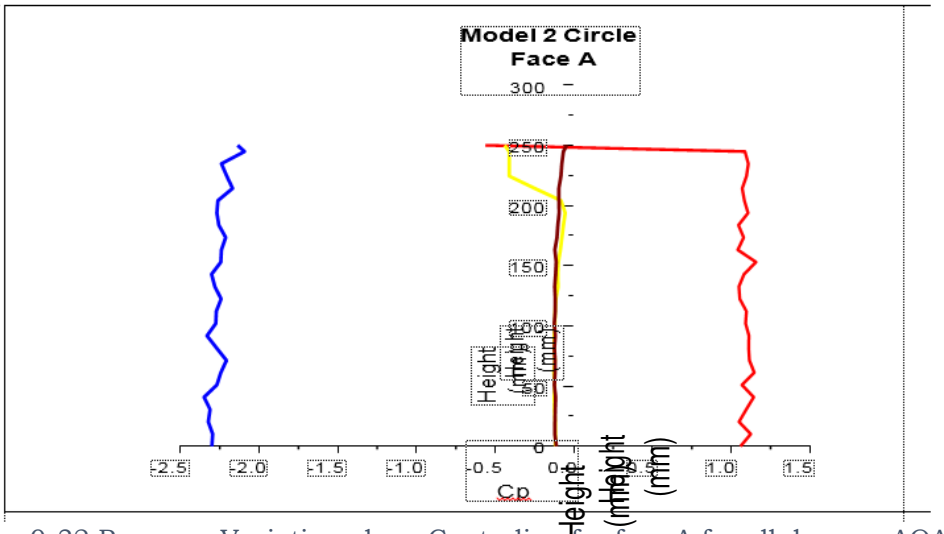
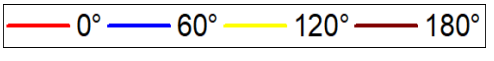



Figure 0-32 Pressure Variation along Centerline for face A for all degrees AOA of model 2



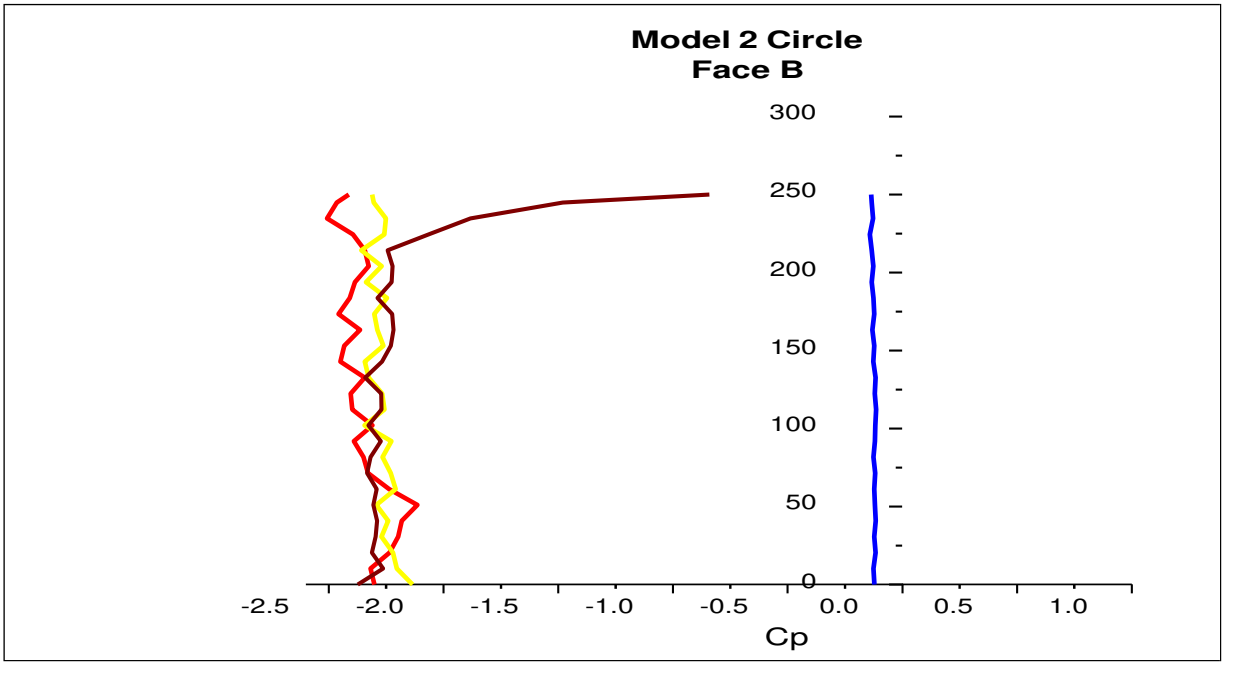
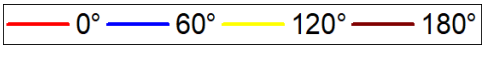




Figure 0-33 Pressure Variation along Centerline for face B for all degrees AOA of model 2



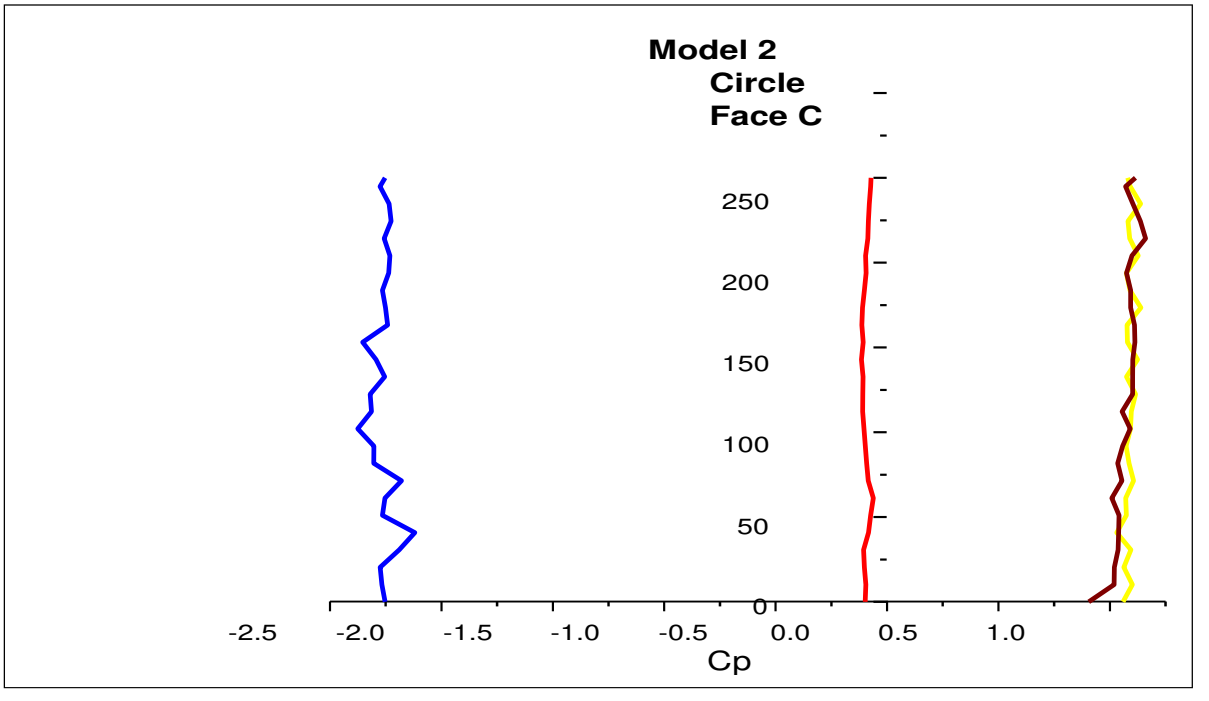



Figure 0-34 Pressure Variation along Centerline for face C for all degrees AOA of model 2

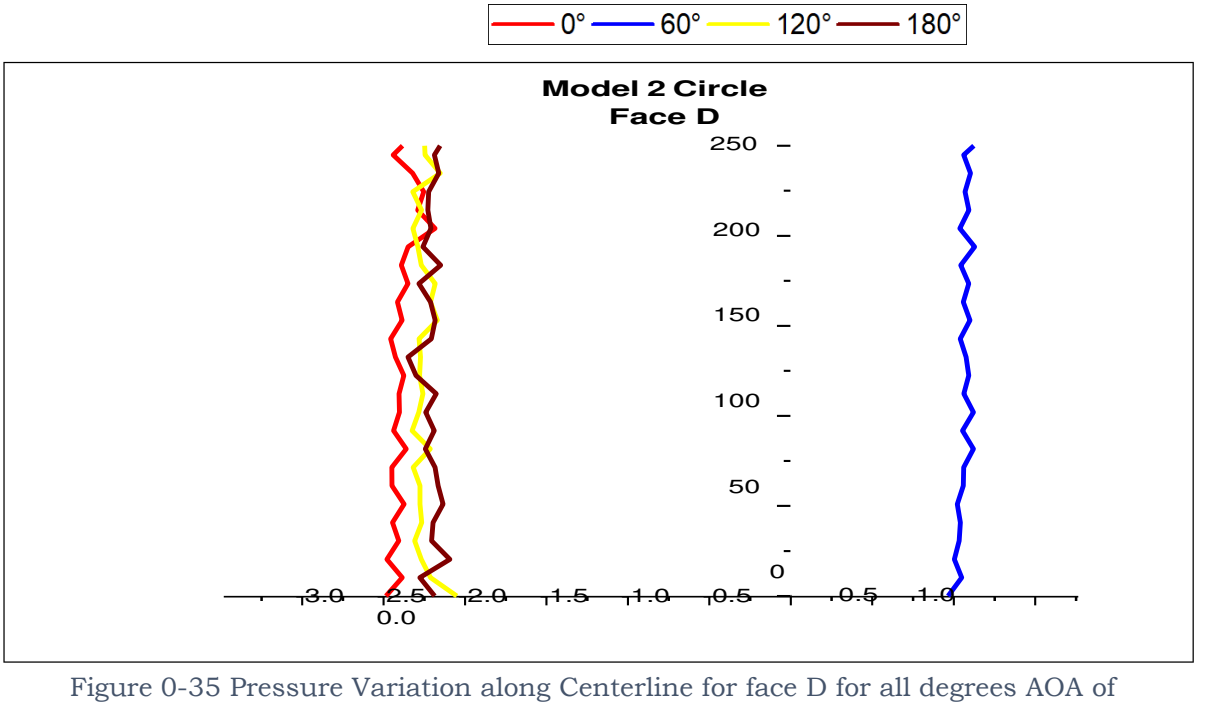


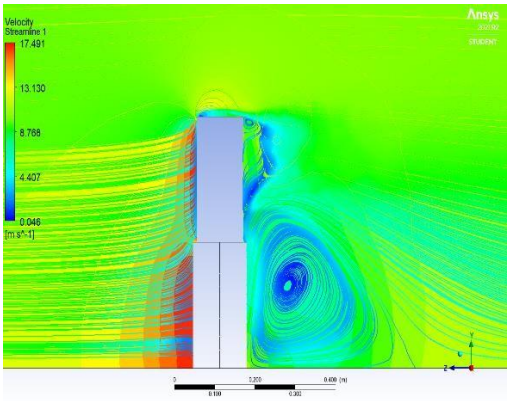

Figure 0-35 Pressure Variation along Centerline for face D for all degrees AOA of model 2

**Velocity Distribution**

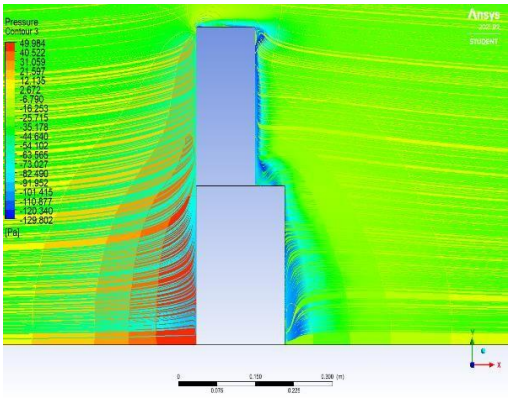
The velocity variation at windward faces of Building are as shown using velocity streamline.

**Model- 1 (250 triangle 250 circle)**

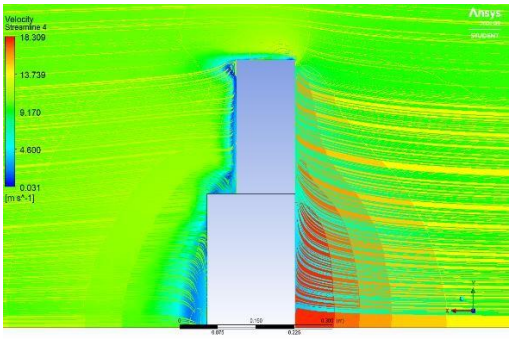
- Figures 6-1 to 6-4 exhibit the velocity distribution on different faces at different anglesof incidence.
- At windward face of the building wind velocity and wind pressure is maximum.
- At 0° and 180° vortex shedding is maximum and can be seen in the fig.
- At 60° and 120° vortex shedding is minimum.
- Wind velocity and pressure is varying at different angle of incidence.
- Red line shows the maximum wind velocity and wind speed whereas blue line showsminimum wind velocity and speed.



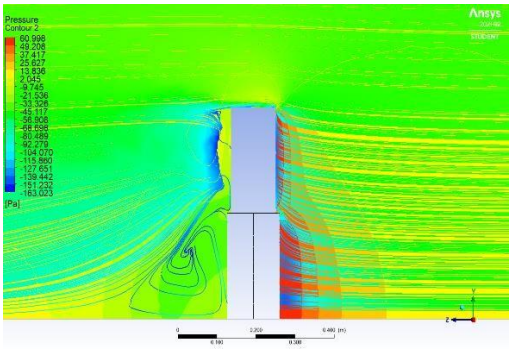

At 0°




At 60°




At 120°




At 180°

**Conclusion**

The pressure contour and mean pressure coefficients for the triangular and circular-shape building model for different height ratios at $0^0$ ,$60^0$ ,$120^0$ and $180^0$ wind incidence angles are compared in this paper. The *k*-ε model is used to simulate the results. The major finding of this research are as follows:

- The influence of height ratios and wind orientations on wind pressure distribution and magnitude of pressure coefficients on triangular and circular building models is identified by numerical study measurement of wind pressures on building models.
- The variation of pressure coefficients on the centreline is discussed and shown graphically.
- Comparison is made for numerical simulation data with various codes.

**References**


1. Nagar, S., & Ritu Raj. (2020). Experimental study of wind-induced pressures on tall buildings of different shapes. *Wind and Structures, Vol. 31, No. 5 (2020) 441-453*.
2. Ahuja, J. A. (n.d.). Effects of Side Ratio on Wind-Induced Pressure Distribution on Rectangular Buildings.
3. bailey, k. a. (1987-88). Effects of Aerodynamic Modifications of Building Shapes on Wind Induced Response of Tall Buildings.
4. Bairagi, A. M. (n.d.). Wind pressure and velocity pattern around 'N' plan shape tall building.
5. Dalui1, N. A. (n.d.). Aerodynamic analysis of pentagon-shaped tall buildings.
6. Dalui2, B. B. (n.d.). Experimental and Numerical Study of Wind-Pressure Distribution on Irregular-Plan-Shaped Building.
7. Dalui2, P. S. (n.d.). Comparison of aerodynamic coefficients of various types of Y-plan-shaped tall buildings.
8. Franke, J., & Hirsch, C. (n.d.). Recommendations on the use of CFD in wind engineering.
9. Kastner, P., & Dogan, T. (2020). Streamlining meshing methodologies for annual urban CFD simulations. *Journal of Energy and Buildings*.
10. M. Mallick, A. M. (n.d.). Modelling of Wind Pressure Coefficients on C-Shaped Building Models.
11. On the domain size for the steady-state CFD modelling of a tall building . (n.d.).
12. Pal, S., & Ritu Raj. (2021). Comparative study of wind induced mutual interference effects on square and fish-plan shape tall buildings. *Elsivier*.
13. Q.S. Li, J.Y. Fu, & Y.Q. Xiao. (n.d.). Wind tunnel and full-scale study of wind effects on China's tallest building.
14. R. Sheng a, b. L. (n.d.). Wind tunnel study of wind effects on a high-rise building at a scale of 1:300.
15. R. Sheng, L. Perret, & I. Calmet. (n.d.). Wind tunnel study of wind effects on a high-rise building at a scale of 1:300.
16. Suresh K Nagar1a, R. R. (n.d.). Experimental study of wind-induced pressures on tall buildings of different shapes.
17. R. Ahlawat and A. Ahuja, "Wind Loads on T Shape Tall Buildings," *J. Acad. Ind. Res.*,vol. 24, no. 1, p. 257922, 2015.
18. R. Ahlawat and A. K. Ahuja, "Wind loads on Y plan shape tall building," *Int. J. Eng. Appl. Sci.*, vol. 2, no. 4, p. 257946, 2015.
19. IS: 875 (2015), *Indian Standard design loads (other than earthquake) for buildings and structures-code of practice,part 3(wind loads)*. 2015.
20. ASCE: 7-10(2013), *Minimum Design Loads for Buildings and Other Structures. Structural Engineering Institute of the American Society of Civil Engineering,*

*Reston.*2013.
21. GB 50009-2001, *NATIONAL STANDARD OF THE PEOPLE'S REPUBLIC OF CHINA*. 2002.
22. AS/NZS:1170.2(2011), *Structural Design Actions - Part 2: Wind actions. Standards Australia/ Standards New Zealand, Sydney*. 2011.
23. Hong Kong Building Department, "Code of Practice on Wind Effects in Hong Kong 2019," 2019.

Technology Analysis & Strategic Management

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ctas20

# Agri-tech: experiential learning from the Agri-tech growth leaders

Sanjay Chaudhary & P. K. Suri

**To cite this article:** Sanjay Chaudhary & P. K. Suri (2022): Agri-tech: experiential learning from the Agri-tech growth leaders, Technology Analysis & Strategic Management, DOI: 10.1080/09537325.2022.2100755

**To link to this article:** https://doi.org/10.1080/09537325.2022.2100755

Published online: 14 Jul 2022.

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=ctas20

TECHNOLOGY ANALYSIS & STRATEGIC MANAGEMENT
https://doi.org/10.1080/09537325.2022.2100755

Routledge
Taylor & Francis Group

Check for updates

# Agri-tech: experiential learning from the Agri-tech growth leaders

Sanjay Chaudhary [a] and P. K. Suri[b]

[a]Delhi Technological University, Vivekananda Institute of Professional Studies, Pitampura, India; [b]Delhi School of Management, Delhi Technological University, Delhi, India

**ABSTRACT**

This research study analyses the critical issues in Agri-tech start-up growth in India and explores ways to improve the process. Research methods are triangulated to understand Agri-tech entrepreneurs' opinions and experiences. After the articles and reports were reviewed, a statistical analysis of a user opinion survey of two hundred and two Agri-tech users was conducted. Post survey, three rounds of semi-structured interviews among the twenty-two Agri-tech entrepreneurs and experts helped in fine-tuning the findings. The key issues highlighted in the findings are the availability of low-interest and creative financing models; Infrastructure availability and incubation support; Cross-domain quality solutions with multiple levels of features; Supportive mindset of business customers and partners; and Skill-building and Training of farmers. This research is one of the few academic studies on Agri-tech entrepreneurship and includes the COVID-19 pandemic effect. The findings help the researchers build the knowledge base of agriculture 4.0 entrepreneurship. Suggestions may help the practitioner boost the growth of Agri-tech solutions. In addition, the study highlights innovative suggestions like setting up an open national database of core parameters, cloud computing-based agriculture 4.0 digital stack, digitalising farmer producer organisations, and monitoring plus organisational evaluation setup.

**ARTICLE HISTORY**
Received 13 January 2022
Revised 3 May 2022
Accepted 29 June 2022

**KEYWORDS**
Agriculture; Agri-Tech; digitalisation; start-up

## 1. Introduction

The agricultural sector is witnessing a digital revolution to increase productivity, lower costs, and reduce environmental hazards. The COVID-19 pandemic has further forced economies to improve the resiliency and security of their agricultural supply chains (Boston Consulting Group [BCG] 2020).

Several initiatives by the large corporate entities have set the stage for massive information and communications technologies (ICT) based digital transformation of the agricultural practices in India. e.g. Bayer, Cargill, IFFCO, ITC, Monsanto, and Unilever. The initiatives are supported by the ease of regulations, and the government-initiated large projects in e-trading, artificial intelligence, digital literacy, financial inclusion, and mobility.

Agri-tech start-ups are at the forefront of the changes as the growth catalyst and acceleration enablers. The subset of these innovative initiatives, such as precision agriculture and agricultural e-commerce platforms/e-marketplace, may significantly affect the Indian economy to the tune of $32 billion and make their presence felt during a pandemic. Agri-Tech start-ups help the farming community increase the farm's produce, yield, and income generation (Mckinsey 2014; Sharma and Mathur 2019).

**CONTACT** Sanjay Chaudhary one.sanjay@gmail.com Delhi Technological University, and Delhi School of Business, Pitampura, Delhi 110034, India

© 2022 Informa UK Limited, trading as Taylor & Francis Group

In India, approximately six hundred agri-tech start-ups are leading the digital transformation. Investment in Indian agri-tech during 2014–2019 amounted to USD 1.7 billion, with a 48% annual compounded growth rate. The Agritech market potential in India is estimated to be USD 24 billion in 2025. Market opportunities are high since only 1% market penetration was achieved in 2021 (Ernst and Young 2020).

The state of the agri-tech ecosystem differs from country to country. Several countries and stakeholders are implementing Agritech, while many have yet to formulate a common approach to address the phenomenon. An early comprehensive approach at the national level may help rationalise resources, address opportunities and challenges, add new revenue streams, and ultimately improve the living standards of the farming community engaged in agriculture.

### *1.1. Research gap*

Akin to the developing countries, the Indian agriculture sector has a sizeable unorganised market, inefficient supply chain, limited logistics facilities, and finance-related problems. These issues may be solved using information technology (IT) in general and Agri-tech in particular (Seth and Ganguly 2017). However, compared to developed countries, India's level of digitalisation of farming activities is low. The skill level and affordability of farmers are also low. Agri-tech start-ups, in general, have a small size, limited financial reserves, and low portfolio diversity. Against this backdrop, it becomes vital to formally analyze the key issues facing Agri-tech start-ups in India and suggest actions for practitioners to bridge the gap in multidimensional support, leading to the survival of Agri-tech start-ups.

Furthermore, compared to the developed countries, there is a considerable gap in policy and institutional support for Agri-tech start-ups, whereas the performance demand is equal or more (Desai et al. 2011). Post-pandemic economic reality may also differ.

### *1.2. Research question*

The research gap leads to the research questions. What are the issues in Agri-tech start-ups' growth? Are these issues prioritised so that limited resources and support may be provided at the time when the COVID-19 pandemic is straining the economic resources? Do Agri-tech start-ups need differentiated help?

### *1.3. Potential contribution*

The study's knowledge base may help a researcher in India and other developing countries. The study also addresses the academic debate by providing insight into the struggles and issues of agri-tech entrepreneurs. The user survey results may help policymakers devise entrepreneur and user-friendly policies that help solve the priority issues. This study is among the few academic studies with successful entrepreneurs' opinions on this topic and tries to address the effects of the COVID-19 pandemic. The practitioners may use the experiential learning from the study to devise a growth strategy.

## 2. Research objectives

The objectives of the study are:

- Analyze and prioritise the critical issues in the growth of Agri-tech start-ups in India.
- Suggest a feasible way ahead for Agri-tech start-ups' growth.

The research article is organised as follows: The research methodology section exhibits the sample details and the statistical analysis to achieve the research objectives. The literature review

section presents the recent development in agri-tech activities and the critical issues in the current scenario. The findings and suggestions section prioritises the issues for agri-tech start-ups and details the implications for the researchers and practitioners. It suggests ways to overcome them. Finally, the article concludes with recommendations in the conclusion section.

## 3. Research methodology

A triangulated approach is adopted. The triangulation approach includes a review of relevant literature, detailed interviews cum opinions of the select start-ups' chief executive officers (CEOs), and a user opinion survey statistical analysis.

Two questionnaires (semi-structured and structured) were developed after reviewing recent scholarly articles, reports, and relevant news articles.

The structured questionnaire is used to survey two hundred and two Agri-tech users in the Indian states of Delhi and Uttar Pradesh. The two hundred and two Agri-tech users were selected for the opinion survey using simple random sampling. The respondents were randomly selected using a lottery system from the database available with the Agri-tech Centre of Excellence at a state university in India. The survey was conducted online using Google Forms in August 2021.

The semi-structured questionnaire-based interview process is used to understand and clarify current issues and explore solutions from twenty-two entrepreneurs and experts (Creswell 2007). The respondents who volunteered the reply formed a subset of fifty agri-tech entrepreneurs and expert lists collated from the Agri-tech reports by Ernst and Young (E&Y), Federation of Indian Chambers of Commerce and Industry (FICCI), and the National Association of Software and Service Companies (NASSCOM), using purposive sampling/researcher judgment (E&Y 2020; FICCI 2018; NASSCOM 2020). The interviews were conducted for one year period up until July 2020.

The 'keywords' were generated from the review of the research articles using the 'Word Map' feature of NVivo software. Again, the themes were conceptualised around the keywords by coding content from the interviews and ranked using the frequency of the keywords based on field notes in NVivo software. The ranking of themes was cross-checked through an opinion survey of Agritech users.

This qualitative study's trustworthiness was established using Lincoln and Guba's Evaluative Criteria of credibility, transferability, dependability, and conformability (Lincoln and Guba 1985).

The credibility of qualitative research was established using referential adequacy. The sample size was kept at twenty-two for interviews and opinions. No new information was elicited by adding more respondents. Overall, the total number of seventeen interview respondents and five expert opinions on the findings is within the suggested range of 15–30 in qualitative information systems research (Astalin 2013; Creswell 2009; Marshall et al. 2013; Morse 2000).

The transferability was established using the subjective explanation along with the meaning obtained through semi-structured interviews of seventeen participants in India.

The study's dependability was established using a second three-member expert panel with members from the Agriculture University, Institutes, and Corporates. The second expert panel helped audit and fine-tune the findings of the study. The feedback of the experts also helped in terms of suggestions. The study's confirmability was established using reflexivity. The study situation was approached from distinct positions or perspectives: entrepreneurs engaged in the various supply chain segments, experts from academia, agricultural development institutes, agricultural research institutes, industry, incubators, and farmers. Using respondents from multiple fields led to varied opinions but an equally accurate understanding of the topic. Two principal researchers' dialogue led to complementary knowledge of the phenomenon, which debated hidden beliefs and perspectives.

## 4. Literature review

Agri-tech start-ups are disruptive and provide a helpful product, service, or application as a solution. The digital solution pertains to Agriculture 4.0 (Food and Agriculture Organization [FAO] 2020; World

Government Summit [WGS] 2018) and changes how the segments work in the agricultural supply chain, from sourcing to production to distribution, processing, and retailing.

### 4.1. Agri-tech solutions

Agri-tech solutions with ICT at its core may improve access to assets, markets, and services across the value chain (Figure 1), which has accelerated adoption during and after the COVID-19 pandemic (ThinkAg 2020).

The data analytics solution involves farm mapping, farm management (soil and crop health diagnostics), drones or tractor-based data on weather/crop to forecast or assess risk using Artificial Intelligence (AI), customer relationship management, blockchain traceability, and compliance. The infrastructure enablers are drone-enabled vertical farming monitoring solutions and aquaponics (Internet of Things (IoT) enabled smart farming). It also includes high-precision crop control, data collection/analysis, automated farming techniques, hydroponics (IoT enabled the farming ecosystem to monitor humidity and air temperature), and drip irrigation. The farming as a service (FAAS) solution involves on-demand harvesting, digital payments, market pricing, and agricultural machinery platform (renting, crop care practices). In contrast, the finance solutions cater to fees, revenue sharing, insurance, and lending features (FICCI 2018).

### 4.2. Status of Agritech

Due to rapid growth in recent years, the number of Indian Agri-tech start-ups grew to approximately 450 start-ups in 2019, witnessing a yearly increase of 25% during the period 2016–2019. These 450 start-ups may be considered average compared to the 3103 start-ups globally that started during five years up to 2019. In 2019, more than 50% of start-ups focused on supply chain efficiency solutions and obtained more than 50% of the investment. Digital supply chain solutions remain an attractive investment option in the COVID-19 economy (AgFunder 2020; NASSCOM 2019).

India's Agri-tech industry can reach USD 24 billion in revenue by 2025 and significantly improve the current 1% market penetration. Technology majors such as Cargill, Bayer, IBM, Microsoft, Monsanto, and Cisco have also launched projects across the agricultural supply chain (Ernst and Young 2020; NASSCOM 2019; Sarangi 2020).

The net effect of the COVID-19 pandemic on Agri-tech is estimated to be positive in the post-pandemic economy. Agri-Tech still provides an attractive opportunity. Large technology companies are

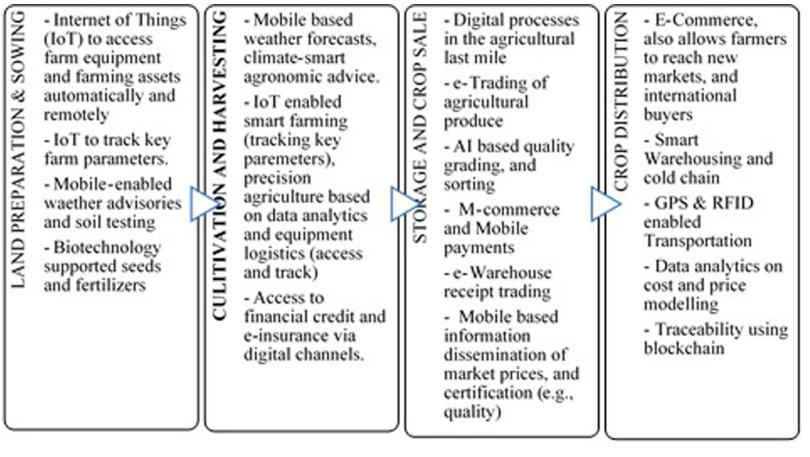



**Figure 1.** Digital solutions in agricultural value chain.

now targeting start-ups to improve their offerings for the post-COVID-19 pandemic market (AgFunder 2020; BCG 2020; MarketsandMarkets 2020).

Major drivers for the increase in agri-tech start-ups include the following: growing rural mobile penetration, increase in rural purchasing capability, the rise of farmer producer organisations, changing food consumption patterns, upgrades in logistics, improvements in digital infrastructure, regionalisation of the supply chain, reduced cost of components (mobile devices, data connectivity, sensors, vernacular content technology, and robotics) and cloud computing technology (Kalaari 2018; Lele and Goswami 2017).

### 4.3. Need for a monitoring and evaluation framework

The Indian agricultural innovation ecosystem consists of 74 Agricultural Universities, 65 Indian Council of Agricultural Research (ICAR) institutes, fourteen national research centres, six bureaus, thirteen project directorates, and 79 All India Coordinated Research Projects (AICRPs), 716 Krishi Vigyan Kendras (KVK). Besides this, there are about twenty-three general universities with agriculture faculty, private organisations, nongovernment organisations (NGO), and scientific societies (National Institute of Agricultural Extension Management [Manage] 2020).

So far, the core coordination work has been done at ICAR. But, when it comes to Agritech, there are a lot of activities outside the ICAR ecosystem. Experts are emphasising the continuous monitoring and evaluation of the Agritech sector initiatives. At the national level, a comprehensive monitoring and evaluation framework (Figure 2) will help policymakers track progress and decide what needs to be done to achieve the intended common objectives for the Agritech sector. The monitoring and evaluation framework will also help keep stakeholders on board to address critical issues, e.g. Agritech growth in the post-COVID-19 era.

In the suggested framework, monitoring refers to using data to measure the extent of progress, utilisation of funds, and achievement of the objectives. At the same time, the evaluation refers to determining the impact of activity per its goals and objectives. The output indicators provide information on the adoption of Agritech. The outcome indicators offer input into the results for stakeholders, e.g. expected improvement. The indicators are objective, quantifiable, comparable, controllable, and linked to activities.

## 5. Key issues in Agritech

The critical issues in the further growth of Agri-tech entrepreneurship in India are explored using content analysis. The 'Word Map' feature in NVivo software-generated keywords in the shortlisted

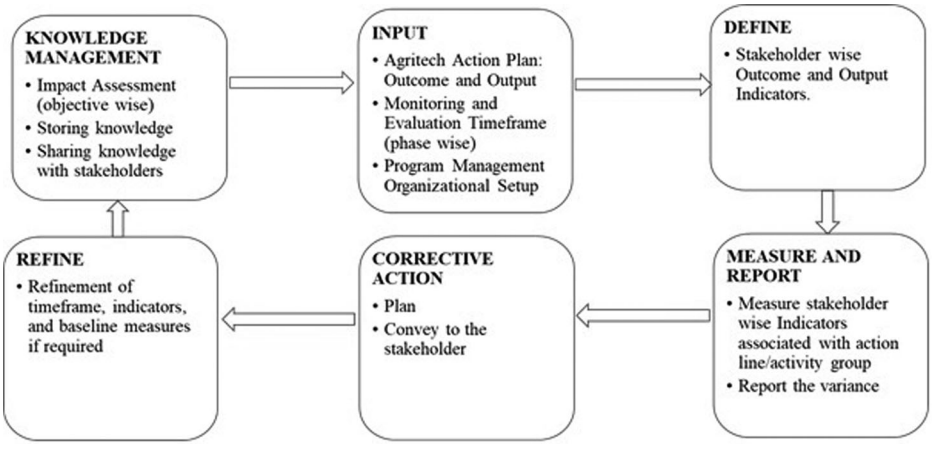



**Figure 2.** Conceptual monitoring and evaluation framework.

**Table 1.** Key issues in the growth of Indian Agri-tech start-ups.

| S. No. | Issue | References |
|---|---|---|
| 1 | Infrastructure availability and incubation support | (Anupam and Saravanan 2019), (Bose and Kiran 2014), (Food and Agriculture Organization [FAO] 2017), (FICCI 2018), (IBEF 2018), (Mor et al. 2015), (NASSCOM 2018, 2019), (Sharma and Mathur 2019) |
| 2 | Availability of low-interest and creative models of financing | (Agfunder 2020), (Kalaari 2018), (Ministry of Agriculture and Farmer Welfare [MOAFW] 2018), (NASSCOM 2020), (Subramaniam 2020) |
| 3 | Supportive mindset of business customers and partners | (Anupam and Saravanan 2019), (Arafat et al. 2020), (India Brand Equity Foundation [IBEF] 2018), (Kundu and Joshi 2014), (NASSCOM 2019), (Sachitanand 2018), (Sharma and Mathur 2019), (Singh 2013) |
| 4 | Cross-domain quality solutions with multiple levels of features | (Chaudhary and Suri 2021), (FICCI 2018), (MOAFW 2018), (NASSCOM 2020), (Sachitanand 2018), (Sarangi 2020), (Subhash et al. 2016) |
| 5 | Skill-building and training of farmers | (Crawford 2020), (D'Cunha 2018), (FAO 2017), (Sarangi 2018), (Sharma and Mathur 2019), (Singh 2013), (Tohidyan Far and Rezaei-Moghaddam 2019) |
| 6 | Lack of subject matter experts/mentors | (Anupam and Saravanan 2019), (Bose and Kiran 2014), (FICCI 2018), (Subramaniam 2020) |
| 7 | Low level of digitalisation and information availability | (Ciruela-Lorenzo et al. 2020), (E&Y 2020), (Sarangi 2018), (Seth and Ganguly 2017) |
| 8 | Fintech integration and support | (Anupam and Saravanan 2019), (FICCI 2018), (Fintech news 2020), (NASSCOM 2020), (Reddy and Kumar 2020), (Shah 2020) |
| 9 | Talent retention | (NASSCOM 2018, 2020) |

academic articles, reports, and case reviews. The themes were conceptualised around the high-frequency keywords as the add-on, but a similar analysis is done for the CEOs and expert Interviews detailed in the following subsection. The revealed themes/issues are listed in Table 1.

There is a need to build awareness among youth about opportunities in agri-tech start-ups. Among existing agri-entrepreneurs, there is an awareness gap between government schemes and policies (Bairwa et al. 2014). The farm management, agricultural consultants, rural real estate appraisers, and agri-property agents may be sensitised with increased awareness of recent agri-tech solutions, even though initial services may seem piecemeal (Grant and Zhang 2019).

The inadequacy of marketing, warehouse, and logistics infrastructure is also felt in the agri-tech space. Farmers now wish to know more about technological advances as mobile phones reach the masses. More than 25% of farmers in India have internet-enabled smartphones. If these devices are backed by affordable mobile broadband tariff plans and training, it may help improve farm activities' workflow related to agri-tech (NASSCOM 2019).

According to Nasscom Analyst, many Indian start-ups, including Agri-tech, face cash flow and funding crunches due to the COVID-19 crisis. The crisis may be felt until mid-2022 when it calls for the government to support access to working capital, ease the taxation process, and ease compliance plus fiscal policy. To help agri-tech businesses deal with prolonged gestation periods and ecosystem hindrances, subsidy benefits, with softer loans under priority sector lending, may help (Godhara, Sharma, and Godhara 2020; NASSCOM 2020).

In need of a cross-domain solution, the gap may be filled by the collaboration of government, the public, start-ups, and enterprises, plus transparency in the design and implementation of solutions not yet seen in the sector. Most start-ups work in silos of agriculture input sales, information sharing services, or e-commerce. Agricultural solutions need to be integrated to provide participatory planning, procurement of inputs, information sharing, produce processing, and output marketing. Such integrated solutions will lead to high adoption if bundled with social marketing in the local language. The rebound in business post-COVID-19 offers numerous opportunities for joint efforts and benefits from the industry's recovery (BCG 2020; Moni and Sharma 2018).

Agri-tech entrepreneurs also felt a shortage of mentors and talent. Experienced mentors do not have enough domain knowledge of Agri-tech. Those who do, lack depth in financial guidance. (Anupam and Saravanan 2019). Agri-tech start-up technologies' principal attraction lies in their farm-size agnostics, which are thus beneficial to Indian farmers with small landholdings. The

government may help them spread their innovations and services across the farming community (Singh 2013). The Agricultural Science Centres (KVK), under the Indian National Agricultural Research System (NARS), may be a vital organisation to achieve this intent.

## 6. Findings and suggestions

### 6.1. Discussions

To further understand the views and opinions of e-commerce platform-based Agri-tech start-ups,[1] Chief executive officers (CEOs) interviews were conducted across India. The selected excerpts are mentioned in the following text.

An executive from Company 'A' said,

> Hand-holding agri-tech start-ups by large corporations in public and private sector is required. The company has onboarded several large retail chains in its customer base. Nevertheless, the operational help beyond purchase shall help to scale up the operations to the next level.

According to the Company 'B' executive, '*the COVID-19 pandemic has affected our business marginally and forced us to innovate using social commerce, where we have enrolled more than 500 new age resellers*'.

Company C's executive emphasised, 'A need for support by the agri-incubators to fine-tune the revenue model and provide grants'. An executive from Company 'D' highlighted: 'There *are a limited number of venture capitalists addressing this sector. Bank funding is also not easily available. Hopefully, the RAFTAAR scheme is beneficial*'.

Executives from Company 'E' and 'F' urged that '*easy and economic finance availability to farmers and customers may increase market size. The government should take a requisite stance at the World Trade Organization in protecting local farmers and increasing exports*'. According to Company 'G' executive, '*financing has become difficult during the COVID-19 pandemic. Start-ups are engaging in the costlier option of venture debt financing*', and the government needs to be aware of the pandemic effect.

According to Company H's CEO, '*the digital platform serves the needs of all stakeholders in the farm-to-fork supply chain*'. He reiterates: '*We are constantly upgrading our business model every three years*'. He added that 'technology *shall play a key role in future success*'. According to Institute 'I' expert, '*Quality management and reliable supply chain are a big challenge*'.

As mentioned by the CEO, the challenges faced by Company 'H' include '*low liquidity in the market post demonetization, average farm productivity, low traceability in the agri-supply chain, the need for quality assurance and credit linkages for farmers*'. To overcome the challenges, the start-up has partnered with several banks for commercial finance for farmers and traders. In addition to promoting credit guarantees and warehouse receipt financing, it has also started a first-of-its-kind agribusiness-focused non-banking financial company (NBFC). Farmers are provided with good-quality hybrid seeds and free advice on good farm practices and crop cultivation to boost productivity.

Traceability in the supply chain is introduced via data science-based IT solutions. Quality assurance via the quality lab is also ensured in every significant sourcing location. The expectation from the government is '*to create a conducive environment for investment in the agri-tech and agri-business sectors and fill the infrastructure gaps in the supply chain*'.

For the COVID-19 pandemic, the CEO, Company 'H', comments, '*Pandemic is an opportunity to showcase how Agri-tech can solve the problem related to yield, procurement and shortening the Agri-supply chain*'. Company 'H' has witnessed an increase in registrations. The start-up facilitated interstate trade and supported more than five thousand trucks of produce across India. He asked the government to '*address the labour issues during and post-pandemic*'.

Company 'J' executive mentioned that

> the Agri-tech ecosystem is not yet organized. Thus, a start-up needs to be flexible and put in extra resources. Additionally, the start-ups' supported incubators are too few, and funding is also low. To make them viable, the funding amount (with low/zero interest rate) must increase significantly.

He also mentioned that

> state-level agricultural trade policies are restrictive. Thus, agri-tech start-up solutions work fine in the pilot stage. However, while commercializing, the problem is scaling them to the national level. There may be a need for a uniform national agricultural trade policy, plus cheaper financing options for farmers and traders. Low-cost finance may help increase users' purchasing power and the long-term viability of start-ups.

The CEO of Company 'K' echoed similar sentiments: 'The Agri sector requires investors with a long-term view. It, along with commercialization support from large corporations, helps start-ups in this sector'. During the COVID-19 pandemic, he shares that 'district agricultural officials cannot venture out due to lockdown. During this time, the company shares climate-smart advisories, education videos, training content, data management, and analytics to farmers, bedside goods, and monetary help'. The government may partner with such applications for weather-sensitive advisories and help farmers with farming inputs.

The entrepreneur from Company 'L' asked for support from the public and private companies to adopt technologies and solutions from start-ups with preferential treatment. The person said, '*Established customers and companies may change their mindset toward the ability of start-up solutions*'. The expert from University 'M' makes a similar comment: '*Business Models of Agri-tech start-ups are presumed to be "rigid" and hence one of the most significant hindrances to being networked with large corporations, incubators, and accelerators*'.

The Company 'N' executive voiced that '*the accelerators shall be able to help start-ups commercialize across geographies*'. According to an executive from Company 'O', '*Activity in the agri-tech sector increases post the COVID-19 lockdown, as they fill the infrastructure and ecosystem gaps with technology*'. Institute 'P' expert expressed that '*accelerators may introduce start-ups to Indian and other Asian companies engaged in large-scale commercial farming in Africa. Such initiatives may help in a massive improvement in African agricultural supply chain and international business*'.

At Institute 'P', the Incubator expert highlighted that '*Academia should encourage more Agri-entrepreneurs in its institutions – the academic initiatives, such as Agri-hackathons aggregate aspiring and early-stage entrepreneurs from diverse sectors*'.

Experts from a think tank (Institute 'Q') mentioned that '*the Agri-tech start-ups' solutions to sync with the infrastructure available on the ground in India and the business model need to be flexible*'. The University 'R' expert verbalises a similar sentiment:

> Many Agri-tech start-ups work on a component of the Agricultural Value System (AVS) and hence face many challenges in produce's backward and forward linkages. Seamless integration of Agri-tech, Digital Tech, and FinTech is missing; thus, Agri-tech start-ups lack an appropriate ecosystem for their growth and proliferation.

On the impact of the COVID-19 pandemic, he says

> Farming operations were exempted from the COVID19 Lockdown restrictions. Initially, there were logistical issues in the transportation of farm produce from farm gate to retailers. Farmers also faced 'demand-side' problems. Those Agri-tech start-ups working in the e-Commerce and Logistics discipline were able to facilitate their operations during the COVID19 Lockdown period.

*An* executive from Company 'S' also felt that 'storage, *transportation, and IT connectivity have to ramp up before such hi-tech solutions have a significant impact*'. The Institute 'T' expert opinions that '*Agri-tech has the potential to increase farmers' income significantly*'. It is also mentioned that '*the challenges linked to the financing of agri-tech are still not addressed by the government start-up financing schemes like Start-Up India*'.

To extend its support, the government may apportion a certain amount for the Agri-tech start-ups in the Small Industries Development Bank of India (SIDBI) start-up funds. Such initiatives may be parked under the Krishi Kalyan Abhiyan (KKA) initiative of the Ministry of Agriculture Cooperation

and Farmers Welfare (MACFW), with Krishi Vigan Kendra (KVK) as the nodal centre. '*The digital skills of users and ecosystem partners also need to be enhanced to match the solution offerings and realize the desired benefits*', as per the opinion of an expert from Company 'U'. An expert from University 'V' verbalises a similar view: '*Farmers with limited landholding/buying capacity need to be made aware of solutions providing access to cheap and reliable technology-driven agri-equipment on a sharing basis*'. The E-Krishi Yantra Anudan scheme (launched in July 2020) may be a positive step in this direction by the government.

### 6.2. Findings

The vital issues in Agri-tech growth are further shortlisted and ranked using the opinion survey of two hundred and two end-users. The seventeen agri-tech entrepreneurs and experts additionally ranked the issues for priority. The critical issues raised are ranked and summarised in Table 2.

The ranking by the CEO survey and the opinion survey of users, experts, and small-scale entrepreneurs are the same. The additional insight from the user opinion survey is that the agri-tech users and experts feel that 'Lack of Subject Matter Experts/Mentors and Talent Retention is as vital as 'Skill-Building and Training of Farmers'.

## 7. Suggestions

The implications of the findings for the researchers, practitioners, and ways to overcome them are suggested below:

### 7.1. Theoretical implications

The findings of the survey and interviews prioritise the critical issues. The study is among the first in the agri-tech sector that puts together the opinions of entrepreneurs and experts. It adds to the existing academic literature and fills the knowledgebase gap concerning the Agritech for India and other similar developing countries. Such knowledge is vital to understanding the global aspect of the phenomenon.

### 7.2. Practitioner implications

The findings may be of interest to practitioners in the entrepreneurial space. If the identified issues are resolved, it will help create a conducive environment for the agri-tech entrepreneurship ecosystem growth. The COVID-19 situation enhances the need for quick and sustainable resolution of the hindrances and structural reforms in the ecosystem.

**Table 2.** Relative importance of issues in the growth of Indian agri-tech start-ups.

| S. No. | Top 5 issues | User opinions rank* | Percentage (out of 7) | CEO opinions rank** | Percentage (out of 17) |
|---|---|---|---|---|---|
| 1 | Availability of low-interest and creative models of financing | 1 | 60 | 1 | 53 |
| 2 | Infrastructure availability and resource support | 2 | 57 | 2 | 41 |
| 3 | Cross-domain quality solutions with multiple levels of features | 3 | 54 | 3 | 41 |
| 4 | A supportive mindset of business customers and partners | 4 | 53 | 4 | 35 |
| 5 | Skill-building and training of farmers | 5 | 51 | 5 | 18 |

*Based on the average of 202 opinions in the survey.
**Based on Frequency of Note out of seventeen.

The monitoring and evaluation of the Agritech start-up ecosystem are currently working in various forms/levels of detail. Still, they are uncoordinated at several policy institutes in India, e.g. ICAR, NASSCOM, and fourteen other industry associations. The Agritech activities are happening in start-ups, large private and public companies, incubators, accelerators, universities, research institutes, NGOs, and scientific bodies. The monitoring, evaluation, and support coordination may be realised in non-bureaucratic practice by the programme management arm of a neutral unit/public organisation. e.g. the National Institute of Agricultural Extension Management (Manage) or ICAR in coordination with the united nations' food and agricultural organization (FAO).

### *7.3. Managerial implications*

The suggestions for improvement in the agri-tech start-up ecosystem for the Agri-tech entrepreneurs based on experts' discussions are as follows:

*Availability of low-interest and creative financing models:* Due to ecosystem limitations, funding terms may be eased as the gestation cycle and payback period in an agri-tech start-up are slightly stretched compared to other sectors. Banks and financing institutions must develop favourable terms and creative models for funding. The funding is vital since agri-tech start-up end-users have low creditworthiness; thus, they have to depend on credit from the agri-tech companies or their partners rather than banks or other financial institutions. The pandemic has also negatively impacted the cash availability in the value chain. Low-interest financing and easy-term funding are also necessitated since agri-tech start-ups may soon shift their focus from large farmers to medium and marginal farmers with small landholdings and limited payment capacity linked to the value chain.

*Incubation Availability and Infrastructure Support*: India has approximately 435 incubators (97% in the academia or public sector) and eighty-five accelerators (89% in the private sector). Among accelerators, only 6% have a programme for agri-tech (NASSCOM 2020). The number of accelerators in the public sector, incubators in the private sector, and agri-tech-focused programmes in the accelerators need to increase significantly to manage the geographical scale and size. Standalone incubators must look beyond their limited capability and coordinate with other incubators, corporate social responsibility departments, and government institutions for grants and support infrastructure. The government may support filling the infrastructure gaps on the Internet, technology (cloud-based data centre, application program interfaces (API) store, digital stack e-platform for Agri-tech companies), and public-private partnerships incubators. The open standard-based digital cloud stack (identity, digital assets, secure datasets, payment, finance, and insurance) and farm-wise databases of core parameters are the two underlying assets on which Agriculture 4.0 needs to operate to minimise the crisis created by a COVID-19-like situation in the future.

*Cross-Domain Quality Solutions with Multiple Levels of Features*: Agri-tech start-ups' technology solutions need to sync their offerings to the current agricultural supply chain realities. Agri-tech start-ups need to work with a structural entity such as Farmer Producers Companies (FPCs)/ Farmer Producers Organizations (FPOs), Agricultural Marketing Cooperative Societies (AMCS), and Primary Agricultural Cooperative Societies (PACS). To have the required Farmers customer base, the business model may include flexibility, e.g. software solution bundling with the leased/rental equipment/hardware may increase the affordability of the innovative solutions.

The offerings and solution versions need to be simpler, e.g. vernacular language content with pictorial representation, integrated platforms, credit linkages, and affordable price points for small farmers.

The COVID-19 lockdown crisis provides both opportunities and challenges. The COVID-19 pandemic situation has forced the government to expedite structural reforms. Access to the open database of core agriculture parameters and national digital stack may help the agri-tech solution scale up to the smart system and eventually connect to the system of systems.

*Supportive Mind-set of Business Customers and Partners*: Medium and large private sector companies need to change their mindset and adopt solutions from a start-up. Large corporations with access to substantial land clusters directly or indirectly in India or abroad may become involved. The large Indian companies allocated only 0.3% of their USD 10 billion Corporate social responsibility (CSR) fund to Agritech from 2014 to 2019 (NASSCOM 2019). Improved allocation of CSR funds may promote agri-tech solutions among small farmers. The government needs to reorient its generic support and upscale its focused support for agri-tech setups in rural India. Local support from community opinion-makers may also induce the mass adoption of solutions during the trial phase.

*Skill Building and Training of Farmers*: The start-up's executives voiced the need to develop additional training programmes to educate farmers and adapt to mobility and new technological advancements. The COVID-19 pandemic has reduced farmers' resistance to adopting Agri-tech digital applications (Mitter 2020).

The user opinion survey also suggested the need for ecosystem development by spreading awareness at the village level, training farmers on a large scale, and increasing knowledge of benefits. Adopting digital Agri-tech solutions by farmer producer organizations (FPOs) could be a game-changer for start-ups. Agri clinics and Kisan Vigan Kendra under public-private partnership may be set up near the farmer community while including recent technologies, e.g. e-commerce, data analytics, and artificial intelligence, in their portfolio. Agri universities' training and skill-set capacity building, curriculum enhancement in relevant technologies such as digital usage competency, artificial intelligence, drone technology, Internet of Things (IoT), data analytics, web development, and modern farm machine maintenance may help support user-and employee-based start-ups.

## 8. Conclusion

The silent digital transformation is progressively transforming agriculture. At the same time, Agri-tech start-ups are grappling with the issues highlighted in this study. The issues that may be addressed on priority include the availability of low-interest financing, infrastructure improvement, incubation support, cross-domain quality solutions, a supportive mindset of business customers and partners, and skill-building and training of farmers/users/FPOs.

The Agritech growth trend may not be visible in 2022 due to the economic fallout of the pandemic. However, investments may rebound in subsequent years and provide an opportunity for careful investors. Now is the need for the government to support the Agri-tech start-ups in the pandemic situation with working capital, funding, compliance, and fiscal policy support, as it is doing for other micro and small enterprises. The other potential initiatives that support the fledgling ecosystem are the open-source digital platform for Agri-tech/Agriculture 4.0, an open database of core agricultural parameters, and low-cost cloud hosting. Recently, large technology companies, such as Bayer, Cargill, Monsanto, Microsoft, and IBM, have also rolled out Agri-tech solutions. It is prudent for start-ups to make their solutions interoperable with technology majors and sync with global platforms. Investors may shift their focus from China to India in the post-pandemic world food marketplace. Thus, start-ups shall show a successful track record concerning the investors' preferred technologies. The integrated pre-harvest and post-harvest e-commerce platform, artificial intelligence-based solutions, biotech, and advanced farm analytics may define the agriculture sector's next growth phase.

Entrepreneurs need to note that not only technologies but also new business models, such as 'Farming as a service' (FAAS), micro franchising, and traceability using blockchain, are of investor interest due to the sustainability and improvement aspects in the 'Farm to Fork' supply chain. Entrepreneurs may even start looking at international growth opportunities in the post-pandemic economy as competitors may still struggle with aftereffects.

The digitalisation wave, led by agri-tech start-ups with their innovative business model, recent technologies, and easy access to marketing, can better transform and lead India to agriculture 4.0.

## Note

1. Twenty-two alphabets 'A', 'B' … . 'V', are used to refer to the twenty-two hidden names of the start-ups, Institutes, and Universities.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Notes on contributor

***Sanjay Chaudhary*** is an Associate Professor of MIS and Operations Management. He is also a Research Scholar at Delhi Technological University. Before seven years in Academics, he has seventeen years of international work experience in the corporate sector. An IIT (India), SUNY (USA), and AIT (Thailand) alumnus have expertise in ICT, Corporate, education, and Training. He is awarded a fellowship by the IETE and a Chartered Engineer designation by the Institution of Engineers, India. He is a member of the All-India Management Association (AIMA). He has published several academic journal articles and a book.

***P. K. Suri*** serves as a Professor at the Delhi School of Management (DSM), Delhi Technological University. His areas of interest for teaching, research and consultancy include E-Governance, IT Strategy and Management, Project Management and Business Analytics. He has served as Head (DSM) during 2013–2016. He is originally from the 1986 batch of the Indian Statistical Service. He has also served in consulting roles for the UN bodies, viz. WTO and FAO. He is a member of the Governing Council of the Global Institute of Flexible Systems Management (GIFT). He has published numerous journal articles.

## ORCID

*Sanjay Chaudhary* http://orcid.org/0000-0002-7380-8487

## References

AgFunder. 2020. *Farm Tech Investment Up 370% in 6 Years. How Will Covid-19 Impact 2020 Trends? 2020*. Farmtech Investment Report. April 30. https://agfundernews.com/farm-tech-investment-up-370-in-6-years-how-will-covid-19-impact-2020-trends.html.

Anupam, Anand, and R. Saravanan. 2019. *Agri-Tech Start-Ups: The Ray of Hope in Indian Agriculture*. Discussion Paper 10. MANAGE-Centre for Agricultural Extension Innovations. https://www.manage.gov.in/publications/discussion%20papers/MANAGE-Discussion%20Paper-10.pdf.

Arafat, Mohd Yasir, Imran Saleem, Amit Kumar Dwivedi, and Adil Khan. 2020. "Determinants of Agricultural Entrepreneurship: A GEM Data-Based Study." *International Entrepreneurship and Management Journal* 16 (1): 345–370.

Astalin, Prashant Kumar. 2013. "Qualitative Research Designs: A Conceptual Framework." *International Journal of Social Science and Interdisciplinary Research* 2 (1): 118–125.

Bairwa, S. L., K. Lakra, S. Kushwaha, L. K. Meena, and P. Kumar. 2014. "Agripreneurship Development as a Tool to the Upliftment of Agriculture." *International Journal of Scientific and Research Publications* 4 (3): 1–4.

Bose, S C, and R Kiran. 2014. "Identification of Success Factors for Business Incubation in Agribusiness for Achieving Higher Productivity." *Productivity* 55 (1): 64–69.

Boston Consulting Group. 2020. "Signs of Rebound Forecast a New Era for Agriculture." Accessed March 8, 2021. https://www.bcg.com/publications/2020/agricultural-industry-to-bounce-back-post-covid-19.aspx.

Chaudhary, Sanjay, and P. K. Suri. 2021. "Modelling the Enablers of E-Trading Adoption in Agricultural Marketing: A TISM Based Analysis of eNAM." *Vision-the Journal of Business Perspective*. doi:10.1177/0972262920977979.

Ciruela-Lorenzo, A. M., A. R. Del-Aguila-Obra, A. Padilla-Meléndez, and J. J. Plaza-Angulo. 2020. "Digitalization of Agri-Cooperatives in the Smart Agriculture Context. Proposal of a Digital Diagnosis Tool." *Sustainability* 12: 1325–1332.

Crawford, Elizabeth. 2020. *Ag-Tech Could Hold the Key to Attracting a New Generation of US Farmers, Returning the Industry to Growth*. William Reed Business Media. Accessed January 2, 2021. https://www.foodnavigator-usa.com/Article/2020/01/02/Ag-tech-could-hold-the-key-to-attracting-a-new-generation-of-US-farmers.

Creswell, J. W. 2007. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. Thousand Oaks, California: Sage.

Creswell, J. W. 2009. *Research Design. Research Design: Qualitative, Quantitative and Mixed Approaches*. Thousand Oaks, California: Sage.

D'Cunha, Suparna Du. 2018. "For India's Farmers, It is Agtech Start-Ups, Not the Government, That is Key." Accessed July 11, 2019. https://www.forbes.com/sites/suparnadutt/2018/01/08/for-indias-farmers-its-agtech-start-ups-not-government-that-is-key/#724d2a3c1c6e.

Desai, B., E. D'Souza, J. W. Mellor, V. P. Sharma, and P. Tamboli. 2011. "Agricultural Policy Strategy, Instruments and Implementation: A Review and the Road Ahead"." *Economic and Political Weekly* 46 (53): 42–50.

Ernst and Young. 2020. "Agritech – Toward Transforming Indian Agriculture." Accessed August 10, 2021. https://assets.ey.com/content/dam/ey-sites/ey-com/en_in/topics/start-ups/2020/09/ey-agritech-towards-transforming-indian-agriculture.pdf?download.

Federation of Indian Chambers of Commerce and Industry. 2018. *Agri Start-Ups: Innovation for Boosting the Future of Agriculture in India*. FICCI-PWC Report. Accessed November 8, 2019. https://ficci.in/spdocument/23049/Agri-start-ups-Knowledge-report-ficci.pdf.

Fintech news. 2020. "6 Agri-Fintech Start-Ups in Asia to Follow in 2020." Accessed February 5, 2020. https://fintechnews.sg/36287/fintech/6-agri-fintech-start-ups-in-asia-to-follow-in-2020/.

Food and Agriculture Organization. 2017. "E-Agriculture Strategy Guide." Accessed June 6, 2020. https://www.fao.org/publications.

Food and Agriculture Organization. 2020. *Digital Innovation for Promoting Agriculture 4.0 in the Near East and North Africa*. Food and Agriculture Organization of the United Nations. Accessed February 18, 2021. http://www.fao.org/3/nd262en/nd262en.pdf.

Global System for Mobile Communications Association. 2020. "The GSMA Agri-Tech Toolkit for the Digitisation of Agricultural Value Chains. GSMA Agri-Tech Programme." https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2020/03/GSMA_Agri-tech_Toolkit_for_the_Digitisation_of_Agricultural_Value_Chains1.pdf.

Godhara, A., P. K. Sharma, and R. L. Godhara. 2020. "The Notion of Start-Up India Initiative Policy Framework: Insights for Start-Up to Scale-Up." *Journal of Economic Policy and Research* 15 (1): 16–28.

Grant, M., and W. Zhang. 2019. "Entrepreneurs and Start-Ups in the Agricultural Industry." *Journal of ASFMRA* 1: 33–44.

India Brand Equity Foundation. 2018. *India – Pushing the Right Levers: Game Changers for the Indian Economy in the Coming*. IBEF. Accessed October 14, 2020. https://www.ibef.org/India-Pushing-the-right-levers/index.html.

Kalaari. 2018. *Agtech Investment Focus*. Kalaari Capital Report. Accessed July 16, 2019. https://www.kalaari.com/wp-content/uploads/2018/12/Kalaari-AgTech-Report-2018-1.pdf.

Kundu, T., and M. C. Joshi. 2014. "An Entrepreneurship-Based Business Model for Providing Better Benefits to Indian Farmers." *Asia Pacific Journal of Management and Entrepreneurship Research* 3 (1): 70–83.

Lele, U., and S. Goswami. 2017. "The Fourth Industrial Revolution, Agricultural and Rural Innovation, and Implications for Public Policy and Investments: A Case of India." *Agricultural Economics* 48: 87–100.

Lincoln, Y. S., and E. G. Guba. 1985. *Naturalistic Inquiry*. Newbury Park, California: Sage.

MarketsandMarkets. 2020. "COVID-19 Impact on Digital Agriculture Market." Accessed June 20, 2021. https://www.marketsandmarkets.com/Market-Reports/covid-19-impact-on-digital-agriculture-market-222616344.html.

Marshall, Bryan, Peter Cardon, Amit Poddar, and Renee Fontenot. 2013. "Does Sample Size Matter in Qualitative Research?: A Review of Qualitative Interviews in is Research." *Journal of Computer Information Systems* 54 (1): 11–22.

McKinsey. 2014. *India's Technology Opportunity: Transforming Work, Empowering People*. McKinsey Global Institute Report. Accessed November 26, 2019. https://tinyurl.com/y5amrj7v.

Ministry of Agriculture and Farmer Welfare. 2018. *Promoting Start-Ups in Agriculture*. The Government of India. Accessed June 21, 2020. http://nmoop.gov.in/conference/docs/Background_Paper_Agri_Start-ups.pdf.

Mitter, Sohini. 2020. "How the Agri-Tech Platform Gramophone Increased Crop Yield for 5L Farmers in MP." Your Story. Accessed June 25, 2020. https://yourstory.com/2020/06/startup-bharat-Agri-tech-gramophone-increased-yield-farmers.

Moni, M., and Sharma S. 2018. *Situational Analysis on the Needs of Farming Community – ICT Enabled Services*. Report of the Committee on Doubling Farmers' Income. Delhi: Department of Agriculture, Cooperation and Farmers' Welfare. http://www.indiaenvironmentportal.org.in/files/file/DFI%20Vol-12A.pdf.

Mor, R. S., Sarbjit Singh, Arvind Bhardwaj, and L. P. Singh. 2015. "Technological Implications of Supply Chain Practices in the Agri-Food Sector: A Review." *International Journal* 2 (2): 720–748.

Morse, J. M. 2000. "Determining Sample Size." *Qualitative Health Research* 10 (1): 3–5.

National Association of Software and Service Companies. 2018. *Maximizing India Farm Output*. Nasscom Report. https://www.nasscom.in/knowledge-center/publications/Agri-tech-india-maxing-india-farm-output.

National Association of Software and Service Companies. 2019. *Agri-Tech in India: Emerging Trends in 2019*. Nasscom Report. https://www.nasscom.in/knowledge-center/publications/Agri-tech-india-emerging-trends-2019.

National Association of Software and Service Companies. 2020. *Reviving the Indian Tech Start-Up Engine During Covid19*. Nasscom Report. https://www.nasscom.in/system/files/secure-pdf/Reviving-the-Indian-Tech-Startup-Engine-During-COVID19_vF.pdf.

National Institute of Agricultural Extension Management. 2020. *Agricultural Extension and Support Systems in India: An Agricultural Innovation Systems (AIS) Perspective*. Discussion Paper 20. https://www.manage.gov.in/publications/discussion%20papers/MANAGE_Discussion%20Paper%2020.pdf.

Reddy, P. Madhu Kumar, and A. Rama Kumar. 2020. "A Study on Fintech in the Indian Agricultural Sector." *Journal of Critical Reviews* 7 (4): 605–607.

Sachitanand, Rahul. 2018. "For India's Agri-Tech Start-Ups, the Wind of Change is Finally Here." *The Economic Times*, July 3. https://economictimes.indiatimes.com/small-biz/start-ups/features/for-indias-agri-tech-start-ups-the-wind-of-change-is-finally-here/articleshow/64714325.

Sarangi, Susil Kumar. 2018. "Digital Farming: The Beacon of New Age Agriculture for Feeding Billions in India." *IOSR Journal of Business and Management* 20 (4): 54–63.

Sarangi, Santosh. 2020. "Blockchain, AI, IoT: How India Can Help Farmers by Leveraging These Technologies." *Financial Express*, May 26. https://www.financialexpress.com/opinion/blockchain-ai-iot-how-india-can-help-farmers-by-leveraging-these-technologies/19700409/.

Seth, Ankur, and Kavery Ganguly. 2017. "Digital Technologies are Transforming Indian Agriculture." *The Global Innovation Index*, 105–111. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2017-chapter5.pdf.

Shah, Jitesh. 2020. *Agri-Tech Trends You Should Look Out for in* 2020. Inc. 42. Accessed April 10, 2021. https://inc42.com/resources/india-Agri-tech-trends-you-should-look-out-for-in-2020/.

Sharma, K, and H P Mathur. 2019. "A Step towards Digital Agronomy by Start-ups." *International Journal of Business Insights and Transformation* 12 (1): 26–31.

Singh, Amarendra Pratap. 2013. "Factors Influencing Entrepreneurship Among the Farming Community in Uttar Pradesh." *Indian Journal of Commerce and Management Studies* 4 (3): 13–20.

Subhash, S. P., K. Srinivas, M. P. Samuel, and R. K. Sastry. 2016. "Evolution of Agribusiness Incubation Ecosystem in NARES for Promoting Agri-Entrepreneurship." *Indian Journal of Agricultural Economics* 71 (3): 235–251.

Subramaniam, M. V. 2020. *Problems Faced by Agri Start-Ups in India – I*. Pulse. Accessed March 14, 2021. https://www.linkedin.com/pulse/problems-faced-agri-start-ups-india-i-m-v-subramanian.

ThinkAg. 2020. "Ag-Tech in India- Investment Landscape Report 2020." Accessed April 2, 2021. https://a0e20efd-4fe9-4537-80f5-0bc1b751db89.usrfiles.com/ugd/a0e20e_da949f63da464c90a0ab4d25c4b280d5.pdf.

Tohidyan Far, S., and K. Rezaei-Moghaddam. 2019. "Multifunctional Agriculture: An Approach for Entrepreneurship Development of the Agricultural Sector." *Journal of Global Entrepreneurship Research* 9: 1–24.

World Government Summit. 2018. "Agriculture 4.0: The Future of Farming Technology." Accessed May 3, 2021. https://www.worldgovernmentsummit.org/api/publications/document?id=95df8ac4-e97c-6578-b2f8-ff0000a7ddb6.

Environmental Earth Sciences (2022) 81:353
https://doi.org/10.1007/s12665-022-10477-8

ORIGINAL ARTICLE

# An assessment of historical and future coastal dynamic response along the Odisha coast

**Badal Mohanty**[1] · **Raju Sarkar**[1]

Received: 30 December 2021 / Accepted: 5 June 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

Due to the overuse of the coastal zones, the effect of climatic crises, and the coastal erosion problem the coastal zones across the globe are at great risk. To understand the severity of the coastal erosion by keeping the usability aspect this study has considered the whole Odisha coast as the study area and has studied the effect of coastal erosion as a statistical analysis to evaluate the Net Shoreline Movement (NSM), End Point Rate (EPR), Linear Regression Rate (LRR) with the help of Digital Shoreline Analysis Tool (DSAS) at the blocks level for all the 22 coastal blocks of the 6 coastal districts. The area that has altered in the past has been determined to better comprehend the impact of the erosion process in the state. Furthermore, the future trend for the coastline position for 2030 and 2040 has been forecasted which showed the estuary positions are going to be affected most. Considering the Landsat satellite data to manually delineate the shoreline position from 1973 to 2020, the analysis showed average NSM, EPR, and LRR values of -84.95 m, -1.81 m/year, and -0.36 m/year respectively with 72.47 sq. km of the eroded coastal area against 42.83 sq. km of newly formed landmass by the coastal dynamic process. The study showed that mitigation and preventive measures are of utmost importance for the coast. The construction of different fixed coastal protection structures or some sustainable approach should be taken to prevent this disaster.

**Keywords** Accretion · Coastal dynamic action · Coasts · Erosion · Odisha

## Introduction

Coastal regions have always been an essential part of human life. It is used as a major trading route across countries, as a zone to construct monumental structures, or as recreational places. For some nations, it is the prime source of income. It is a general trend to observe coastal areas to be more populated than other areas worldwide. On a global scale, 20% of the land can be categorized as the coastal area which is inhabited by 41% of the global population as of the year 2003. The distribution of population is not uniform but rather depends on the weather condition, availability of opportunities for economic growth, etc. Also, among the 33 megacities with a population greater than 8 million, 21 are present in and within a distance of 100 km from the seashore (Martínez et al. 2007). Along with the adverse effect of climatic changes, the increase in these numbers is creating pressure on the coastal zones and changing the coastal dynamic actions (Rizzo and Anfuso 2020). The coastal dynamic action implies the process of erosion of landmass and the deposition or accretion of the same as a cyclic process of sediment circulation. But, the global trends are suggesting the depletion of the accretion process which is causing severe erosion (Rizzo and Anfuso 2020). A study (Mentaschi et al. 2018) has concluded that, among different causes of erosion, the human-induced factors are the major driving factor. The developments of dams and hydroelectric plants are halting the transportation of sediments and the cutting of coastal mangrove resources is causing the coastal landmass to lose its strength against the effect of waves. Another contributing factor to coastal erosion is global warming or more specifically the rise of the sea level (Zhang e al. 2004). Due to these stated causes, about 28,000 sq. km of land have been eroded while the figure for gained land is 50% less than the loss, i.e., 14,000 sq. km from the year 1984 to 2015 (Mentaschi et al. 2018). As reported by the IPCC 6th Assessment Report (Kemper et al. 2021), by

✉ Raju Sarkar
rajusarkar@dce.ac.in

Badal Mohanty
badalmohanty809@gmail.com

[1] Department of Civil Engineering, Delhi Technological University, Bawana Road, Delhi, India

Published online: 28 June 2022

the year 2150 following the SSP1 pathway the sea level will rise by $0.59 \pm 0.28$ m whereas by following the SSP5 pathway it will be $1.32 \pm 0.56$ m. SSP stands for shared socioeconomic pathways which describe different scenarios or changes that may happen in the future. SSP1 represents the most sustainable way of development and SSP5 represents the maximum use of fossil fuels causing different climate crises (O'Neill et al. 2014). With the sea-level change, the process of coastal erosion is bound to increase.

Over the years, the shorelines get changed due to the combined consequences of coastal erosion and accretion. By observing the shoreline positions of the past years, the estimation and severity of this dynamic action can be quantified (Mentaschi et al. 2018). With the help of GIS technology and multi-temporal satellite data, much research has been carried out. Studies have proven the efficacy of statistical measures including linear regression which can be applied to the coastline position to find the rate of the erosion process (Maiti and Bhattacharya 2009; Jana et al. 2014). Specifically, in the context of coastal erosion study, the digital shoreline analysis tool or DSAS (Himmelstoss et al. 2018) has been used extensively across the globe. For the Bay of Bengal, this tool has been applied to the coast of Indian states like West Bengal (Mondal et al. 2020), Odisha (Barik et al. 2021; Jana et al. 2014), Andhra Pradesh (Baig et al. 2020), Tamil Nadu (Velsamy et al. 2020; Natarajan et al. 2021) as well as the coast of Bangladesh (Ghosh et al. 2015; Hossain et al. 2021).

The research typically follows two broad steps, first is to extract the position of the shoreline using the satellite images. Images from satellites such as the Landsat series have been used to derive the shoreline. Automatic shoreline extraction by band rationing (Ghosh et al. 2015) and image edge detection (Hossain et al. 2021) as well as manual digitization by image interpretation (Barik et al. 2021; Jana et al. 2014) have been implemented for the task. Then the second step is to find the shoreline erosion rate. Statistics like linear regression rate or end point rate have been derived by the DSAS tool to quantify the changes. Further, some studies have tried to predict the future shoreline position using Gaussian process (Kupilik et al. 2018), the advanced machine learning approach (Yin et al. 2021), and the DSAS tool also (Natarajan et al. 2021).

These studies mainly focus on a relatively smaller part of the coastal length and generally do not cover any state entirely (Velsamy et al. 2020; Natarajan et al. 2021). Also, these study usually focuses on the district as the smallest individual unit (Barik et al. 2021) while neglecting the variation in the dynamic effect at even smaller administrative units. This is especially the case for the state of Odisha which is the study area of this research work where the preventive and protective measures taken against the horrific effect of this dynamic action by sea are minimal as compared to its magnitude. Many studies have been conducted to address this issue but failed at its usability as they have concentrated on a small region for their investigation and have rarely considered areas below the district level to make their analysis. This study has tried to address these problems by considering the entire coast of Odisha state of India. Further, the projection of the coastline and the areas likely to show the most change have been identified for the first time for the region.

This paper, in the following sections, gives an overview of the study area and then further goes up to describe the adopted methodology and the derived results with the concluding remarks on the effort made for the study along with its findings.

## Study area

The taken study area in this research is the coastal region of Odisha (Fig. 1). Odisha is an eastern state of India, situated on the coast of the Bay of Bengal. Out of 30 districts of the state, 6 coastal districts, namely Baleshwar, Bhadrak, Kendrapara, Jagatsinghpur, Puri, and Ganjam make up the 588-km-long coastline of the state. At the block level, a total of 22 blocks with an area of 8189 square kilometers of these 6 districts are home to 3.61 million people. This population translates to a density of 441 people in square kilometers of land in the study area which is 64% higher than the population density of 270 for the state (The 2011 census report) (Fig. 2).

Rajanagar block of Kendrapara district has the longest coast, Krushnaprasad block of Puri district has the maximum land area, and Chandbali block of Kendrapara district has the maximum population. The details about the coastal length, area, and population (as per the 2011 census report) have been presented in Table 1. There are three ports in the state which are all in the study area. One port is managed by the national government at Paradeep of Jagatsinghpur district, the other two at Gopalpur of Ganjam district and Dhamra of Bhadrak district are under state government jurisdiction. Six major rivers form estuaries in the coastal districts namely, Budhabalanga and Subarnarekha river at Baleswar, Brahmani and Baitarani rivers at Bhadrak, Mahanadi river at Jagatsinghpur, and the Rushikulya river at Ganjam. The Chilika Lake which is the largest brackish lake in Asia is also present in the Puri district. Further, all the district of the study area comes under the cyclone-prone region (Iwasaki 2016).

## Data used

To detect the shoreline position for the analysis the Landsat series images have been used (Table 2). Since the lunch of the Landsat 1 satellite in 1972, the Landsat

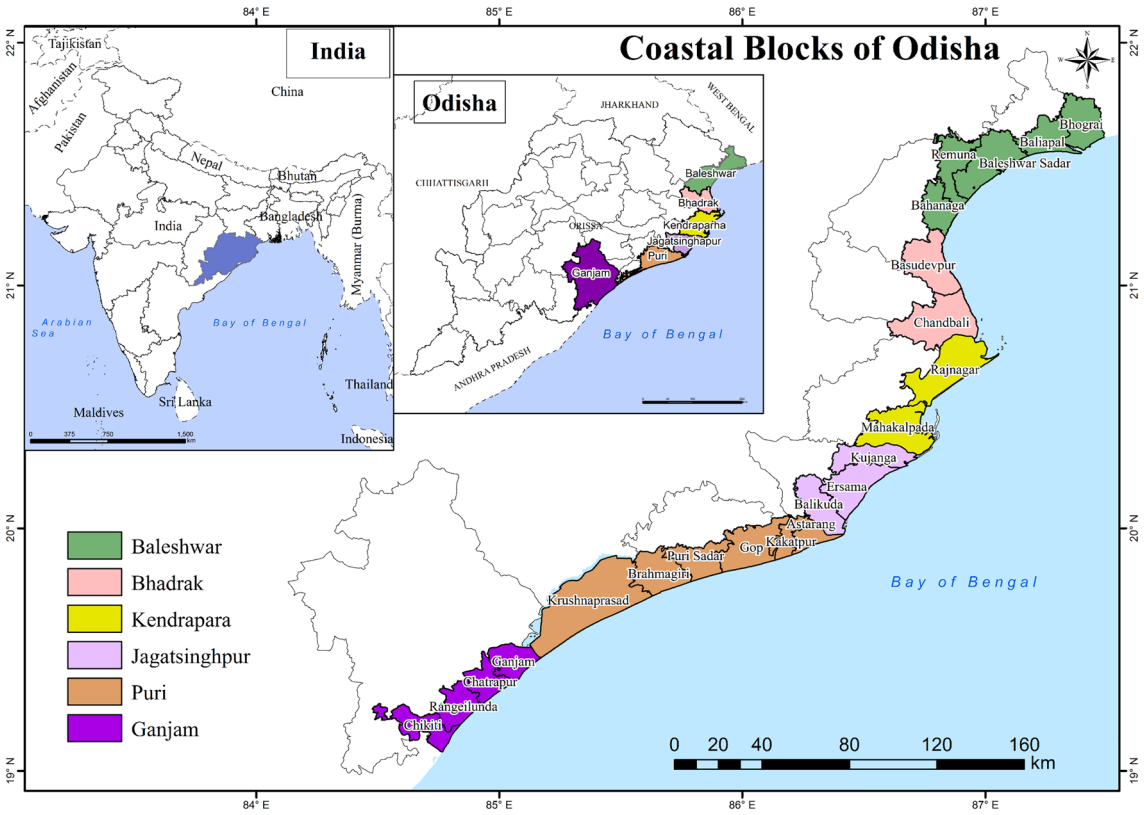



**Fig. 1** Study area

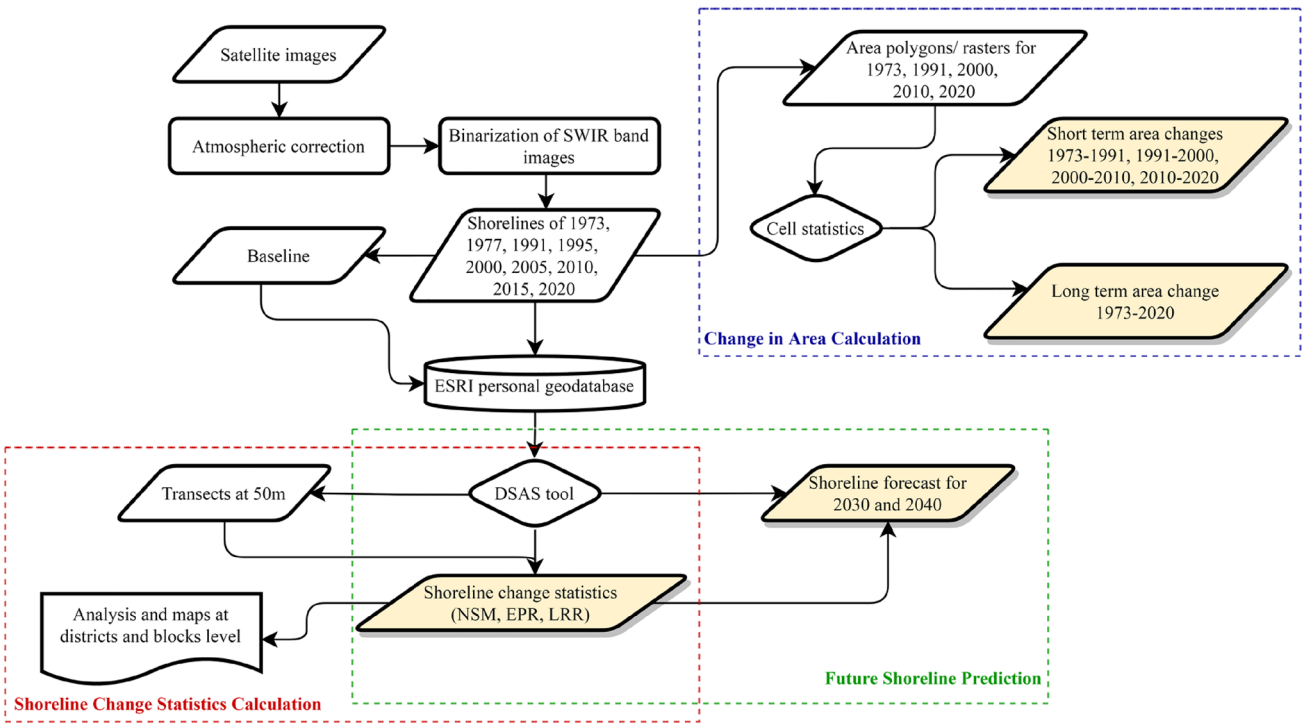



**Fig. 2** Methodology of the work

**Table 1** Details of the study area

| S No. | Block | Coast length (km) | Area (sq. km) | Population (as per the 2011 census report) |
|---|---|---|---|---|
| Baleshwar district | | | | |
| 1 | Baleshwar Sadar | 44.17 | 471.95 | 247,047 |
| 2 | Baliapal | 35.82 | 275.40 | 197,259 |
| 3 | Bhograi | 21.80 | 332.37 | 283,586 |
| 4 | Remuna | 68.06 | 315.49 | 180,044 |
| 5 | Bahanaga | 17.38 | 250.51 | 138,369 |
| Bhadrak district | | | | |
| 6 | Basudevpur | 36.89 | 427.28 | 219,108 |
| 7 | Chandbali | 28.41 | 586.52 | 250,037 |
| Kendrapara district | | | | |
| 8 | Rajnagar | 69.83 | 559.45 | 170,110 |
| 9 | Mahakalpada | 60.95 | 480.53 | 212,463 |
| Jagatsinghpur district | | | | |
| 10 | Balikuda | 8.27 | 313.68 | 165,275 |
| 11 | Ersama | 35.25 | 406.91 | 146,273 |
| 12 | Kujanga | 17.95 | 304 | 176,065 |
| Puri district | | | | |
| 13 | Kakatpur | 7.67 | 163 | 107,406 |
| 14 | Astarang | 24.57 | 147.93 | 82,176 |
| 15 | Brahmagiri | 17.87 | 348.54 | 139,449 |
| 16 | Gop | 23.83 | 425.08 | 165,952 |
| 17 | Krushnaprasad | 56.06 | 1062.27 | 89,371 |
| 18 | Puri Sadar | 14.36 | 289.59 | 150,800 |
| Ganjam district | | | | |
| 19 | Chikiti | 14.75 | 295.43 | 104,572 |
| 20 | Chatrapur | 17.16 | 223.02 | 135,751 |
| 21 | Rangeilunda | 13.16 | 266.33 | 161,372 |
| 22 | Ganjam | 16.80 | 243.98 | 89,170 |

satellite program is providing continuous temporal coverage of the earth. The Landsat images have been used in studies related to aquatic science, surface water mapping, vegetation phenology, surface temperature, agriculture, forest monitoring, etc. (Wulder et al. 2019). Previous studies also prove the efficacy of Landsat satellite images with spatial resolution up to 30 m to be suitable for the coastal erosion analysis (Barik et al. 2021; Natarajan et al. 2021). Images corresponding to different paths/rows covering the entire study area from the year 1972 to 2020 have been taken keeping the gap of 5 years. This gap could not be maintained in the earlier years due to the absence of cloud-free images. Images of post-monsoon and pre-monsoon periods have been considered for the analysis.

**Table 2** Data used

| Year | Date | Path/ row | Satellite/ sensor | Spatial resolution |
|---|---|---|---|---|
| 1973 | 17th Jan | 149/46 | Landsat 1 MSS | 60 m |
| | 18th Jan | 150/46 | | |
| 1977 | 5th Jan | 149/45 | Landsat 2 MSS | |
| | | 149/46 | | |
| | 6th Jan | 150/46 | | |
| | | 150/47 | | |
| 1991 | 30th Apr | 139/45 | Landsat 5 TM | 30 m |
| | 20th Mar | 140/46 | | |
| | | 140/47 | | |
| | 29th Mar | 139/46 | | |
| 1995 | 19th Jan | 139/45 | | |
| | | 139/46 | | |
| | 26th Jan | 140/46 | | |
| | | 140/47 | | |
| 2000 | 10th Dec | 139/45 | Landsat 7 ETM + | |
| | | 139/46 | | |
| | 17th Dec | 140/46 | | |
| | | 140/47 | | |
| 2005 | 15th Feb | 139/45 | Landsat 5 TM | |
| | | 139/46 | | |
| | 22nd Feb | 140/46 | | |
| | | 140/47 | | |
| 2010 | 28th Jan | 139/46 | | |
| | 4th Feb | 140/46 | | |
| | | 140/47 | | |
| | 13th Feb | 139/45 | | |
| 2015 | 10th Jan | 139/45 | Landsat 8 OLI | |
| | | 139/46 | | |
| | 17th Jan | 140/46 | | |
| | | 140/47 | | |
| 2020 | 3rd Mar | 140/46 | | |
| | | 140/47 | | |
| | 28th Mar | 139/45 | | |
| | | 139/46 | | |

## Methodology

### Shoreline extraction

The absorption spectra of water in infra-red wavelength bands are very low as compared to visible wavelength bands. So, the near infrared (NIR) and shortwave infrared (SWIR) bands of the Landsat satellites are most useful for differentiating land from the sea. Further, the presence of sediments in water can cause the seawater to have slightly higher reflectance in the NIR band which can create problems in the differentiation process. But, turbid water generally does not

cause the same problem for the SWIR band (Wang 2007). So for the shoreline extraction task, SWIR band images have been chosen. Before this, the satellite images have been atmospherically corrected. The satellite images are subjected to some errors due to the absorption and scattering of particles in the atmosphere while the images were collected by the sensor. Atmospheric correction is a necessary task that is done before any analysis of the images (Nazeer et al. 2014). In this case, the atmospheric corrections were done using the semi-automated classification plugin of QGIS (Congedo 2021). Along with the SWIR band, other optical band images for each year were also corrected to facilitate a better understanding of the study area by creating color composites. After the correction, the image tiles were mosaicked to cover the entire study area. Then the SWIR band image was binaries by visual interpretation of the image to make the demarcation of the land and sea better. The shorelines for every considered year were digitized by the heads-up digitization method at a constant map scale of 1:100,000 in ArcMap 10.7. False-color composites with NIR, red, and green bands in RGB channels were created to help in better visualization of the process. To maintain the similar scale while digitizing the coast line, the Landsat 1 and 2 MSS images have been resampled from 60 to 30 m with cubic convolution resampling technique as done by Barik et al. (2021).

## Shoreline change statistics calculation in DSAS tool

The digital shoreline analysis system (DSAS) tool is free software that works as an extension to ArcGIS or ArcMap and helps in calculating the change statistics from the time series data of the shoreline positions (Hossain et al. 2021). This tool required two inputs, one is the shoreline line feature layer which has to contain all the shorelines as the feature vector. The second one is the baseline. These two have to be stored in an ESRI personal geodatabase file. The shorelines change rates are calculated at the transect intersections along the shoreline. The transects are the lines drawn at a regular user-defined interval on the baseline and cross the shorelines. The attribute field of the shorelines contains the date information and the uncertainty value for each shoreline to facilitate the calculations of change statistics (Jana et al. 2014). For this analysis, the shoreline position for 2020 was buffered at some distance landward and reconstructed as the baseline. A total of 9645 transects were drawn at 50-m intervals with a default uncertainty of 10 m and smoothening distance of 2500 m. High smoothening distance has been taken to make the transects close to parallel to each other (Himmelstoss et al. 2018). Afterward, change statistics such as net shoreline movement (NSM), end point rate (EPR), and linear regression rate (LRR) have been calculated at each transect location.

The NSM depicts the separation between the youngest or newest coastline and the oldest. A positive NSM number indicates accretion, whereas a negative value indicates erosion. Considering $d_{2020}$ and $d_{1972}$ as the distances of the shorelines from the baseline in 2020 and 1972 respectively, NSM in meters is calculated as Eq. 1 (Natarajan et al. 2021),

$$NSM = d_{2020} - d_{1973}\# \quad (1)$$

Further, the EPR in m/year is calculated from the NSM as Eq. 2. EPR shows the rate of change of the shoreline and like NSM, the Positive value shows accretion and the negative value shows erosion (Natarajan et al. 2021).

$$EPR = \frac{\text{NSM}}{\text{number of years between the oldest and youngest shoreline}}\# \quad (2)$$

The LRR is calculated by considering the shoreline positions of all the years unlike that of NSM and EPR which are derived with only the youngest and the oldest shoreline positions. The regression rate is calculated with the ordinary least square (OLS) method (Himmelstoss et al. 2018). OLS finds a linear fit line by minimizing the residual sum of squares (Weisberg 2005). The slope of this line is represented by the LRR value.

The NSM, EPR, and LRR values were also categorized into five classes as shown in Table 3. The geographical coastline of the study area was split into smaller segments and all three categorized values (NSM, EPR, and LRR) were extracted to those lines at the blocks level.

## Future shoreline prediction

The DSAS tool provides a shoreline forecasting option for 10 to 20 years in the future. This method is an abstract form of the Kalman filtering which was derived as a linear filtering and prediction solution (Kalman 1960), moreover uses the concept of shoreline forecasting incorporating both long-term and short-term shoreline positions (Long and Plant 2012). This model requires at least four previous shoreline positions and calculate in a successive approach (Himmelstoss et al. 2018). Along with the future shoreline position, this tool also calculates an uncertainty layer for

**Table 3** Classification of DSAS statistics

| S.No | Class | NSM values (m) | EPR and LRR rates (m/yr) |
|---|---|---|---|
| 1 | High erosion | $\leq -100$ | $\leq -10$ |
| 2 | Moderate erosion | $> -100$ and $\leq -1$ | $> -10$ and $\leq -1$ |
| 3 | Almost no change | $> -1$ and $\leq 1$ | $> -1$ and $\leq 1$ |
| 4 | Moderate accretion | $> 1$ and $< 100$ | $> 1$ and $< 10$ |
| 5 | High accretion | $\geq 100$ | $\geq 10$ |

every forecasted year showing the reliability of the prediction at each transect location.

Using this process the shoreline positions for 2030 and 2040 have been derived. The quantitative measure for the movement of the shorelines in both the predicted years has been derived individually. For this, a feature layer with the shoreline of 2020 and a predicted year has been designed and using the DSAS tool, the statistics on NSM have been calculated with transects drawn at 500-m intervals along with other inputs as before.

### Change in area calculation

The rate of change in land area in sq. km/ year due to coastal erosion or accretion has been derived for all the blocks in the study area. This change has been calculated starting from 1973 as both short-term change considering years as 1973–1991, 1991–2000, 2000–2010, 2010–2020 and long-term change taking years as 1973–2020.

For the calculation, first, the area polygon for any particular year was created by considering the coastline of that year as the seaward boundary and the geographic separation of the block line as the landward boundary. This process is done for all the years. In different years, as the geographic separation line remains constant, the change in the coastal area would be only due to the change in the coastline. Then the polygons were converted to raster of constant cell size and cell values of 1 or 2. For example, considering the short-term area change between 1991 and 2000, the area raster was created with a cell value of 1 for 1991 and 2 for 2000. Further, these two rasters were combined with the sum operation of cell statistics. The process is shown in Fig. 3. The raster produced from this operation got three values, 1, 2 and 3 which implied erosion, accretion, and no change, respectively. Then the portion of the raster with cell values of 1 and 2 was converted to feature or polygon layer and separated at the block level to quantify the rate of changes in the area.

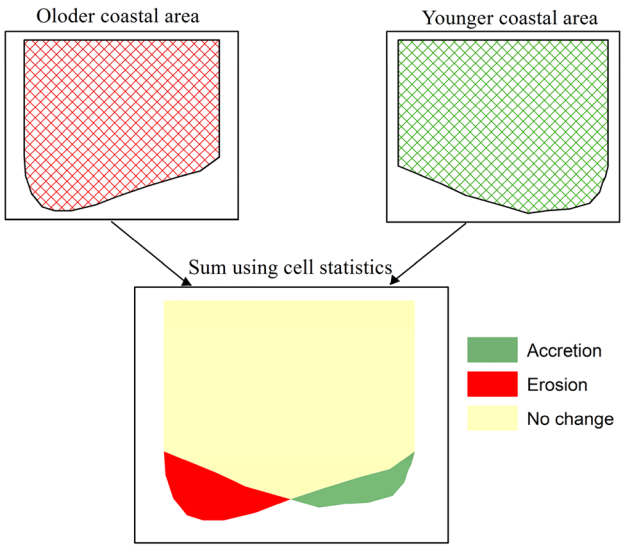



**Fig. 3** Change in area calculation process

## Results

Different calculations for the whole study area have been done by considering the six districts as separate units. Also, some statistics have been observed at the block level. The aforementioned approach intends to better quantify the statistical findings and proper visualization of the same.

### Shoreline change statistics

On the entire coastline of Odisha, among the 9645 transects, 5866 (60.82%) and 5549 (39.18%) transects registered land retreat and deposition, respectively. The average change in shoreline position or mean NSM is − 84.95 m with a maximum negative change of − 3508.4 m and a maximum positive change of 4042.9 m. The calculated EPR and LRR showed corresponding average values of − 1.81 m/year and − 0.36 m/year. The transects have the maximum EPR values for erosion as − 74.45 m/year with an average of − 5.34 m/year and accretion as 85.79 m/year with an average of 3.67 m/year. Also, the maximum erosion rate is − 36.54 m/year with an average rate of − 4.74 m/year and the maximum accretion rate is 72.83 m/year with an average rate of 4.32 m/year for LRR.

1810 number of transects for the Baleshwar district showed depletion of coastal length by an average of − 53.6 m with the maximum retreat of − 2382.25 m and the maximum deposition of 1291.44 m. The average EPR value is − 1.14 m/year with maximum erosion and accretion rates of − 50.55 and 27.41, respectively. Similarly, the average LRR value for this district is 2.46 m/year with a rate of − 36.54 m/year and 25.95 m/year concerning the maximum erosion and accretion. The south side shoreline covering the Bahanaga, Remuna, and part of Baleshwar Sadar blocks showed higher erosion rates in all three measured statistics as compared to the north-east side shoreline.

For the Bhadrak district, the average NSM length is observed to be − 257.4 m from the 1089 transects. The maximum erosion length is − 1206.33 m and that of accretion is 1039.42 m. − 5.46 m/year of average EPR is measured along with the maximum erosion rate of − 25.6 m/year and maximum accretion rate of 22.06 m/year. The maximum erosion rate is − 14.52 m/year and the maximum accretion rate of 19.06 m/year for LRR with an average rate of 1.26 m/year. Along the coastal stretch of this district, the two sides show a high erosion rate in comparison with the middle part covering both blocks.

A total 1504 number of transects have been generated in the Kendrapara district which showed an average NSM value of − 224.99 m with a maximum retreat of − 3508.4 m and a maximum deposition of 4042.9 m. The average EPR is − 4.77 m/year for the districts in addition to the maximum erosion rate is − 74.45 m/year and the maximum accretion rate of 85.79 m/year. By considering the LRR, the maximum erosion rate and maximum accretion rate are found to be − 33.65 m/year and 72.83 m/year. The average LRR value is 4.33 m/year. Between the two blocks, almost the whole coastal length of the Rajnagar block is found to be under erosion along with the south part of the Mahakalpada block.

The average change in shoreline position is − 206.26 m for the Jagatsinghpur district calculated by 1168 numbers of transects. The maximum withdraws of − 1628.88 m and maximum accumulation of 662.29 m are observed in the area. The average EPR value is − 4.37 m/year and the average LRR value is − 4.19 m/year. The maximum erosion rate is − 34.57 m/year and the maximum accretion rate of 14.05 m/year for the EPR. Further, by taking LRR into account, the maximum erosion rate is − 33.65 m/year, and the maximum accretion rate of 18.31 m/year. The erosion rate is found to be higher on the southeast side of the Ersama block and the Balikuda block of the district.

For the Puri district, a total 2876 number of transects have been used to calculate the DSAS statistics. The average change in coastal line position is found to be − 32.09 m which is the lowest among all the districts on the negative side, i.e., for average shoreline retreat. As derived from the NSM, the maximum retreat is − 634.89 m and the maximum deposition is 262.24 m. Both EPR and LRR showed an average value of − 0.68 m/year and − 0.61 m/year, respectively. Also, from the EPR and LRR values, the maximum erosion rates are − 13.47 m/year and − 17.67 m/year. Further, the maximum accretion rates are 5.56 m/year and 10.46 m/year. Among the six blocks of the district, the Kakatpur, Gop, and parts of Krushnaprasad block fell on transects showing erosion.

A total of 1263 transects for the Ganjam district showed an average change in coastal distance of 71.92 m as NSM. This is the only district to show the average outward movement of the shoreline as derived from the NSM. The maximum shoreline retreat length is − 226.51 m and that for accretion is 695.9 m. 1.52 m/year of average EPR is measured along with the maximum erosion rate of − 4.81 m/year and maximum accretion rate of 14.77 m/year. The maximum erosion rate is − 4.02 m/year and the maximum accretion rate of 9.9 m/year for LRR with an average rate of 0.93 m/year. The southeast coastline of the Ganjam block and the middle part of the coastline of the Chatrapur block showed erosion.

The areas near the estuary of Subarnarekha river which is present towards the north-east side of the Baleshwar district and the estuary of Devi river which is present at the border of the Jagatsinghpur and the Puri district are found to be under erosion. Also, the area near the Dhamra port of the Bhadrak district and the northeast side of the Paradeep Port is under erosion condition.

The DSAS statistics were categorized into five classes based on their values and transferred to the geographical coastal line of the study area as described in the methodology section (Fig. 4). Also, the length of each class at the district level in percentage has been calculated to give an estimate of the severity of the erosion (Fig. 5).

Taking the class division as per the NSM, for the Baleshwar district about 37% of coastal length is under high erosion (less than − 100m) which is balanced by almost the same (40%) amount of high accretion (more than 100m) length. The EPR classes also show a similar trend for this district with 6.28% length under high erosion (− 10 m/year) and 6.89% under high accretion (10 m/year) unlike the class division by LRR, in which 3.27% length under high erosion (10 m/year) and 19.21% length under high accretion (10 m/year). In the Bhadrak district, about 60% of the coastline has been retreated by less than − 100m and 50% of the coastline shows an EPR rate in the moderate erosion class (between − 10 m/year and − 1 m/year). About 43–44% of the coastline of both the Kendrapara and Jagatsinghpur districts have retreated by less than −100m. For the Puri district, 52.88% of coastal length has witnessed moderate erosion (between -100 m/year and − 1 m/year) as per NSM values, and considering the LRR classes almost similar percentage of length shows no change (between − 1 m/year and 1 m/year). Ganjam is the only district for which the accretion and almost no change classes have higher values considering all three statistical measures in comparison with the erosion classes.

The contrast in the percentage lengths under different categories considering both EPR and LRR values can be attributed to the calculation considerations of both. Calculation of the EPR uses only the oldest and the youngest shoreline whereas, for the LRR, shorelines of all the years under investigation are taken into account. So in this study, EPR has been calculated from the shoreline of 1973 to 2020 however in the calculation of the LRR, the shorelines of all 9 years have been considered. Considering all three statistics and all six districts, around 50% of the shorelines of the Bhadrak, Kendrapara, and Jagatsinghpur (Barik et al. 2021) districts are either under high or moderate erosion categories. Both the Baleshwar (Jana et al. 2014) and Ganjam (Mishra et al. 2021) districts are showing more accretion.

## Shoreline prediction

Using the forecasting option of the DSAS tool, the shoreline positions of 2030 and 2040 have been identified. The shoreline movement (NSM) for 2030–2020 and 2040–2020 for the entire study area have been derived using the DSAS tool

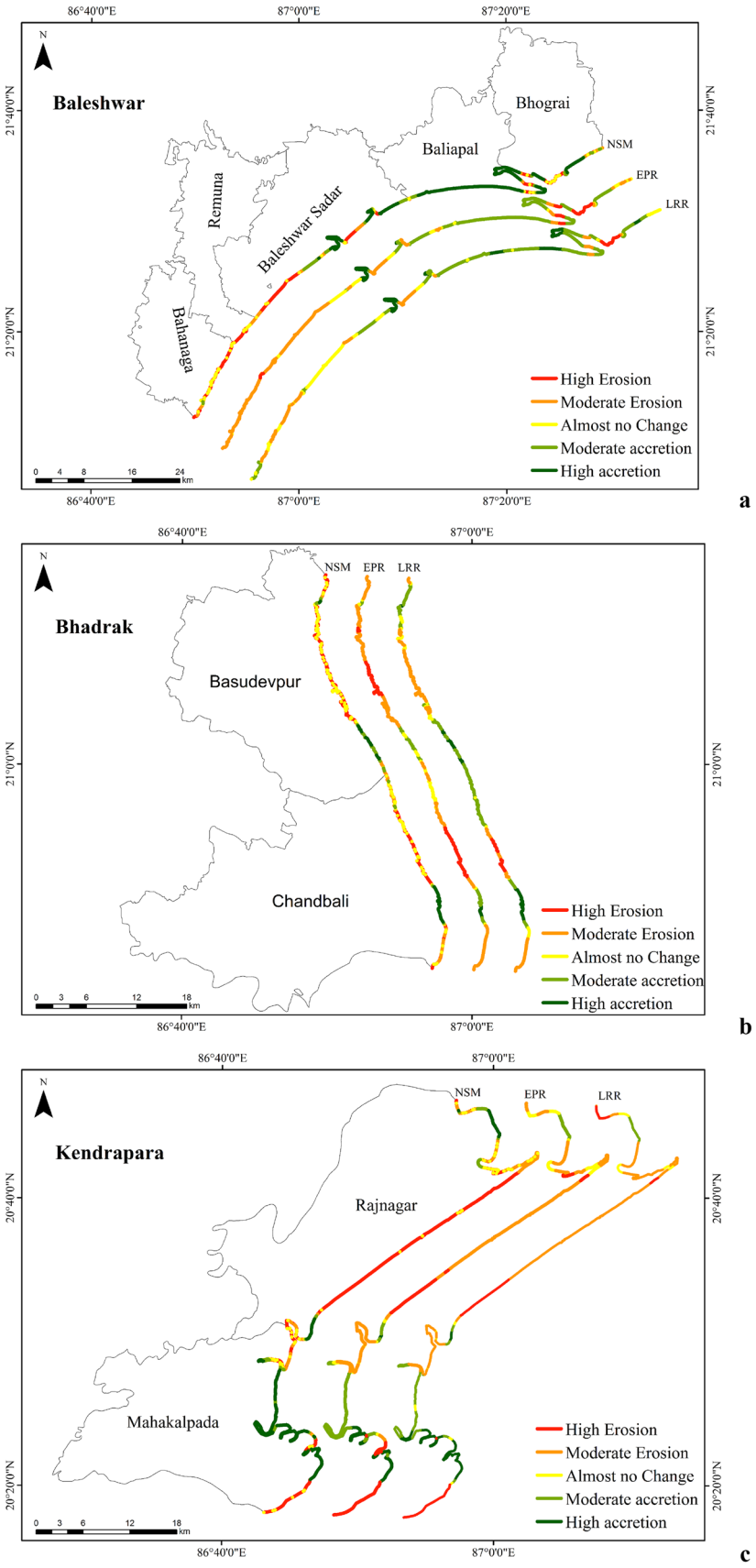
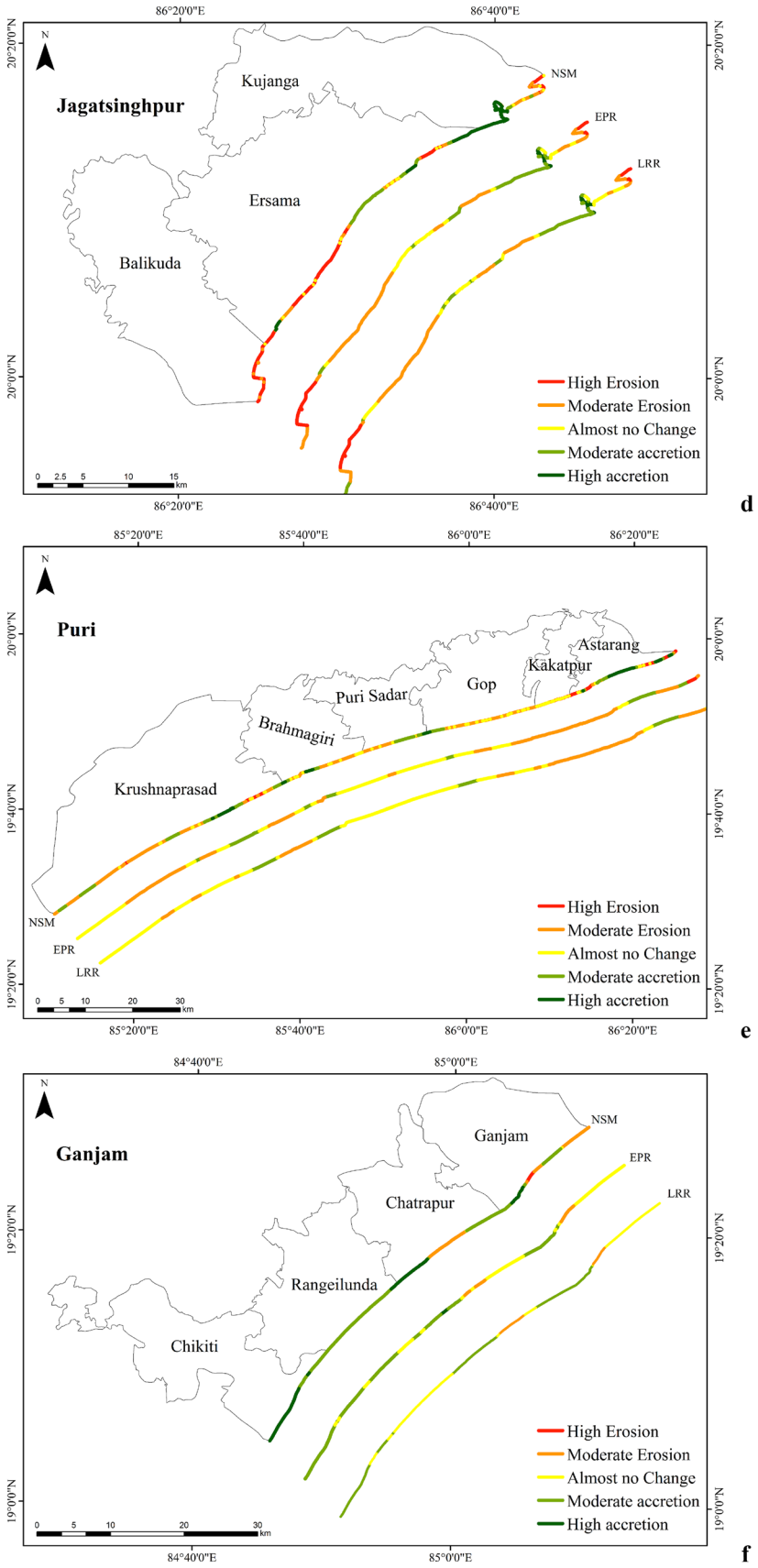



**Fig. 4** NSM, EPR, and LRR classified values of **a**. Baleshwar. NSM, EPR, and LRR classified values of **b** Bhadrak and **c** Kendrapara. NSM, EPR, and LRR classified values of **d**. Jagatsinghpur and **e**. Puri. NSM, EPR, and LRR classified values of **f**. Ganjam

**Fig. 4** (continued)



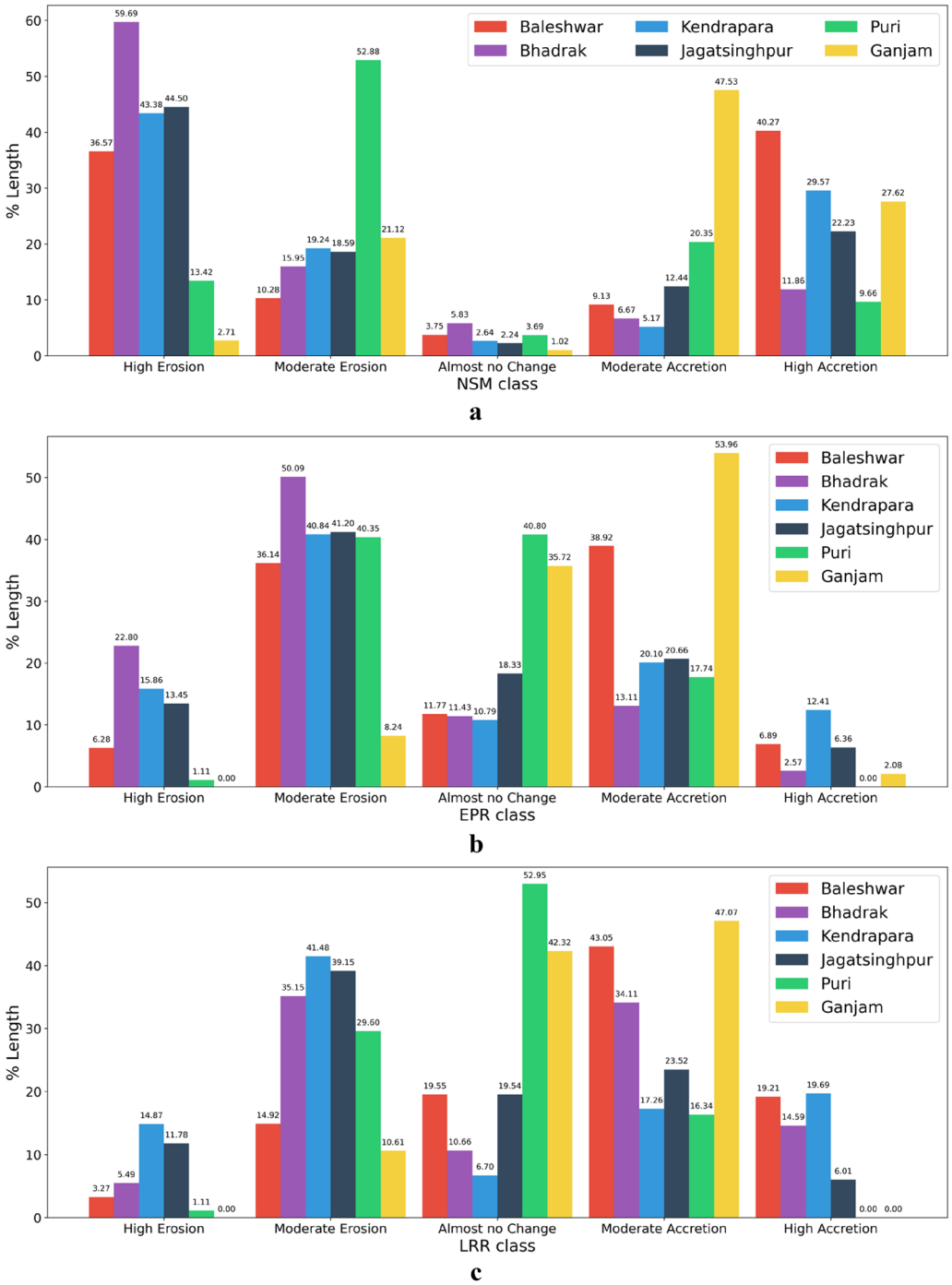



**Fig. 5** % coastline at district level under different classes of NSM, EPR, and LRR

with 885 numbers of transects drawn at 500 m intervals. For 2030, 41% of the transects showed retreat in the shoreline distance with an average change in distance of 22.8 m. The predicted maximum retreat is − 3534 m and the maximum deposition is 2421 m. By 2040, the maximum retreat and maximum deposition are forecasted to be − 3176 m and 2573 m, respectively. Also, the average change for this year is 17.9 m. The plot of the NSM values at different transects in Fig. 6 shows, that comparatively the change is low for the Ganjam and Puri districts.

Figure 7 shows the locations with high erosion or accretion as predicted by the tool along with their uncertainty level. Notably, the changes in shoreline positions are higher near the estuary positions. The estuary of Devi river

(Fig. 7a), the east side of Subarnarekha river (Fig. 7d), and the coastline near the Bhitarakanika national park or mangrove wetland of the Kendrapara district (Fig. 7c) are predicted to show the erosion of the sand deposits. The south-east side of Kendrapara district near the border of Jagatsinghpur district (Fig. 7b) will show erosion as well as the shift in sand deposit position which is predicted to cause accretion. The area towards the north side of the Bhitarakanika national park (Fig. 7c), as well as the west side of Subarnarekha river (Fig. 7d), are predicted to be under erosion in the coming 20 years.

## Change in area

Both the short-term and long-term area change rates due to both erosion and accretion processes in sq. km/year have been calculated at the block level and aggregated at the district level (Table 4). Considering short-term changes, the Baleshwar has the maximum rate of change in area for erosion in the year range of 1973–1991 (2.04 sq. km/year) and 2000–2010 (0.578 sq. km/year) but the same district is found to has the maximum rate of change in area for accretion in the year range of 1991–2000 (2.993 sq. km/year) and 2010–2020 (0.558 sq. km/year). For the year interval of 2010–2020, Puri district also shared the same area change due to accretion as with Baleshwar. For the earlier year interval of 1991–2000, Puri was found to have the maximum rate of change in the area considering erosion conditions (0.601 sq. km/year). For a consecutive year span of 2000–2010 and 2010–2020, Jagatsinghpur district showed the maximum area change due to accretion (1.059 sq. km/year) and erosion (1.017 sq. km/year), respectively. In long term, i.e., from 1973 to 2020, Kendrapara district has both maximum area change by erosion (0.516 sq. km/year) as well as accretion (0.415 sq. km/year).

Only for Ganjam district, the net change in area is positive, in other words, only this district has witnessed a gain in the landmass of 4.63 sq. km whereas the other five districts have lost land area (Fig. 8) between 1973 and 2020. The loss of land is maximum for Bhadrak district (10.69 sq. km) which is followed by Jagatsinghpur (8.78 sq. km) and Puri (5.12 sq. km). Both Baleshwar and Kendrapara districts have lost a similar amount of area (4.8 sq. km). For the entire study area and at large of the coastal area of Odisha, the

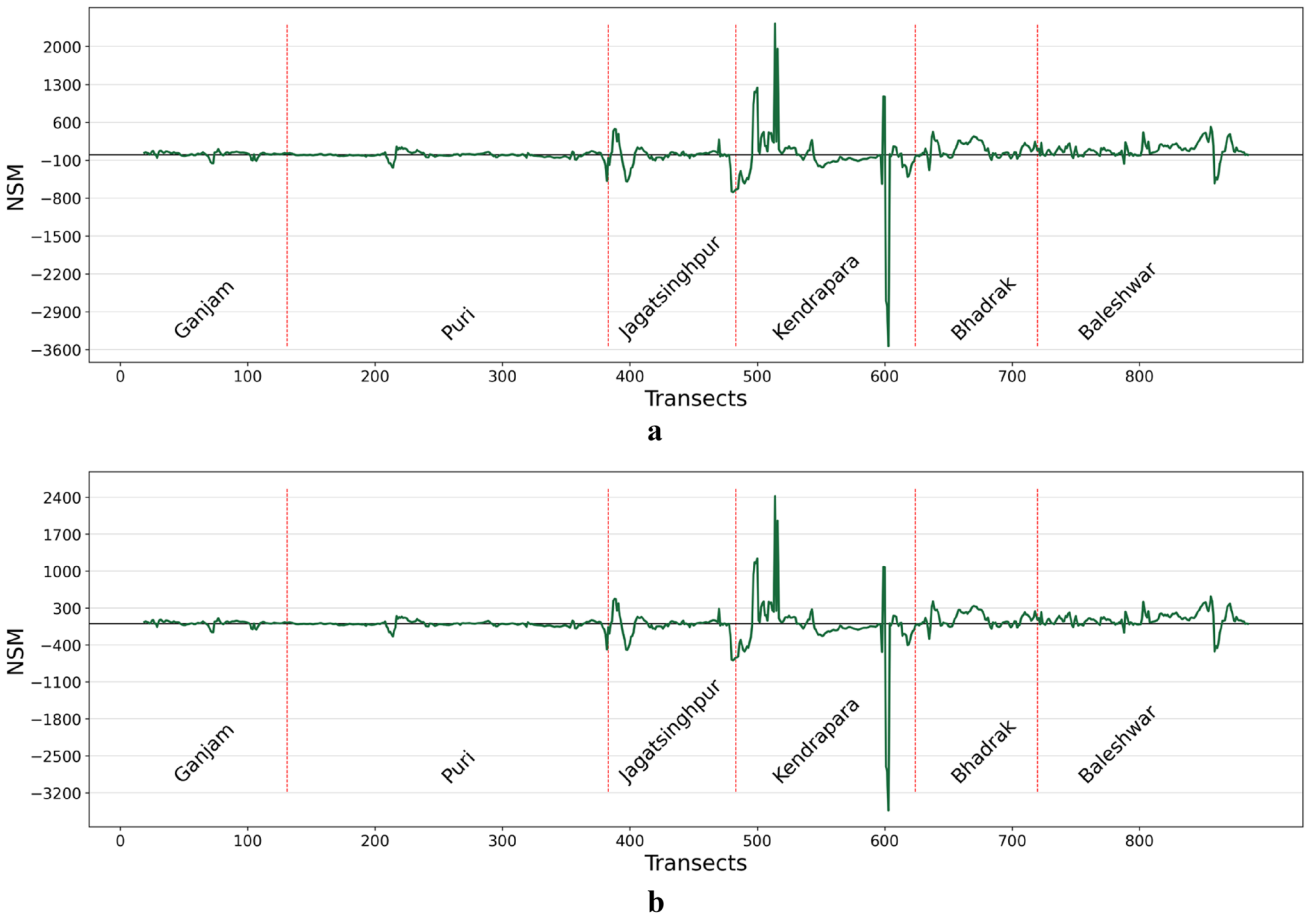



**Fig. 6** NSM of **a** 2030 and **b** 2040

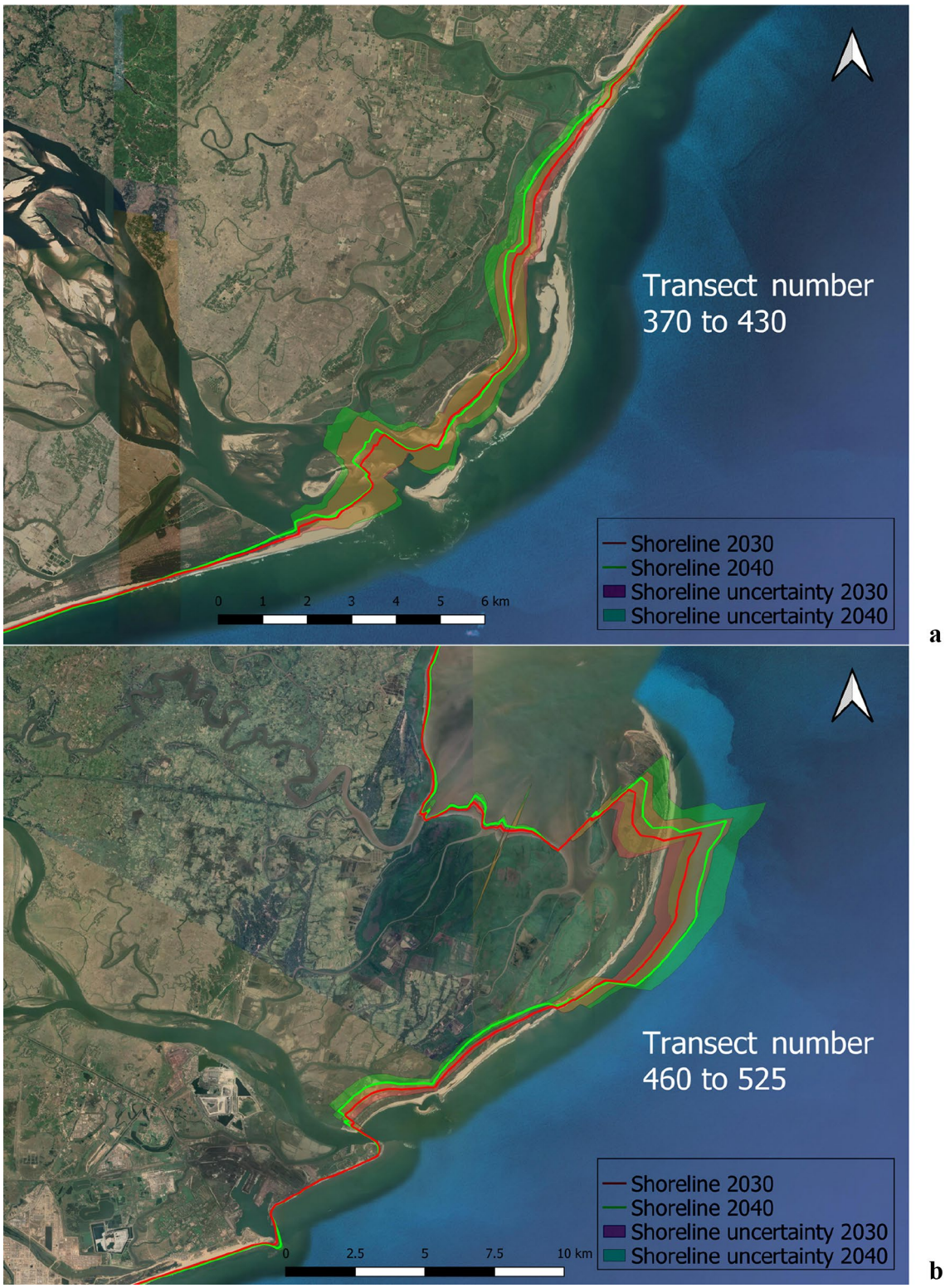



**Fig. 7** Predicted shoreline positions for 2030 and 2040 at **a**. estuary of Devi River and **b**. the border of Kendrapara district and Jagatsinghpur district. Predicted shoreline positions for 2030 and 2040 at **c**. coastline near the Bhitarakanika national park or mangrove wetland and **d**. estuary of Subarnarekha River

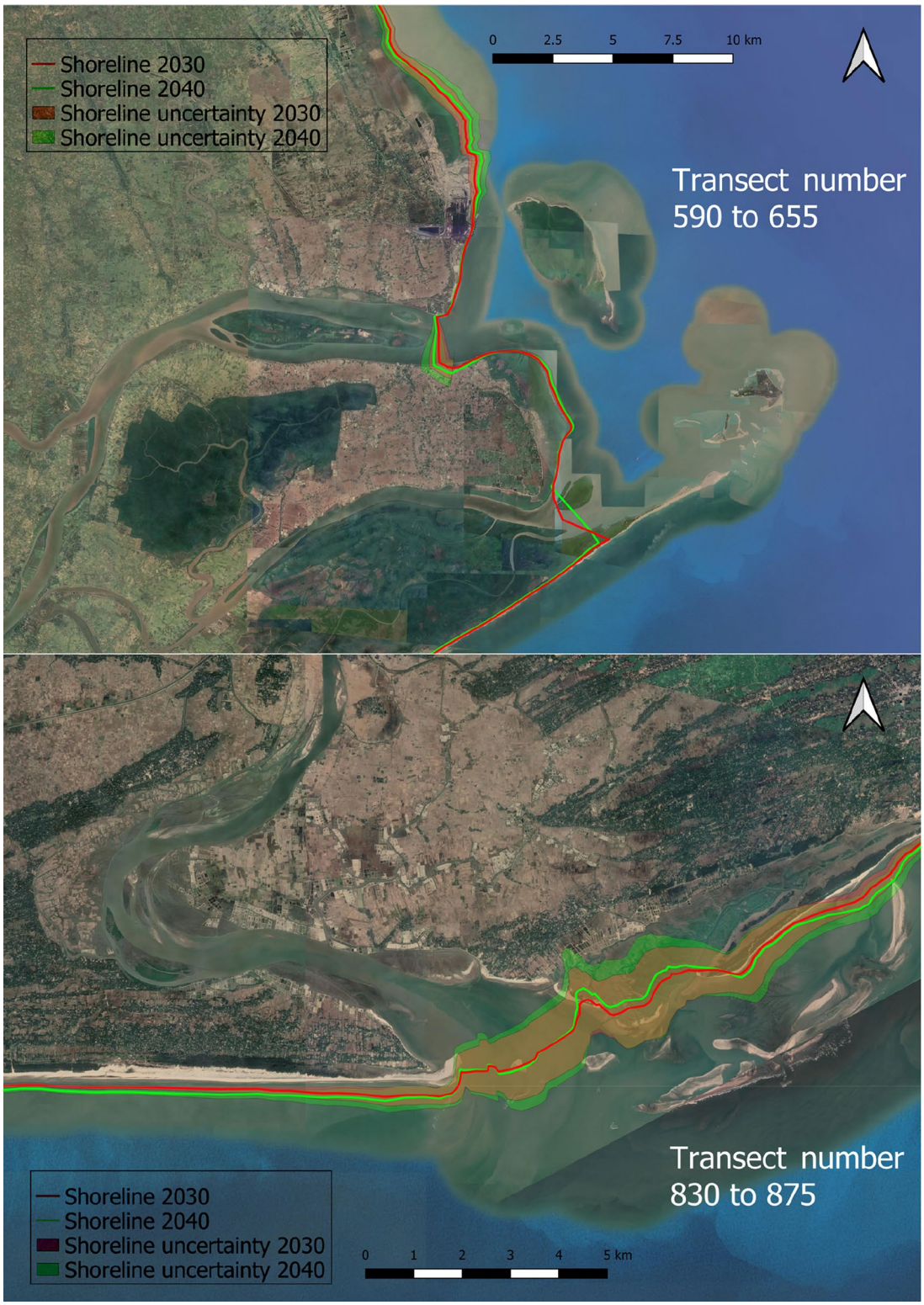



**Fig. 7** (continued)

eroded and gained land masses area amount to be 72.47 sq. km and 42.83 sq. km making the net area have been lost due to coastal erosion to be 29.64 sq. km. The Odisha coast has experienced about 60% more erosion than accretion which is higher than the 50% mark for the same at the global level (Mentaschi et al. 2018).

**Table 4** Rate of change in area for different districts

| Year range | Baleshwar | | Bhadrak | | Kendrapara | | Jagatsinghpur | | Puri | | Ganjam | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Er* | Ac# | Er | Ac | Er | Ac | Er | Ac | Er | Ac | Er | Ac |
| | Short-term change rate (sq. km/ year) | | | | | | | | | | | |
| 1973–1991 | 2.04 | 0.034 | 1.715 | 0 | 1.121 | 0.327 | 0.437 | 0.163 | 0.323 | 0.218 | 0.4 | 0.239 |
| 1991–2000 | 0.094 | 2.993 | 0.062 | 1.351 | 0.494 | 1.64 | 0.472 | 0.149 | 0.601 | 0.304 | 0.122 | 0.14 |
| 2000–2010 | 0.578 | 0.743 | 0.022 | 0.744 | 0.518 | 0.857 | 0.173 | 1.059 | 0.164 | 0.639 | 0.091 | 0.174 |
| 2010–2020 | 0.49 | 0.558 | 0.194 | 0.352 | 0.707 | 0.404 | 1.017 | 0.092 | 0.558 | 0.212 | 0.157 | 0.2 |
| | Long term change rate (sq. km/year) | | | | | | | | | | | |
| 1973–2020 | 0.317 | 0.212 | 0.283 | 0.055 | 0.516 | 0.415 | 0.242 | 0.055 | 0.167 | 0.058 | 0.018 | 0.117 |

*Erosion

#Accretion

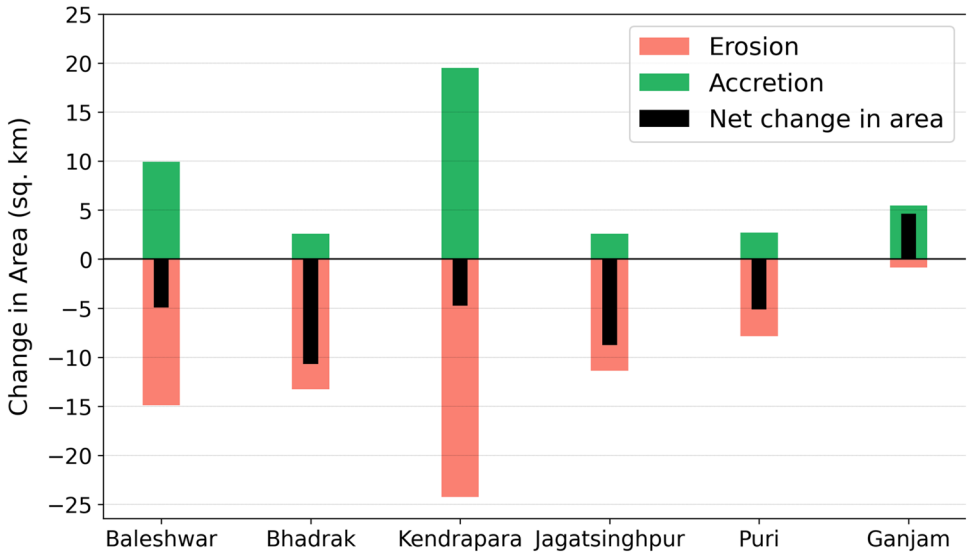



**Fig. 8** Change in area of different district from 1973 to 2020

At the block level (Table 5), the Rajnagar block of Kendrapara has lost 10.189 sq. km of land which is 34% of the entire land lost whereas the other block of the same district, Mahakalapada has gained 5.449 sq. km of land. Other blocks which have lost major land area are, the Bhograi block (− 4.418 sq. km) of Baleshwar district, Basudevpur (− 5.712 sq. km) and Chandbali (− 4.976 sq. km) blocks of Bhadrak district, and Balikuda (− 4.83), and the Ersma (− 3.609) blocks of Jagatsinghpur district. All four blocks of the Ganjam district have gained land by the coastal dynamic actions.

## Discussions

Considering the combined effect of coastal erosion and accretion process, the coastline gets changed due to the compound outcome of several natural or geographical as well as anthropogenic factors (Xu et al. 2016). The constant effect of wave energy which shapes the coast can affect it as the wave pattern or height increases or changes periodically. The effect of sea-level rise, storm surges, and tsunamis also influence the coastline position (Mujabar and Chandrasekar 2013). The geological features or properties of the coastal soil present can control the process of erosion as the sandy beach tends to be eroded more than that rocky or clayey beaches (Valvo et al. 2006). Similarly, various anthropogenic contributors such as spatial distribution of population, the developmental activities like the creation of roads or ports near the coast, construction of dams at rivers upstream, cutting down coastal vegetation, etc., are escalating the coastal erosion conditions (Xu et al. 2016). Analogous reasons for the change in coastal zones have also been found in various studies conducted on the Indian coastline. A study done for the coastline along south Gujarat (Misra and Balaji 2015) showed that changing land use and land cover patterns are major factors. Furthermore, for the eastern coastline of India, the cyclonic storm surge plays a major role in the erosion process (Mishra et al. 2019).

**Table 5** Change in area at block level from 1973 to 2020

| Block name | Area lost | Area gained | Net area changed |
|---|---|---|---|
| *Baleshwar district* | | | |
| Baleshwar Sadar | 3.138 | 3.353 | 0.215 |
| Baliapal | 0.602 | 5.304 | 4.702 |
| Bhograi | 5.694 | 1.276 | – 4.418 |
| Remuna | 1.048 | 0.013 | – 1.035 |
| Bahanaga | 4.414 | 0.004 | – 4.41 |
| *Bhadrak district* | | | |
| Basudevpur | 6.35 | 0.638 | – 5.712 |
| Chandbali | 6.929 | 1.953 | – 4.976 |
| *Kendrapara district* | | | |
| Rajnagar | 13.588 | 3.399 | – 10.189 |
| Mahakalpada | 10.657 | 16.106 | 5.449 |
| *Jagatsinghpur district* | | | |
| Balikuda | 5.019 | 0.189 | – 4.83 |
| Ersama | 4.709 | 1.1 | – 3.609 |
| Kujanga | 1.643 | 1.301 | – 0.342 |
| *Puri district* | | | |
| Kakatpur | 0.777 | 0 | – 0.777 |
| Astarang | 2.09 | 1.158 | – 0.932 |
| Brahmagiri | 0.502 | 0.318 | – 0.184 |
| Gop | 1.268 | 0.15 | – 1.118 |
| Krushnaprasad | 2.874 | 0.996 | – 1.878 |
| Puri Sadar | 0.333 | 0.097 | – 0.236 |
| *Ganjam district* | | | |
| Chikiti | 0.001 | 1.799 | 1.798 |
| Chatrapur | 0.307 | 2.072 | 1.765 |
| Rangeilunda | 0 | 0.852 | 0.852 |
| Ganjam | 0.539 | 0.754 | 0.215 |

The data show, that the prevention, as well as mitigation of the erosion process, is very important in the study area. There are numerous methods available and have been adopted in different parts of the world to mitigate this problem. Many fixed coastal protection structures such as seawalls made up of stone or concrete, groyne created with stone or wood or breakwater structures have beenused (Foti et al. 2020). Some coastal protection is present in a few parts of the study area. For instance, in the Siali beach of the Ersama block. But these brittle structures do not always withstand the constant dynamic force created by the waves (Foti et al. 2020). The breakwater structures (Fig. 9a) and the concrete sea walls (Fig. 9b) used at the Siali beach have failed due to the same. This shows that even if the same methods have been applied successfully to some other coastal areas, these failed to be translated in practice to this particular area. Also, hard or fixed protection like these creates erosion in the nearby areas where these protections are absent. Another possible solution includes more sustainable approaches such as the creation of artificial dunes, nourishment of eroded beaches, and restoration of vegetation near the coastal areas (Foti et al. 2020). Figure 9c shows that sand erosion has exposed the roots of the trees near the Siali beach. Decreasing the anthropogenic activities by different restrictions in constructions or excessive movement or activities by people (Foti et al. 2020) and/or adopting strategies to do effective coastal retreats (Haasnoot et al. 2021) may prove effective solutions in this case. Further, the river mouths in the study area are showing signs of excessive erosion, the causes of this need to be further analyzed and managed.

## Conclusion

The shoreline positions derived from the Landsat satellites from the year 1973–2020 have been analyzed with the help of the DSAS tool. The NSM, EPR, and LRR values calculated at the district level shows the Bhadrak, Kendrapara, and Jagatsinghpur district are in maximum danger due to erosion. The estuaries of Subarnarekha river and Devi river, also the areas near Paradeep and Dhamra port are some of the hotspots for erosion. The predicted future shoreline shows except for the Ganjam district; all five coastal districts will show the overall landward movement in the next 20 years. Specifically the estuaries will witness the maximum negative change according to the forecast. The whole state has lost 72.47 sq. km of land against the gain of 42.83 sq. km. The Rajnagar block of Kendrapara has lost a maximum of 10.189 sq. km of land area. This study concludes that the Odisha coast is under high erosion and showing signs of escalation in the near future. Also as evident by similar research, it is evident that various natural, as well as anthropogenic conditions, are causing the erosion. The constructed coastal protective structures have failed to fulfil their objective in some places. These statistics or the study can be taken as reference by policymakers or institutions in making a proper and robust work plan for the prevention as well as mitigation of this unfortunate disaster. Furthermore by a thorough analysis of the locations showing high erosions different protective structures or natural barriers such as plantation of coastal vegetation can be done to control and mitigate the erosion process.

**Fig. 9** **a** The breakwater structures, **b** the concrete sea walls, **c** at the Siali beach of Ersama block

**Funding** No funding received for this study.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

Baig MRI, Ahmad IA, Shahfahad T, M., & Rahman, A. (2020) Analysis of shoreline changes in Vishakhapatnam coastal tract of Andhra Pradesh, India: an application of digital shoreline analysis system (DSAS). Ann GIS 26(4):361–376. https://doi.org/10.1080/19475683.2020.1815839

Barik G, Guru B, Sangma F (2021) shoreline changes analysis and forecast using digital shoreline assessment system 50: evidences from parts of east coast of India. J Indian Soc Remote Sens 49(11):2815

Congedo L (2021) Semi-automatic classification plugin: a python tool for the download and processing of remote sensing images in QGIS. J Open Sour Softw. 6(64) 3172 https://doi.org/10.21105/joss.03172

Foti E, Musumeci RE, Stagnitti M (2020) Coastal defence techniques and climate change: a review. Rendiconti Lincei Scienze Fisiche e Naturali 31(1):123–138. https://doi.org/10.1007/s12210-020-00877-y

Fox-Kemper B, HT Hewitt, C Xiao, G Aðalgeirsdóttir, SS Drijfhout, TL Edwards NR Golledge, M Hemer, RE Kopp, G Krinner, A Mix, D Notz, S Nowicki, IS Nurhati, L Ruiz, J-B Sallée, ABA Slangen, Y Yu (2021), Ocean, cryosphere and sea level change. In: Climate change 2021: The Physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change [MassonDelmotte V, P Zhai, A Pirani, SL. Connors C Péan, S Berger, N Caud, Y Chen, L Goldfarb, MI Gomis, M Huang, K Leitzell, E Lonnoy, J BR Matthews, TK Maycock, T Waterfield, O Yelekçi, R Yu and B. Zhou (eds.)]. Cambridge University Press

Ghosh MK, Kumar L, Roy C (2015) Monitoring the coastline change of Hatiya Island in Bangladesh using remote sensing techniques. ISPRS J Photogramm Remote Sens 101:137–144. https://doi.org/10.1016/j.isprsjprs.2014.12.009

Haasnoot M, Lawrence J, Magnan AK (2021) Pathways to coastal retreat. Science 372(6548):1287–1290. https://doi.org/10.1126/science.abi6594

Himmelstoss EA, Henderson R E, Kratzmann MG, Farris A S (2018) Digital Shoreline Analysis System (DSAS) Version 5.0 User Guide(No. 2018-1179). US Geological Survey. https://doi.org/10.3133/ofr20211091

Hossain MS, Yasir M, Wang P, Ullah S, Jahan M, Hui S, Zhao Z (2021) Automatic shoreline extraction and change detection: a study on the southeast coast of Bangladesh. Mar Geol 441:106628. https://doi.org/10.1016/j.margeo.2021.106628

Iwasaki S (2016) Linking disaster management to livelihood security against tropical cyclones: a case study on Odisha state in India. International Journal of Disaster Risk Reduction 19:57–63. https://doi.org/10.1016/j.ijdrr.2016.08.019

Jana A, Biswas A, Maiti S, Bhattacharya AK (2014) Shoreline changes in response to sea level rise along Digha coast, Eastern India: an analytical approach of remote sensing, GIS and statistical techniques. J Coast Conserv 18(3):145–155. https://doi.org/10.1007/s11852-013-0297-5

Kalman RE (1960). A New Approach to Linear Filtering and Prediction Problems. https://doi.org/10.1115/1.3662552

Kupilik M, Witmer FD, MacLeod EA, Wang C, Ravens T (2018) Gaussian process regression for Arctic coastal erosion forecasting. IEEE Trans Geosci Remote Sens 57(3):1256–1264. https://doi.org/10.1109/TGRS.2018.2865429

Long JW, Plant NG (2012) Extended Kalman filter framework for forecasting shoreline evolution. Geophys Res Lett. https://doi.org/10.1029/2012GL052180

Maiti S, Bhattacharya AK (2009) Shoreline change analysis and its application to prediction: a remote sensing and statistics based approach. Mar Geol 257(1–4):11–23. https://doi.org/10.1016/j.margeo.2008.10.006

Martínez ML, Intralawan A, Vázquez G, Pérez-Maqueo O, Sutton P, Landgrave R (2007) The coasts of our world: ecological, economic and social I mportance. Ecol Econ 63(2–3):254–272. https://doi.org/10.1016/j.ecolecon.2006.10.022

Mentaschi L, Vousdoukas MI, Pekel JF, Voukouvalas E, Feyen L (2018) Global long-term observations of coastal erosion and accretion. Sci Rep 8(1):1–11. https://doi.org/10.1038/s41598-018-30904-w

Mishra M, Acharyya T, Chand P, Santos CAG, Kar D, Das PP, Nascimento TVMD (2021) Analyzing shoreline dynamicity and the associated socioecological risk along the Southern Odisha coast of India using remote sensing-based and statistical approaches. Geocarto Int. https://doi.org/10.1080/10106049.2021.1882005

Mishra M, Chand P, Pattnaik N, Kattel DB, Panda GK, Mohanti M, Mohanty T (2019) Response of long-to short-term changes of the Puri coastline of Odisha (India) to natural and anthropogenic factors: a remote sensing and statistical assessment. Environ Earth Sci 78(11):1–23. https://doi.org/10.1007/s12665-019-8336-7

Misra A, Balaji RA (2015) A study on the shoreline changes and Land-use/land-cover along the South Gujarat coastline. Procedia Eng 116:381–389. https://doi.org/10.1016/j.proeng.2015.08.311

Mondal I, Thakur S, Juliev M, Bandyopadhyay J, De TK (2020) Spatio-temporal modelling of shoreline migration in Sagar Island, West Bengal. India J Coast ConserV 24(4):1–20. https://doi.org/10.1007/s11852-020-00768-2

Mujabar PS, Chandrasekar N (2013) Shoreline change analysis along the coast between Kanyakumari and Tuticorin of India using remote sensing and GIS. Arab J Geosci 6(3):647–664. https://doi.org/10.1007/s12517-011-0394-4

Natarajan L, Sivagnanam N, Usha T, Chokkalingam L, Sundar S, Gowrappan M, Roy PD (2021) Shoreline changes over last five decades and predictions for 2030 and 2040: a case study from Cuddalore, southeast coast of India. Earth Sci Inf 14(3):1315–1325. https://doi.org/10.1007/s12145-021-00668-5

Nazeer M, Nichol JE, Yung YK (2014) Evaluation of atmospheric correction models and Landsat surface reflectance product in an urban coastal environment. Int J Remote Sens 35(16):6271–6291. https://doi.org/10.1080/01431161.2014.951742

O'Neill BC, Kriegler E, Riahi K, Ebi KL, Hallegatte S, Carter TR, van Vuuren DP (2014) A new scenario framework for climate change research: the concept of shared socioeconomic pathways. Clim Change 122(3):387–400. https://doi.org/10.1007/s10584-013-0905-2

Rizzo A, Anfuso G (2020) Coastal dynamic and evolution: case studies from different sites around the world. Water 12(10):2829. https://doi.org/10.3390/w12102829

Valvo L. M, Murray A. Brad, Ashton A (2006). How does underlying geology affect coastline change? An initial modeling investigation. Journal of Geophysical Research Earth Surface. https://doi.org/10.1029/2005JF000340

Velsamy S, Balasubramaniyan G, Swaminathan B, Kesavan D (2020) Multi-decadal shoreline change analysis in coast of Thiruchendur Taluk, Thoothukudi district, Tamil Nadu, India, using remote sensing and DSAS techniques. Arab J Geosci 13(17):1–12. https://doi.org/10.1007/s12517-020-05800-1

Wang M (2007) Remote sensing of the ocean contributions from ultraviolet to near-infrared using the shortwave infrared bands: simulations. Appl Opt 46(9):1535–1547. https://doi.org/10.1364/AO.46.001535

Weisberg, S. (2005). Applied linear regression. John Wiley and Sons. 528

Wulder MA, Loveland TR, Roy DP, Crawford CJ, Masek JG, Woodcock CE, Zhu Z (2019) Current status of Landsat program, science, and applications. Remote Sens Environ 225:127–147. https://doi.org/10.1016/j.rse.2019.02.015

Xu N, Gao Z, Ning J (2016) Analysis of the characteristics and causes of coastline variation in the Bohai Rim (1980–2010). Environ Earth Sci 75(8):1–11. https://doi.org/10.1007/s12665-016-5452-5

Yin C, Anh DT, Mai ST, Le A, Nguyen VH, Nguyen VC, Duong TQ (2021) Advanced machine learning techniques for predicting nha trang shorelines. IEEE Access 9:98132–98149. https://doi.org/10.1109/ACCESS.2021.3095339

Zhang K, Douglas BC, Leatherman SP (2004) Global warming and coastal erosion. Clim Change 64(1):41–58. https://doi.org/10.1023/B:CLIM.0000024690.32682.48

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

Chemical Science

ROYAL SOCIETY OF CHEMISTRY

REVIEW

View Article Online
View Journal

Check for updates

Cite this: DOI: 10.1039/d2sc03441g

# Applications of metal–organic framework-based bioelectrodes

Vidushi Aggarwal,[a] Shipra Solanki[ab] and Bansi D. Malhotra *[a]

Metal–organic frameworks (MOFs) are an emerging class of porous nanomaterials that have opened new research possibilities. The inherent characteristics of MOFs such as their large surface area, high porosity, tunable pore size, stability, facile synthetic strategies and catalytic nature have made them promising materials for enormous number of applications, including fuel storage, energy conversion, separation, and gas purification. Recently, their high potential as ideal platforms for biomolecule immobilization has been discovered. MOF-enzyme-based materials have attracted the attention of researchers from all fields with the expansion of MOFs development, paving way for the fabrication of bioelectrochemical devices with unique characteristics. MOFs-based bioelectrodes have steadily gained interest, wherein MOFs can be utilized for improved biomolecule immobilization, electrolyte membranes, fuel storage, biocatalysis and biosensing. Likewise, applications of MOFs in point-of-care diagnostics, including self-powered biosensors, are exponentially increasing. This paper reviews the current trends in the fabrication of MOFs-based bioelectrodes with emphasis on their applications in biosensors and biofuel cells.

Received 20th June 2022
Accepted 4th July 2022

DOI: 10.1039/d2sc03441g

rsc.li/chemical-science

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

## 1 Introduction

Molecule-based porous compounds have successfully found utility in diverse areas, including gas adsorption/storage, separation, catalysis, environmental remediation, energy, optoelectronics, and health. Metal–organic frameworks (MOFs), zeolites and covalent organic frameworks (COFs) are the three most common types of 3D porous covalent polymers (PCPs) reported

[a]Nanobioelectronics Laboratory, Department of Biotechnology, Delhi Technological University, Shahbad Daulatpur, Delhi 110042, India. E-mail: bansi.malhotra@gmail.com

[b]Department of Applied Chemistry, Delhi Technological University, Shahbad Daulatpur, Delhi 110042, India. E-mail: bansi.malhotra@gmail.com

*Vidushi Aggarwal has recently obtained her Bachelor of Technology (B. Tech) in Biotechnology from Delhi Technological University (DTU), India. She is a recipient of the IASc-INSA-NASI Summer Research Fellowship 2021 at the Institute of Bioinformatics (IOB), Bengaluru, India. Recently, she was shortlisted for sponsored Idea development in the Tata Steel MaterialNEXT 3.0 challenge for a research project based on nanomaterials for water remediation. She has been a finalist in the NASA International Space Apps Challenge 2020. Her main areas of interest include metal–organic frameworks, nanomaterials, biosensors and cancer therapeutic targeting. She is highly motivated to pursue a career in translational research with the aim of contributing to patient-centered care and innovative medicine.*

*Shipra Solanki received her PhD in Chemical Sciences from AcSIR, CSIR-National Physical Laboratory, in 2019. She is presently working as CSIR-RA at the Department of Applied Chemistry, Delhi Technological University, Delhi, on the development of self-powered biosensors for disease detection. She has published more than eleven research articles in international journals and a book chapter. Her main areas of interest include nanomaterials, thin films, electrochemistry and biosensors.*

© 2022 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

in the literature. Compared to zeolites, MOFs and COFs can be readily tuned in terms of their pore size, aperture shape and functionality.[1] COFs constitute versatile networks constructed using covalent bonds (B–O, C–C, C–H, C–N, *etc.*) between the organic linkers and MOFs. These can be formed *via* coordinate bonding between metal ions and organic ligands.[2] Both classes of compounds have offered nanoporous materials with high surface areas and diverse pore dimensions, topologies and chemical functionalities. We will only discuss MOFs-based applications in this review.

Generally, d-block transition elements with high coordination numbers such as Zn, Cu, Co and Fe are chosen as metal ions, while multi-dentate organic ligands with a rigid backbone such as benzene-1,4-dicarboxylic acid (BDC), benzene-1,3,5-tricarboxylic acid (BTC), naphthalene-1,4-dicarboxylic acid and 2-ethyl-1*H*-imidazole are used as linkers. The main attraction of MOFs is their well-defined and tunable pore sizes conferring large internal pore volumes, which serve as a hub for guest molecules like proteins, nucleic acids, and carbohydrates. More recently, the concept of secondary porosity in MOFs has been explored by engineering its surface chemistry in terms of particle size, shape and texture.[3] Another captivating feature of MOFs is their ultrahigh specific surface area (possibly up to 7000 $m^2$ $g^{-1}$), which is measured using Brunauer–Emmett–Teller (BET) analysis. Their high BET area makes MOFs promising candidates for application in hydrogen storage, carbon capture and drug delivery. Mechanical flexibility is another lucrative property of MOFs, which allows them to be sized and structured according to specific applications.[4] A myriad of MOFs have been used for applications in gas storage,[5] electrocatalysis,[6] biosensing,[7] drug delivery[8] and environmental remediation.[9]

In this sense, MOFs have emerged as suitable materials in the fabrication of electrodes due to their rigid and malleable structure with redox activity and low cost. They have been extensively explored as anodes, cathodes and electrolyte materials in electrochemical systems. The high porosity and surface area of MOFs-based electrodes are highly favorable for an electrocatalytic role in oxygen reduction reaction (ORR) with great control over structural and chemical properties. Recently, the application of MOFs-based bioelectrodes in electrochemical biosensors and biofuel cells has flourished. Their advantageous properties such as tunable structure, ultrahigh porosity along with stabilizing host–guest interactions adeptly fit the design criteria of bioelectrodes. The characteristic porous framework favors high enzyme loading with improved structural and functional stability. MOFs serve as an ideal microenvironment to protect an enzyme from denaturation and also reduce leaching through hydrophobic interactions and covalent linkages. MOFs pores act as mini-reaction cells, optimizing the mass and charge transfer in biocatalysis. MOFs-based bioelectrodes are suitable for high electrocatalytic current density and low charge resistance as well. Some typical examples include zeolitic imidazolate frameworks (ZIFs),[10] and porous coordination networks (PCNs),[11,12] Fe-based MOFs[13] Mn-based MOFs,[14] Co-based MOFs,[15] and Ni-based MOFs[16] as promising electrode materials in electrochemical biosensors and enzymatic biofuel cells (EBFCs).

In this paper, we discuss the application of MOFs and their derivatives and various components of EBFCs followed by brief applications in biosensors. Fig. 1 shows a summary of the different properties and applications of MOFs. We expect to promote knowledge transfer from the previous studies of MOFs as electrocatalysts, electrolyte membranes, candidates for biomolecule retention, and fuel storage. We present an overview and discuss the recent advances in the development of MOFs in bioelectrocatalytic applications to generate new thoughts.

*B. D. Malhotra received his PhD from the University of Delhi, Delhi, in 1980. He has published more than 330 papers in refereed international journals (citations: 25711; index: 85), has filed 11 patents (in India and overseas), and has co-authored a textbook on Nanomaterials for Biosensors: Fundamentals and Applications and Biosensors: Fundamentals and Applications. He is the recipient of the National Research Development Corporation Award 2005 for the invention of a 'Blood Glucose Biochemical Analyzer' and is a Fellow of the Indian National Science Academy, the National Academy of Sciences, India, and an Academician of the Asia Pacific Academy of Materials (APAM). His current research activities include biosensors, point-of-care diagnostics, nano-biomaterials, biofuel cells, ordered molecular assemblies, conducting polymers, Langmuir–Blodgett films, self-assembled monolayers, advanced functionalized nanosystems, hybrid nanosystems, nano-biotechnology, biomedical engineering, and biomolecular electronics. Dr Malhotra is a former DST-SERB (Govt. of India) Distinguished Fellow and a Chief Scientist, CSIR-National Physical Laboratory(NPL), Delhi, India. Since 1994, he has been exploring functional materials for biosensing applications.*

## 2 Synthesis and modification of MOFs for electrochemical applications

More than 90 000 MOFs have been reported worldwide while around 500 000 more are predicted to be synthesized. Common routes adopted for the synthesis of MOFs include solvothermal, hydrothermal, and solvent-free techniques,[17] as summarized in Fig. 2.[17] The conventional solvothermal synthesis is based on the dissolution of the metal precursor and organic ligand in a suitable solvent such as *N*,*N*-dimethylformamide (DMF) and ethanol while water is used in case of hydrothermal synthesis. The structure and properties of MOFs are carefully governed by a combination of metal ion, organic ligand, solvent as well as synthesis method. Some of the advantages of the different synthetic methods used in MOF synthesis are summarized in Table 1.


© 2022 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

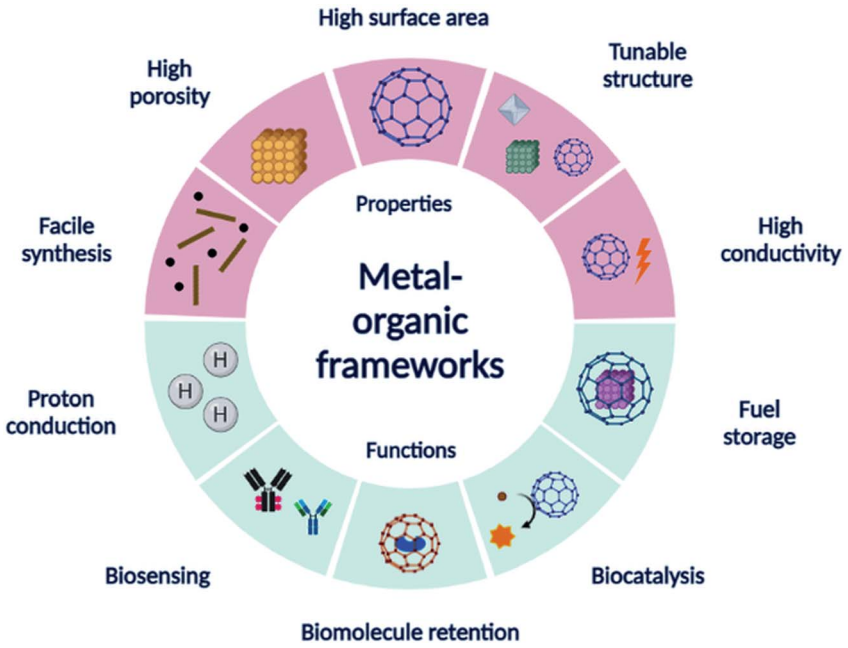



Fig. 1 Various properties and applications of MOFs.

While MOFs have been the front runners in gas storage, separation, and catalysis, applications harnessing their electronic capabilities have been recently realized. Since most MOFs do not have free charge carriers and have low-energy barriers, MOFs were believed to be inherently insulating, limiting their electrochemical applications. Despite having conductivity lower than $10^{-10}$ S cm$^{-1}$, which makes a material insulating in nature, MOFs can be tailored to achieve superior ionic and protonic conductivity by use of appropriate redox-inactive nodes and short linkers.[18] Many methods, including *in situ* and post synthetic modifications, have been adopted to improve their electrical conductivity, which include composite formation, use of functional linkers, doping,[19] and carbonization. Most methodologies include introducing functionality into MOFs *via* direct synthesis approaches. However, these may not be perhaps suitable sometimes because of challenges like insolubility of the linker, thermal and chemical stability, functional group compatibility, and interference between metal ions and linkers during the assembly of MOFs. To overcome these problems, post-synthetic approaches have been suggested. These can be *via* three modes, including post-synthetic modification (PSM), post-synthetic deprotection (PSD), and post-synthetic exchange (PSE).[20] Doping with nitrogen, sulfur, and phosphorus on MOF precursors is found to increase charge delocalization, catalytic centers, and electron transfer, thereby improving overall ORR performance.[21] Another approach commonly used is the carbonization of MOFs with great scope in electrochemical storage devices.[22] It has the advantage of no additional requirement of primary carbon source as it is fulfilled by the organic ligand. Carbonized MOFs demonstrate remarkable surface area and electrochemical performance with facile synthesis techniques. Different synthetic approaches adopted to synthesize MOFs for electrochemical applications are summarized in Fig. 3.

Precisely functionalized MOFs can accelerate the electrochemical processes as compared to CNT, graphene, or other nanomaterials based electrodes by also acting as active redox centres.[18] In MOFs, electrical conductivity depends on its metal-cation nature and coordination chemistry while internal ionic conductivity depends on ligand–ion interactions, hydrophilicity, functional groups, and interactions with guest molecules.[23] The nature of counter ions also elicits the electrochemical response of MOFs known as the ion-size effect. As observed in cyclic voltammetry (CV), smaller anions and cations provide stronger response through a series of faradaic reactions.[24] Electrochemical stability and reversibility are other important parameters to be considered in MOF-based systems. Stability is measured as the retained electrochemical response after repeated cycles, whereas reversibility is rapid electron transfer at the interface of electrode and electrolyte, both are crucial for good electrochemical performance. Charge transfer in MOFs can take place through three routes, namely in-plane $\pi$-conjugation,[25] through-bond[26] and through-space[27] charge transfer approaches. In-plane $\pi$-conjugated MOFs demonstrate high charge mobility and efficient charge delocalization that can be beneficial for many electrochemical applications. On the other hand, the through-bond approach increases electrical conductivity by the participation of delocalized electrons of metal/linker in charge conduction while through-space by $\pi$–$\pi$ interaction or electron hopping with guest molecules. These approaches allow better orbital overlapping and thus small band gaps, which results in superior electrical conductivity of MOFs.

Despite all the advantages, pristine MOFs display superior electrochemical properties only when combined with conduct

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

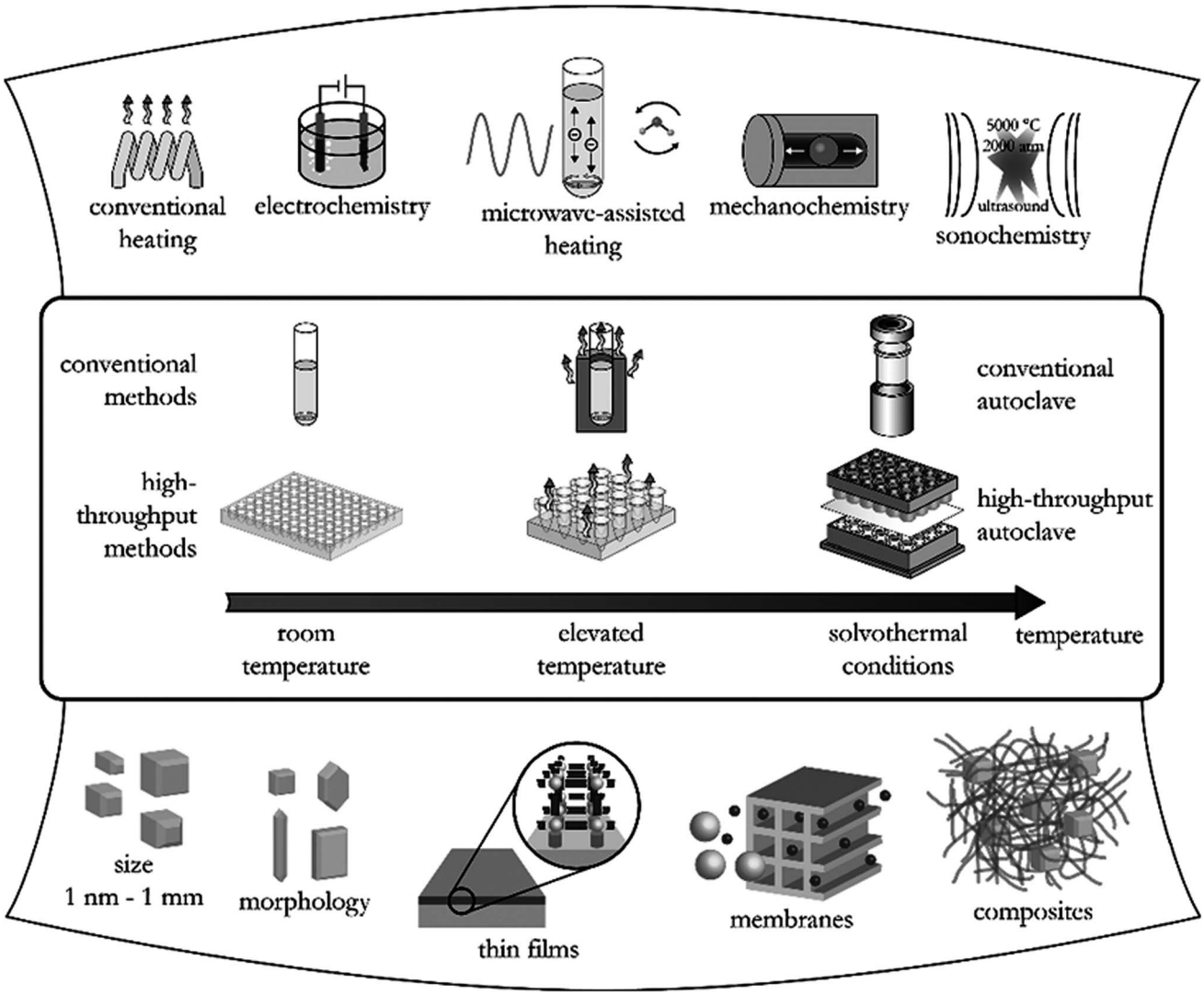



Fig. 2 Different synthetic methods, possible reaction temperatures, and final reaction products in MOFs synthesis.[17]

materials. Conducting MOFs is an upcoming field of research with high expectations in electrochemical applications. While there are limited studies on the commercial potential of conducting MOFs, promising outcomes of MOFs as electrochemical sensors and energy storage are being extensively explored.

## 3 MOFs and biomolecule retention

Immobilization of biomolecules is an important aspect in the fabrication of ideal bioelectrodes to minimize leaching and functional stability. The ultrahigh porosity and surface area of MOFs have been exploited for biomolecule retention. A wide range of biomolecules such as proteins (enzymes, antibodies), carbohydrates, and nucleic acids can be accommodated in MOFs pores as well as linked to their outer surface.[28] Various MOFs are immobilized on the surface of bioelectrodes through covalent, electrostatic, hydrophobic, and van der Waals forces.[29] It results in increased loading of biocatalysts due to increased surface area. The four common approaches used for the retention of biomolecules in MOFs are adsorption, bioconjugation, encapsulation, and diffusion, as discussed below.

### 3.1 Adsorption

Surface attachment of biomolecules to the MOFs can be accomplished in two ways – through surface bonding and by general adsorption. It is a common method for preparing biomolecule-MOFs composites. Biomolecules are anchored to the MOFs surface through weak interactions with no requirement of specific functional groups. Recently, Ma *et al.* used a series of ZIFs as adsorption matrices for immobilization of electrocatalysts, namely glucose dehydrogenase (GDH) and methylene green (MG) in glucose biosensor.[30] ZIF-70, with the largest pore size, was the most efficient matrix for co-immobilization amongst ZIF-7, ZIF-8, ZIF-67, and ZIF-68, in the fabricated bioelectrode as shown in Fig. 4(a). ZIF-70 demonstrated outstanding adsorption capacities towards electrocatalysts and therefore higher sensitivity and selectivity


© 2022 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Table 1 Overview of the synthetic methods of MOFs along with their features, temperature-duration and advantages

| S. no. | Synthetic methods | Features | Temperature-duration | Advantages |
|---|---|---|---|---|
| 1 | Solvothermal (hydrothermal) | Solvents such as water, DMF used | At 353–453 K for 2–3 days | • High yield of MOF<br>• High surface area and porosity of MOF<br>• Highly crystalline MOF |
| 2 | Non-solvothermal | Under the boiling point of solvent | At 298 K for several days to months | • Carried out at room temperature or simple heating<br>• Simple equipment required |
| 3 | Electrochemical | Through anodic dissolution or cathodic deposition | At 273–303 K for 10–30 min | • Suitable for synthesis of large quantity of MOF<br>• No formation or separation of anions required |
| 4 | Mechanochemical | Chemical transformation through milling or grinding | At 298 K for 30 $min^{-2}$ h | • No washing or activation required<br>• Suitable for metal precursor with low solubility |
| 5 | Microfluidics | Reaction in microfluidic channel | At 323–423 K for few minutes | • Fast crystallization rate<br>• Great control over morphology of MOF crystal |
| 6 | Microwave-assisted | Interaction between reactants and radiation | At 303–373 K for 4 $min^{-4}$ h | • Good efficiency in short duration<br>• Great control over reaction parameters and morphology of MOF crystals |
| 7 | Ionothermal | Ionic liquids used as solvent and template | At 333–373 K for 6 h | • Environmentally-friendly method<br>• Great control over morphology of MOF crystals |
| 8 | Sonochemical | Ultrasonic waves used for acoustic cavitation effect | At 272–313 K for 30–180 min | • Fast crystallization rate<br>• Suitable for small particle size |
| 9 | Spray drying | Atomization of MOF precursor solution using spray drier | At 423–453 K for 5–10 min | • Fast and simple technique<br>• Suitable for multi-metallic MOFs |
| 10 | Flow chemistry | Continuous MOF synthesis in tube reactors | At 353 K for 5–10 min | • Low material and energy consumption<br>• Ease in down streaming |

towards glucose with a linear range of 0.1–2 mM. The approach was highly beneficial due to various advantages offered by highly porous ZIF-70, such as better adsorption capability and chemical stability due to interaction with functional groups. Interestingly, MOFs have also been used in the development of biosensors for SARS-COV-2 detection.[31] MOF surface could provide an efficient adsorption platform for fluorophore-labeled probes through electrostatic interactions, hydrogen bonding, and π–π stacking. Moreover, there is possibility of increased recovery of fluorescence due to the weak affinity of hybridized DNA towards MOFs. MOFs as adsorption and fluorescence quenching platform could therefore increase the sensitivity and rapidity of detection in SARS-COV-2. The adsorption technique is quicker and easier as compared to other techniques with no strict requirement on MOF's pore size. This technique also allows reaction parameters to exceed the denaturizing limits of biomolecules as it could be adsorbed after MOFs synthesis. Adsorption also enables better recovery and recyclability of biomolecules in repeated cycles. However, due to the weak protection effect of MOFs, exposed biomolecules could easily detach from the MOFs surface and are prone to leaching. Therefore, there is a growing need for alternative immobilization methods using porous nanomaterials.

### 3.2 Conjugation with linker

Bioconjugation refers to either covalent linkage or electrostatic force of attraction to anchor biomolecules on the MOFs surface. The bio-functionalized MOFs synthesised through this approach are advantageous for simplistic electrochemical biosensing and signal amplification. Biomolecules covalently attached with MOFs functional groups are highly stable and exhibit improved catalytic activity. A porphyrin-functionalized MOFs bioconjugated with streptavidin (SA) named as FeTCPP@MOF-SA using HKUST-1(Cu) as shown in Fig. 4(b) was fabricated for DNA sensing.[32] The setup enabled highly sensitive and specific recognition through the allosteric switch of hairpin DNA. The LOD of 0.48 fM was recorded with the linear range of 10 fM to 10 nM. HKUST-1(Cu) supported a label-free detection

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

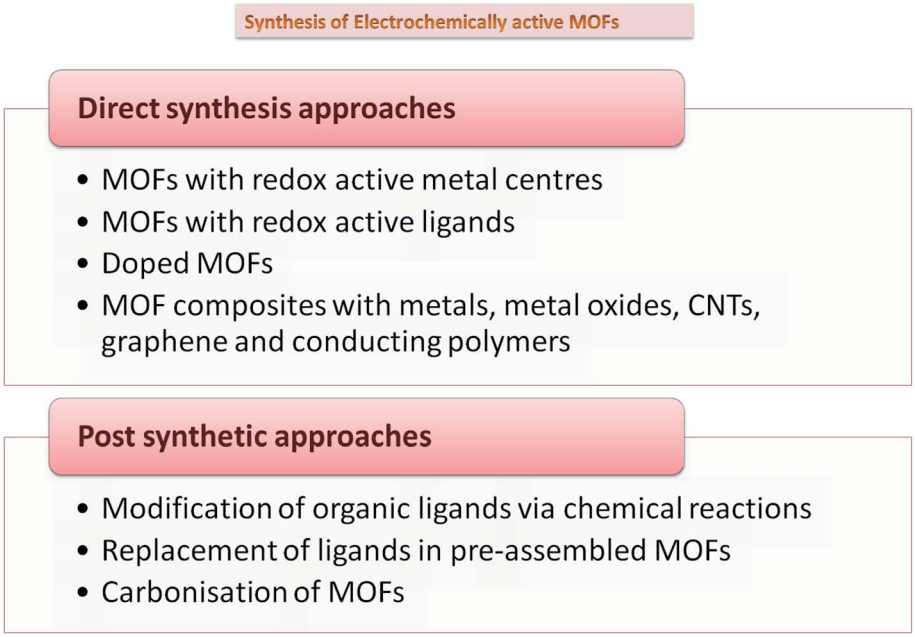



Fig. 3 Different approaches to synthesize electrochemically active MOFs.

approach by working as a signal probe for mediating recognition between SA and aptamer. Bioconjugated MOFs present new avenues for mimetic catalysis as a promising signal transduction platform. In another work by Kumar *et al.*, MOF-5 was bioconjugated with anti-bovine serum albumin (anti-BSA) for electroluminescence biosensing.[33] The pendent (side)

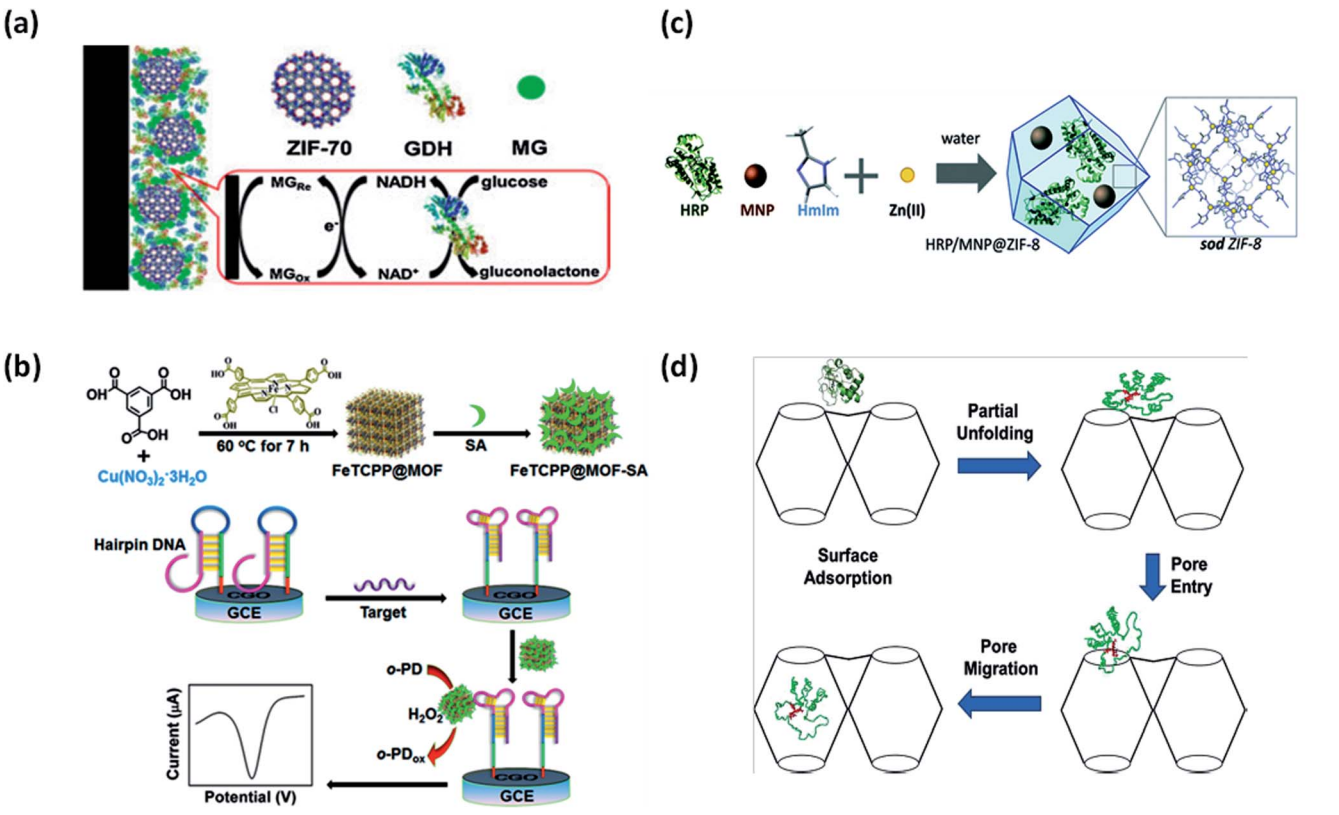



Fig. 4 Different immobilization strategies using MOFs. (a) Adsorption of glucose dehydrogenase (GDH) at ZIF-70 surface.[30] (b) Bioconjugation of FeTCPP@MOF with streptavidin (SA).[32] (c) Encapsulation of horseradish peroxidase (HRP) and magnetic nanoparticles (NPs) in ZIF-8.[34] (d) Diffusion of Cyt c into the cavities of Tb-mesoMOF.[35]


© 2022 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

group on MOF-5 surface was activated through simple organic reactions followed by covalent protein conjugation with anti-BSA. The pendent group of the organic linker was activated using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC). Despite strong bioconjugation of anti-BSA on MOF-5, no loss of enantioselectivity was observed. This study further encouraged the application of MOFs in molecular sensing and immunohistochemistry assays. One major advantage of using this approach is strong interaction between biomolecule and MOFs, which guarantees structural stability. The formation of robust peptide bonds between amino group of enzyme and carboxylate group of MOFs is highly feasible in bioconjugation. Moreover, higher recyclability of biomolecules is assured since there are minimal chances of leaching. However, in this particular approach, the biomolecule linked to the surface is susceptible to denaturation under harsh conditions, which can be overcome by encapsulation, as discussed in the next section.

### 3.3 Encapsulation

Encapsulation involves capturing the biomolecule within MOFs, which can be *in situ* or *ex situ*. The *in situ* encapsulation has emerged as a powerful tool wherein MOFs grow around biomolecules leading to high loading efficiency and negligible leaching. Encapsulation, being a rapid and low-cost approach, is likely to accelerate the application of important industrial biomolecules in the commercial sector. However, the *in situ* strategy is limited to only operation under mild conditions in aqueous solutions due to the risk of denaturation of biomolecules, greatly restricting the diversity of MOFs. ZIFs can be adopted as suitable nanomaterials for biomolecule encapsulation. Recently, magnetically and catalytically active ZIFs have been proposed as recyclable MOF-based biocatalysts. Ricco *et al.* reported a one-pot synthesis method for dual encapsulation of horseradish peroxidase (HRP) and magnetic NPs in ZIF-8, as shown in Fig. 4(c).[34] A five-fold increase in HRP activity was observed along with improved reusability and specific activity of the enzyme. Encapsulation offers various advantages over other immobilization techniques with respect to stability and prevents the leaching of biomolecules from the host matrix. Additional benefits of size selectivity and substrate specificity can be achieved due to the specific pore size of MOF, which is not possible in surface immobilized biomolecules. MOF surface acts as a protective shell against competitive inhibitors and denaturation agents. Moreover, it physically separates biomolecules, thereby preventing aggregation. However, a major drawbacks of encapsulation are the inability to accommodate large biomolecules, hindrance in mass transfer efficiency, and accessibility to the biomolecule. Efforts should be made to study the underlying mechanisms and host–guest interactions to develop standard procedures for MOF encapsulation to remediate these limiting issues.

### 3.4 Diffusion

Diffusion is defined as the infiltration of biomolecules within the MOF pores. This technique is especially useful for the immobilization of large biomolecules such as proteins inside MOFs with relatively small pores. For example, cytochrome c (Cyt c) has been successfully diffused through partial folding within Tb-mesoMOF, as shown in Fig. 4(d).[35] An interesting observation was made regarding the translocation of enzymes through unique conformational changes in enzyme structure at the MOF surface during immobilization. It mimics the activity of proteins undergoing conformational changes during translocation in organelles with narrow pores. This strategy can further open new avenues for MOFs in protein translocation research. Nevertheless, this approach is greatly limited due to concerns over the slow rate of diffusion and enzyme leaching. Unfortunately, limited biomolecules could be immobilized in MOFs through the diffusion process. Therefore, a growing need for better alternatives to overcome these limitations is felt.

All the approaches mentioned above provide better stability to biomolecules. MOFs serve as ideal hosts for biological guests attracting applications in drug delivery, biosensing, bioimaging, bioremediation, genetic manipulation, and preservation of biological specimens.[36–38] With promising applications in biomolecule detection, immunoassays, and tumour detection, various researches focus on achieving technical excellence through size, cost, and materials. Fine tuning of structure and porosity of MOFs for specific applications is also a challenging task. Since high outcomes are anticipated in this field of research, a deeper understanding of internal interactions and mechanisms should be explored. Some bottlenecks such as stability in water and mass transfer limitations also hinder the industrial production of MOFs. Therefore, efforts should be directed to replicate the results from lab to pilot scale.

## 4 MOF-based bioelectrodes for biosensors

Electrochemical biosensors utilize working electrodes with biological recognition elements (BRE) attached for their function. Bioelectrodes act as an interface between biological and electrochemical phenomena, which ultimately determines the overall performance of the biosensor.[39] With the paradigm shift towards point-of-care diagnostics, there is a growing need to achieve exceptional structural and functional stability, biocatalysis, sensitivity, and specificity in bioelectrodes through the use of superior nanomaterials.[8,40] MOFs can be explored as host matrix, biocatalyst, and BRE in bioelectrodes for biosensors as discussed. Table 2 provides a summary of recently reported MOFs-based bioelectrodes in electrochemical biosensors.

### 4.1 MOFs as host matrix

Enzyme-based biosensing devices require optimum thermal, chemical, and mechanical environment for proper action and tend to denature under harsh conditions. Therefore, there is a need to provide a suitable working environment for enzymes for their optimum functioning. MOFs have emerged as suitable host matrices for enzyme immobilization due to their remarkable surface area, crystalline nature, porous framework, and high flexibility. MOFs with a large free volume easily

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Table 2 MOFs-based bioelectrodes used in electrochemical biosensors[a]

| S. no. | MOF | Biological recognition element | Target analyte | LOD | Linear range | Reference |
|---|---|---|---|---|---|---|
| 1 | Co-MOF, ZIF-67 | Cobalt oxide hollow nanododecahedra ($Co_3O_4$-HND) | Glucose | 0.58 μM | 2.0 μM to 6.06 mM | 88 |
| 2 | Cr-MOF (MIL-53-CrIII) | — | $H_2O_2$ | 3.52 μM | 25 to 500 μM | 89 |
| 3 | Cu-MOF | Hemin | $H_2O_2$ | 0.14 μM | 10 to 24 400 μM | 90 |
| 4 | Cu-MOF | GOx | Glucose | 14.77 μM | 44.9 μM to 4.0 mM and from 4.0 to 19 mM | 91 |
| 5 | Cu-MOF | Tyrosinase | BPA | 13 nmol $l^{-1}$ | $5.0 \times 10^{-8}$ to $3.0 \times 10^{-6}$ mol $l^{-1}$ | 92 |
| 6 | Cu-MOF | — | miRNA | 0.35 fM | 1.0 fM to 10 nM | 93 |
| 7 | Fe-MIL-88-$NH_2$ | Hemin | ADRB1 gene | 0.21 fM | 1 fM to 10 nM | 46 |
| 8 | Fe-MOF | $Pb^{2+}$-specific DNAzyme | $Pb^{2+}$ | 2 pM | 0.005 to 1000 nmol $L^{-1}$ | 94 |
| 9 | Ni-MOF | — | Urea | 3 μM | 0.01 to 1.12 mM | 95 |
| 10 | Ni-MOF | — | Glucose | 4.6 μM | 20 μM to 4.4 mM | 96 |
| 11 | Pb-BDC-$NH_2$ and Cd-BDC-$NH_2$ | anti-CEA | Carcinoembryonic antigen (CEA) | 0.03 pg $mL^{-1}$ | 0.3 pg $mL^{-1}$ to 3 ng $mL^{-1}$ | 97 |
| 12 | Pb-BDC-$NH_2$ and Cd-BDC-$NH_2$ | anti-AFP | Alpha-fetoprotein (AFP) | 0.1 pg $mL^{-1}$ | 0.3 pg $mL^{-1}$ to 3 ng $mL^{-1}$ | 97 |
| 13 | PCN-333(Al) | Microperoxidase-11 (MP-11) | $H_2O_2$ | 0.127 μM | 0.387 μM to 1.725 mM | 11 |
| 14 | Y-1, 4-NDC-MOF | AgNPs | $H_2O_2$ | 0.18 μM | 4 to 11 000 μM | 98 |
| 15 | Y-1, 4-NDC-MOF | CuNPs | $H_2O_2$ | 0.43 μM | 4 to 8500 μM | 98 |
| 16 | ZIF-67 | — | Glucose | 0.99 μM | 48 μm to 1 mM | 99 |
| 17 | ZIF-67 | — | Glucose | 0.66 μM | 2 to 1000 μM | 100 |
| 18 | ZIF-8 | HRP | $H_2O_2$ | 0–800 μM | 1 μM | 101 |
| 19 | ZIF-8 | LAC and GOx | Glucose | 5.347 μM | 1 to 10 mM | 42 |
| 20 | ZIF-8 | BPA | LAC | 5.347 μM | 1 to 20 mM | 41 |

[a] GDH – glucose dehydrogenase, GOx – glucose oxidase, HRP – horseradish peroxidase, ADH – alcohol dehydrogenase, LAC – laccase, BPA – bisphenol A.

accommodate large as well as small enzymes with great control over surface topology. The enzymes adsorbed or entrapped inside MOFs have increased stability and display great tolerance to dynamic pH changes.

Li *et al.* constructed a ZIF-8@LAC (LAC = laccase) based bioelectrode (cathode as well as anode) through encapsulation and surface immobilization techniques for the detection of bisphenol A (BPA).[41] ZIF-8 was combined with bacterial cellulose (BC)/carboxylated multi-walled carbon nanotubes (c-MWCNTs) to provide flexibility. ZIF-8@LAC showed improved biocatalysis, structural stability, and enzyme recyclability. No hindrance in the diffusion process of substrate and product through ZIF-8 was encountered upon loading of LAC. The ZIF-based bioelectrode exhibited LOD of 5.347 μM and a maximum power density of 3.68 W $m^{-3}$. It was successful in removing 98% of BPA within 6 h. In a similar work by the same group, they proposed an EBFC-based self-powered biosensor using ZIF-8 as the host matrix, as shown in Fig. 5(a).[42] They adopted *in situ* growth of ZIF-8 encapsulated with glucose oxidase (GOx) and LAC on cellulose acetate (CA) membrane, followed by MWCNTs and AuNPs absorption. CA/ZIF-8@GOx/MWCNTs/Au was used as bioanode and CA/ZIF-8@LAC/MWCNTs/Au as biocathode in a dual chamber biosensor to monitor glucose levels with LOD of 5.347 μM. The electrodes showed large surface area, porosity, flexibility, and thermal stability due to the use of ZIF-8. The study presents promising results as a non-invasive glucose biosensor with continuous detection and structural stability of up to 15 h.

MOFs can be utilized to obtain increased recovery and reusability of enzymes for long term usage. MOFs provide a biocompatible microenvironment for enzymatic biosensing of a wide range of analytes. Despite major advantages, leaching out of toxic metal ions poses a great risk to the safe application of MOF-based bioelectrodes. Therefore, there is a great demand for biocompatible MOFs with synergistic effects of both MOFs and enzymes. The biocompatibility of MOFs is determined by factors such as the nature of metal ions and organic linkers along with some physical parameters. Several metals such as Zr and Ti are non-toxic to humans and are suitable for the synthesis of biocompatible MOFs. Likewise, many organic ligands naturally present in the human body, such as aspartate and adenine, could be explored as linkers. Physical properties of MOFs, including shape, size, and charge, also influence their biocompatibility. Moreover, the water stability and biodegradability of MOFs are of utmost importance for ensuring their biocompatibility. On the biomedical front, biocompatible MOFs have the potential to augment drug delivery and translational research as smart nanocarrier systems. The intelligent application of MOFs as an ideal host matrix is expected to open new avenues for the synthesis of robust bioelectrode for biosensors.


© 2022 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

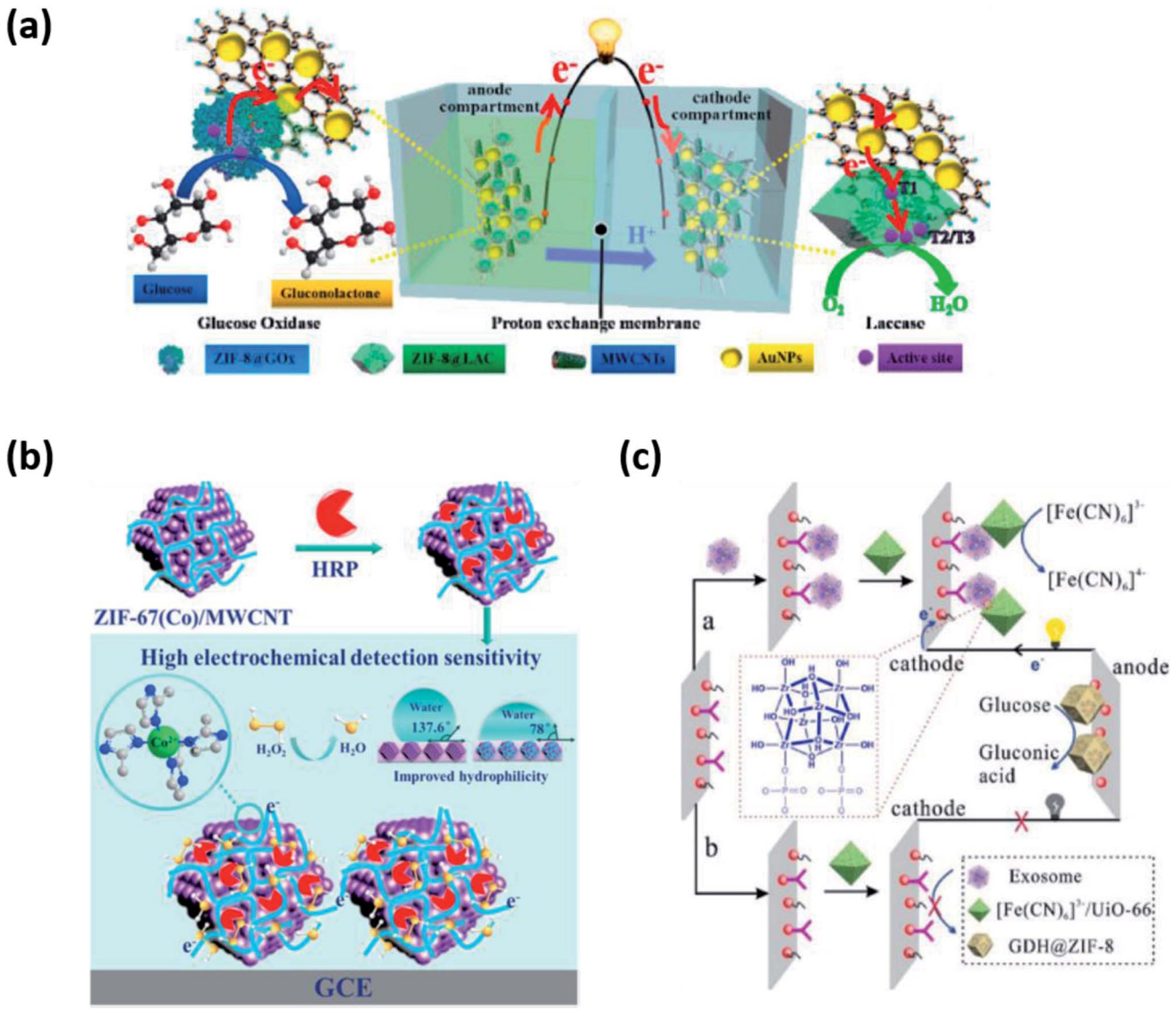



Fig. 5 Applications of MOFs in biosensors (a) ZIF-8 as a host matrix for GOx.[42] (b) PCN-333(Fe) as a synergistic biocatalyst.[44] (c) ZIF-8 as a biological recognition element for exosome detection.[45]

### 4.2 MOFs as electrocatalysts

MOFs not only serve as ideal host matrices but also contribute to synergistic biocatalysis along with enzymes. Several MOFs such as ZIF-8, ZIF-70, and MIL-100(Fe) have been explored to act as catalysts and facilitate electron transfer without compromising the enzyme activity. They help establish electrical communication between the redox enzymes and the bioelectrodes, thereby promoting biocatalysis at low over potentials and high anodic current densities. As an excellent example, a water stable iron-based MOF known as PCN-333(Fe) was employed in the electrochemical biosensing of $H_2O_2$.[43] It demonstrated an LOD of 0.09 μM in the detection range of 0.5 μM to 1.5 mM. PCN-333(Fe) immobilized with HRP showed synergistic catalytic properties of both MOF and enzyme. PCN-333(Fe) proved to be advantageous in many aspects (i) the size similarity of PCN-333(Fe) and HRP ensured highly effective enzyme loading (ii) strong interactions prevented conformational changes of HRP and eliminated the risk of aggregation and leaching (iii) ultrahigh porosity and surface area of PCN-333(Fe) easily accommodated the enzyme without restricting its activity and (iv) most importantly, the presence of $Fe^{3+}$ ions in PCN-333(Fe) with intrinsic peroxidase-like activity acted as biocatalyst further improving the sensitivity of electrochemical biosensing. It was observed that PCN-333(Fe) not only ensured more effective shuttling of electrons between the electrode and HRP redox centers but also increased the collision frequency between the analyte and enzyme, thereby contributing to improved biocatalysis. Moreover, the HRP@PCN-333(Fe) modified electrode functioned well in harsh environments (extreme temperature and pH), showing improved operational stability. In another example, ZIF-67 immobilized with HRP was used to increase the sensitivity of electrochemical biosensing of $H_2O_2$, as shown in Fig. 5(b).[44] The MOF-enzyme system

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

witnessed a 1.3-fold increase in sensitivity recorded as 315 μA $mM^{-1}$ $cm^{-2}$ with a greater affinity towards $H_2O_2$. The LOD was recorded as 0.09 μM in the linear range of 0.5 μM to 1.5 mM. The researchers have emphasized the synergistic biocatalysis of MOF and enzyme, rendering superior performance to electrochemical biosensors.

Hence, the role of MOFs is quite evident in improving the catalytic efficiency of enzymes and as a result, the sensitivity of electrochemical biosensors. MOFs not only confer structural benefits but also the presence of a metal center fastens the rate of electron transfer further increasing the biocatalytic efficiency of bioelectrodes. The positive impact of MOFs on electrochemical biocatalysis is commendable and holds a great future for advancements in biosensing technologies. Therefore, MOF-based bioelectrodes are gaining increased interest for feasible applications in miniaturized and highly stable biosensors.

### 4.3 MOFs enhancing the sensitivity and selectivity of bioelectrodes

In addition to improving the stability and sensitivity of biosensors, MOFs such as ZIF-8, HKUST-1, Cu-MOF, and Ru-MOF have been reported to function as BRE for capturing and sensing diverse biomolecules such as DNA, RNA along with cancer biomarkers such as exosomes and microRNAs (miRNAs). MOFs are commonly used as a sensing unit to amplify electrochemical signals exploiting their biomimetic catalytic feature. It has been observed that MOFs simultaneously provide enhanced sensitivity and selectivity through strong covalent bonding, electrostatic attractions, hydrogen bonding, and hydrophobic interactions between the analyte and metal node/linker.

Recently, a self-powered biosensor based on dual MOFs bioelectrodes for the detection of exosomes, a cancer biomarker, has been reported. A GDH@ZIF-8 modified anode and a $[Fe(CN)_6]^{3-}$/UiO-66 modified cathode were fabricated in order to achieve more stable and sensitive exosome biosensing, as shown in Fig. 5(c).[45] It was observed that ZIF-8 provided better catalytic efficiency and stability to GDH. On the other hand, $[Fe(CN)_6]^{3-}$/UiO-66 served as nano-enrichment carriers by accepting electrons from the anodic oxidation, generating electrical power. Dual MOF-based bioelectrodes enabled additional sensitivity through the formation of Zr–O–P bonds between the metal node of UiO-66 and phosphate groups of an exosome phospholipid bilayer. The LOD was noted to be 300 particles per mL. The aforementioned dual MOFs bioelectrodes based SPB was also successful in identifying exosomes from different cell lines conferring specificity. Further improvement in the anti-interference ability of the biosensor by use of nanomaterials holds great promise in the early diagnosis and treatment of cancer.[9] Yuan *et al.* modified glassy carbon electrode attached with Fe-MIL-88-$NH_2$ probe loaded with Cu(II) ions, hemin, and PtNP.[46] The fabricated bioelectrode Pt/Hemin@Fe-MIL-88-$NH_2$/$Cu^{2+}$ not only assisted in the synergistic electrocatalytic oxidation of $H_2O_2$ but also improved binding affinity towards the substrate. The LOD was noted to be 0.21 fM in the range of 1 fM to 10 nM. The technique proved promising from a clinical point of view, with no longer needing to label or pre-amplify samples. However, the fabrication process is quite long. There is thus a need to achieve a consistent methodology for better stability of biosensors.

MOF-based biosensors are commercially promising in terms of rapid detection and facile synthesis and present an exciting method for real time and point-of-care diagnostics. Their tunable properties, wide range of working conditions (extreme temperature and pH), and low cost of fabrication make them suitable for scale-up in the biomedical industry. However, some limitations in the area of binding kinetics, electric communication between enzyme and electrode, multiplexed analyses, and MOF-coupled transducers remain to be addressed. There are limited studies regarding biocompatibility, toxicity, stability in water, and false-positive results are some bottlenecks limiting the mass production of MOF-based bioelectrodes for biosensors. Nevertheless, efforts should be made to circumvent these issues for improved performance of MOF-based bioelectrodes in biosensors.

## 5 MOF-based bioelectrodes for biofuel cells

Fuel cells are regarded as one of the best green renewable technologies with the ability to harvest electrical energy from the chemical energy of biological processes.[47,48] Since its invention in 1964 by Yahiro *et al.*, EBFCs have witnessed structural and functional advancements for the past six decades.[49] Nanomaterials such as carbon nanotubes and metal NPs are the fore-runners and hence are the ideal choice for various uses in EBFCs and biosensors. More recently, MOFs have found versatile roles in EBFCs in and as bioelectrodes, electrocatalysts, electrolyte membrane, biomolecule retention, and fuel storage. Table 3 provides a summary of recently reported MOFs-based bioelectrodes in biofuel cells.

### 5.1 MOFs as electrocatalysts

In the past decade, enzyme@MOFs combination has been thoroughly explored as promising electrobiocatalysts. Lower stability, reusability, and recovery of enzymes are limitations in mass production. Enzymes such as lipase,[12] hydrolase,[50] catalase,[20] and GOx,[51] when immobilized on MOF surface, offer various advantages to overcome these limitations. The high surface area of MOFs is perfect for high enzyme loading, and its porous structure supports enzyme ingress and diffusion of substrates and products. However, poor conductivity and instability during solvent extraction in MOFs are some functional limitations. With improved electrocatalytic activity and stability, functionalized MOFs offer dual functionality in fabrication as well as encapsulation of NPs.

In an effort to obtain superior properties of MOFs in MOF-derived NPs, several researchers have used MOF templates as precursors for the synthesis of transition metal/metal oxide carbon and carbon NPs such as M–N–C, M–$MO_x$–N–C, N-doped-C and NP-C as electrocatalysts.[22,52,53]. ZIF-67, a Co-based MOF was reported as a precursor for the fabrication of doped carbon nanotubes of hollow frameworks (NCNTF-700).[54] ZIF-67 has


© 2022 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Table 3 MOF-based bioelectrodes in biofuel cells[a]

| S. no. | MOF | Function of MOF used | Fuel (oxidised) | Power density/current density | Reference |
|---|---|---|---|---|---|
| 1 | ZIF-8 | Nanocarriers | Glucose | 23 μW $cm^{-2}$ | 82 |
| 2 | Al-PCP | Immobilization of GOx | Glucose | 0.548 mW $cm^{-2}$ | 102 |
| 3 | IRMOF-8 | Bioanode catalyst | Alcohol | 0.25 ± 0.03 mA $cm^{-2}$ (current density) | 103 |
| 4 | MAF-7 | Biocatalyst | Glucose | 119 μW $cm^{-2}$ | 104 |
| 5 | Co-MOF, ZIF-67 | Electrocatalyst | Glucose | 0.3 mA $cm^{-2}$ (current density) | 88 |
| 6 | MOF derived *meso*-$CuCo_2O_4$ microspheres | Template precursor | Glucose | 0.33 mW $cm^{-2}$ | 105 |
| 7 | ZIF-8 | Electrocatalyst | Oxygen | — | 106 |
| 8 | MIL-100(Fe) | Immobilization matrix | 2,2′-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) | — | 107 |
| 9 | Cu-MOF | Electrocatalysts | $H_2O_2$ | — | 108 |

[a] GDH – glucose dehydrogenase, GOx – glucose oxidase, HRP – horseradish peroxidase, ADH – alcohol dehydrogenase, ABTS – 2,2′-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid).

been found to provide structural advantage over traditional Pt/C electrocatalysts with improvement in ORR and OER. It exhibited a current density of 10 mA $cm^{-2}$ at 1.60 V. The bi-functional electrocatalysts exhibited remarkable stability and functionality. It is also worth noting that pyrolysis temperature was indirectly proportional to ORR activity of NCNTF-700, with the best performance at low temperatures. Similarly, many researchers have harnessed the structural advantage of MOFs to produce transition metal/metal oxide based nanocarbon electrocatalysts.[55–57] Jahan *et al.* have reported a tri-functional catalyst based on graphene oxide (GO)/Cu-MOF. It showed stability and catalytic activity towards ORR, OER, and HER.[58] The system had better charge transport and porosity due to synergy between Cu-MOF and GO. The GO/Cu-MOF offered a maximum power density of 110.5 mW $cm^{-2}$, presenting itself as an economical alternative to Pt/C catalyst. MOF-based carbon nanotubes and graphene have also been studied several times owing to their high electrocatalytic efficiency.[59,60]

A recent study addressed this issue by introducing microperoxidase-11 (MP-11) into a recently discovered mesoporous MOF named as Tb-mesoMOF, thereby introducing novel MP-11@mesoMOF.[61] MP-11 with a heme center is a key player in peroxidase reactions and has been extensively used in biosensors, energy conversion, and biofuel cells. The application of MP-11 is limited due to its aggregation in solution, thus hindering accessibility of the heme group. The 3.0 and 4.1 nm pore size of Tb-mesoMOF easily accommodated MP-11 inside its nanoscopic framework, which was confirmed through decreased BET surface area from 1935 $m^2$ $g^{-1}$ to 400 $m^2$ $g^{-1}$. Also, a color change from colorless to dark red was observed on saturation. A significant increase in the catalytic activity of MP-11@mesoMOF was observed as 7.58 × $10^{-5}$ mM $s^{-1}$ for approximately 30 min along with a higher substrate conversion of 48.7%. Furthermore, the recyclability of MP-11@Tb-mesoMOF is tested at different cycles. The MOF-enzyme system maintained its activity up to six cycles with no evidence of leaching due to strong hydrophobic interactions between the MOF and enzyme, as shown in Fig. 6(a). This study is the first of its kind to use mesoporous MOF for enzyme immobilization and shows the tremendous potential of mesoporous MOFs as an ideal host matrix material for future use.

### 5.2 MOFs as electrolyte membranes

MOFs as electrolyte membranes is one of the largest research areas of MOF's applications in fuel cells. Exceptional super protonic conductivity as high as 7.89 × $10^{-2}$ S $cm^{-1}$ at 60 °C and 95% relative humidity (RH) with the help of MOF-808 has been achieved.[62] Conventionally, an ideal electrolyte membrane should be proton conducting, electrically insulating, and prevent fuel crossover. MOFs, despite their porous and crystalline structure, which contradicts proton-conducting membranes in fuel cells, emerged as one of the most promising electrolyte membranes due to several reasons.[63] A major reason being the presence of counter ions in MOFs, which favor proton conduction by both Grotthuss (proton hopping) and vehicle (self-diffusion) mechanisms.[64] Different counter ions containing hydroxyl, carboxylic, and ammonium-based functional groups have been used to functionalize MOFs for superior proton conduction through hydrogen bonding. Moreover, MOFs present an additional advantage through loading of suitable guest molecules in their pores in the prevention of fuel crossover and thus voltage loss.[65] It is also challenging to develop anhydrous proton conducting materials operational above 80 °C. To overcome this, imidazole has been encapsulated in aluminum MOF to form a hybridized polymer with superior proton conductivity in anhydrous conditions.[66]

Two MOFs with a common metal ion as aluminum but different organic linkers, namely NDC (1,4-naphthalenedicarboxylate) and BDC, were studied for proton conductivity with imidazole as a guest molecule. Imidazoles have an intermolecular proton transport pathway due to the two tautomeric forms alternating between two nitrogen atoms. Al(μ2-OH)(ndc) showed a higher proton conductivity of 2.2 × $10^{-5}$ S $cm^{-1}$ that of Al(μ2-OH)(bdc), which is 1.0 × $10^{-7}$ S $cm^{-1}$

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

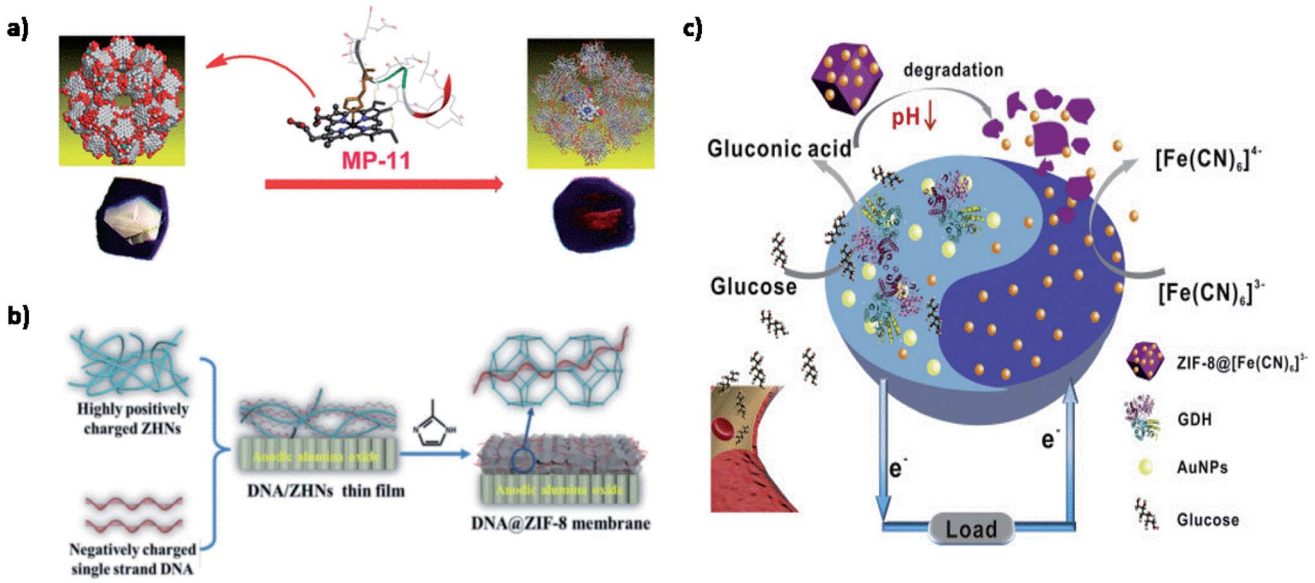



Fig. 6 Applications of MOFs in enzymatic biofuel cells. (a) MP-11@mesoMOF as an electrocatalyst.[61] (b) DNA@ZIF-8 as an electrolyte membrane.[70] (c) Anode-driven cathodic fuel release *via* pH-responsive ZIF-8 nanocarriers.[82]

at 120 °C. This can be explained by the hydrophobic pore surfaces of Al(μ2-OH)(bdc), leading to increased host–guest interaction and hence decreased mobility. Similarly, histamine has been used as a guest molecule to achieve excellent proton conduction due to its unique structural behavior.[67] Al(OH)(ndc) encapsulated with histamine showcased proton conductivity of $3.0 \times 10^{-3}$ S cm$^{-1}$. This is due to intramolecular proton exchange in histamine facilitating reorientation, thereby increasing the rate of the Grotthuss mechanism. High concentration, approximately one histamine per aluminum ion, and its dense packaging in MOF contributed to increased proton conductivity. Better electrode kinetics and lower electrode poisoning can be achieved at higher temperatures but since many existing fuel cells rely on water for proton conduction, one major limiting factor in electrolyte membrane is their dehydration temperature. Based on Nafion, the gold standard of polymer electrolyte membrane, a sulfonated MOF called $Na_3$(thbts) (thbts = 2,4,6-trihydroxy-1,3,5-benzene trisulfonate) was synthesized for potential anhydrous operation.[63] It was similar to Nafion, which is a sulfonated tetrafluoroethylene based fluoropolymer–copolymer having proton conductivity of $2.0 \times 10^{-3}$ S cm$^{-1}$. To make it operational at a high temperature of about 150 °C, water molecules in its structure were replaced with 1*H*-1,2,4-triazole (Tz) that achieved proton conduction in the range of 2 to $5 \times 10^{-4}$ S cm$^{-1}$.

Interestingly, the proton-conducting metal–organic framework 2 (PCMOF2) has been fabricated as a gas separator to prevent fuel crossover. Recently, a hybrid membrane of Zn-MOF/Nafion has been synthesized to improve the electrochemical performance of Nafion electrolyte membrane.[68] Better proton conductivity of $7.29 \times 10^{-3}$ S cm$^{-1}$ has been achieved with 5% doping and 58% RH at 80 °C. The 1.87 times increase as compared to pristine Nafion membrane can be attributed to the host–guest hydrogen bonding, uncoordinated carboxyl oxygen atoms, and sulfonic acid groups to enhance the acidity and hydrophilicity. Further, $NH_3$ as a guest molecule was incorporated into this hybrid membrane to study its promotion effect. Zn-MOF-$NH_3$/Nafion-5 exhibited nearly 5.47 times proton conductivity of the pristine Nafion membrane, which was $2.13 \times 10^{-2}$ S cm$^{-1}$. In addition to forming hydrogen bonds with the guest 1,2-bis(4-pyridyl)ethane molecule, Zn ions combined with intercalated water, resulting in the formation of different types of hydrogen bonds. Good single cell performance was recorded with a maximum power density of 212 mW cm$^{-2}$ and a current density of 630 mA cm$^{-2}$, making it a promising high-performance Nafion hybrid membrane. Despite advancements in this field, the preparation of cost effective, non-toxic, and long-term structural stable electrolyte membranes is still a challenge.

Biocompatible MOFs are gaining interest due to their low toxicity, facile synthesis, and outstanding stability. Biomolecules/MOFs have been recently explored for their intrinsic proton conduction properties. Proteins and DNA found in the lipid bilayer are known to play an important role in proton and ion transmembrane transport. Inspired by this natural phenomenon, biomolecules have been used as proton carriers and proton donors for the successful generation of hydrogen bonds and coordination networks. Wang *et al.* have synthesized an amino acid-based Zr-MOF using L-aspartate spacer.[69] The MIP-202(Zr) had proton conductivity up to 0.011 S cm$^{-1}$ at 363 K with 95% RH. This can be explained by accessible $NH_3$ groups on aspartate linker acting as a proton source and mesoporous structure favoring hydrogen bond formation. The robust 3D microporous framework displayed good hydrolytic and chemical stability suitable for commercial production. Moreover, the tunable properties such as particle size and biocompatible linkers have opened up new avenues in the biotechnological and pharmaceutical industries. Similarly,


© 2022 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

a DNA@ZIF-8 proton conducting membrane was developed by solid confinement conversion process.[70] The hydrophilic nature of DNA due to the presence of amidogen and phosphate groups improved the proton conductivity under 97% RH up to $3.40 \times 10^{-4}$ S cm$^{-1}$ at 25 °C and 0.17 S cm$^{-1}$ at 75 °C. Additionally, methanol crossover was significantly prevented due to a decrease in pore size upon loading of ssDNA (Fig. 6(b)). The system was successful in the construction of direct methanol fuel cells (DMFC) as MOFs-based PEM with a maximum power density of 9.87 mW cm$^{-2}$. Therefore, biomolecule modified MOFs as electrolyte membranes have encouraging results for improving electrochemical performance and stability. Besides this, the eco-friendly synthesis approaches should be explored for commercial outreach.

### 5.3 MOFs for fuel storage

Various porous materials like zeolites, ammonium phosphates, and coordination polymers have been extensively studied for application in fuel storage and transport. However, challenges in the development of robust synthesis methods and structurally stable porous networks over a wide range of temperature and pressure are still faced. Also, it is of key importance to maintain the interpenetrating network of porous materials that do not disintegrate upon the removal of guest molecules. MOFs present one such promising class of materials for the storage of fuels such as hydrogen, ethane, and methane with unique features of porous structure, high surface area, low density, and large-scale flexibility in comparison to the conventionally used zeolites[71–73] [$\{[CuSiF_6(4,4\text{-bipyridine})_2]\cdot 8H_2O\}_n$] was the first MOF to be reported as a novel methane adsorbent in 2000.[74] Based on argon adsorption, it had a specific surface area of 1337 m$^2$ g$^{-1}$, micropore volume of 0.56 mL g$^{-1}$ and was successful in similar adsorption capacity to that of zeolites and activated carbon.

MOFs have started to gain popularity as an ideal candidate for hydrogen storage in fuel cell technology due to their high energy density capability. The hydrogen uptake in MOFs largely depends on the heat of adsorption at low pressure, surface area at moderate pressure, and free volume at high pressure.[75] MOF-5 consisting of $Zn_4O(BDC)_3$ was the first to achieve reversible hydrogen adsorption up to 4.5 weight percent at 78 K, which is equivalent to 17.2 hydrogen molecules per formula unit.[76] At a pressure of 20 bar, 1.0 weight percent hydrogen adsorption at room temperature was observed, and increased loading capacity was found owing to two well defined hydrogen binding sites formed with zinc and the BDC linker. Similar experiments were replicated at room temperature and 10 bar pressure with IRMOF-6 and IRMOF-8, which interestingly gave doubled and quadrupled hydrogen uptake results, respectively, comparable to the highest capacity attained for carbon nanotubes at cryogenic temperatures.

There is a growing need to optimize hydrogen interactions, gravimetric (mass), and volumetric (size) surface area densities. It has been found that smaller pores in MOFs have higher hydrogen uptake capacity. In the case of large pores, the large voids may be reduced by interpenetration or catenation for better efficiency. Strategies such as synthesis like topological modification in pore size,[77] framework catenation,[78] usage of additional adsorbate surfaces,[79] open metal sites,[80] and light metal MOFs[81] are being adopted to upscale the hydrogen affinity. Recently, ZIF-8 was explored for application in a smart membrane-less EBFC for the controlled release of cathodic fuel, as shown in Fig. 6(c).[82] The cathodic acceptor $[Fe(CN)_6]^{3-}$ was encapsulated within ZIF-8 to form $[Fe(CN)_6]^{3-}$@ZIF-8, functioning as a pH responsive nanocarrier. ZIF-8 provided excellent pH sensitivity, stability and also decreased the internal resistance from the membrane. The LOD was recorded to be 1.6 mM with a maximum power density of 23 μW cm$^{-2}$. The EBFC was driven by the production of gluconic acid up on glucose anodic oxidation. $[Fe(CN)_6]^{3-}$@ZIF-8 nanocarriers were degraded due to a change in pH to acidic, resulting in the regulated release of $[Fe(CN)_6]^{3-}$ into the electrolyte, thereby driving the electrochemical reaction between the anode and cathode. Further upgrades in the structure of the smart energy conversion devices could make a positive impact on the point-of-care technologies. Membrane-less systems are particularly attractive due to their low cost, high scalability, and superior power densities. However, care must be taken to ensure high reaction selectivity and maintain the optimum efficiency of bioelectrodes.

The porous structures can be utilized for advancements in fuel-cell powered vehicles and next generation clean energy storage. However, further studies are needed to balance between pore size and surface area of MOFs for maximum fuel storage at room temperature in fuel cells. Also, the complexity of synthetic modifications, poor stability, and narrow range of parameters limit the industrial applicability of MOFs.

## 6 Challenges and future prospects

MOFs have been explored extensively as promising materials for future EBFCs and self-powered biosensors. Researchers have concentrated their efforts to maximize the structural and functional advantages of MOFs into potential solutions for green energy production. Significant milestones regarding the use of MOFs and their derivatives in and as various components of EBFCs such as electrocatalyst, electrolyte membrane, biomolecule retention, and fuel storage have been achieved. MOF composites have shown tremendous potential in a variety of bioelectrochemical applications, as shown in Fig. 7. MOFs not only contribute to structural stability but also improve charge transfer and overall EBFC performance. MOFs established a strong foundation as substitutes for bioelectrode materials with superior biomolecule immobilization, fuel storage, and electrobiocatalysis to yield exceptional power output. The envisioned approaches are currently limited due to the high cost of commercially available MOFs. To address this, polyethylene terephthalate waste is hydrolyzed to obtain terephthalic acid for high value-added and cost-effective MOF synthesis. Few examples of MOFs such as UiO-66, MIL-101, and MOF-5 have already been synthesized through this approach.[83] It is noteworthy that MOFs can be used to deliver the utmost performance with facile synthesis. However, it is still

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

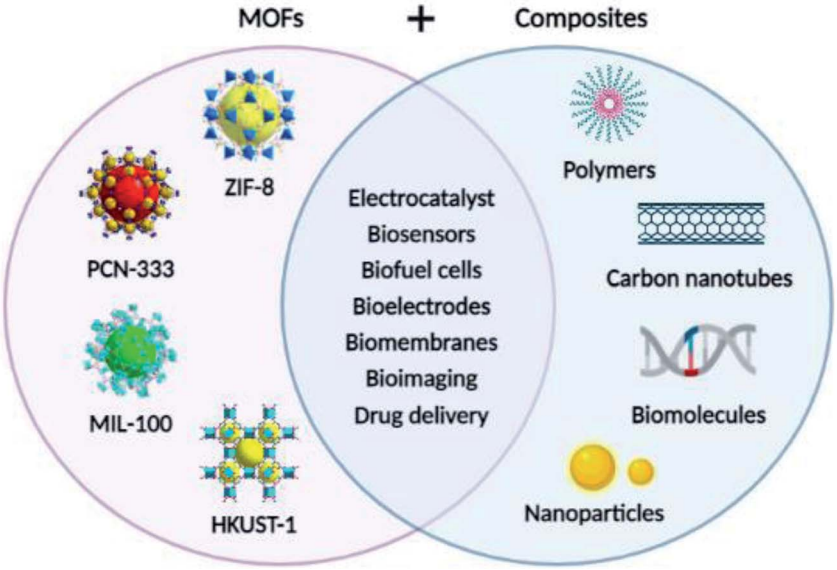



Fig. 7 Bioelectrochemical applications of MOFs and their composites.

a challenging task to carefully regulate their fine structure according to specific applications without affecting their advantageous properties. Advanced synthesis techniques are highly desirable in order to expand their industrial scalability by exploring MOFs with high hydrothermal stability. Post synthesis techniques such as solvent exchange and vacuum activation are commonly practiced that can be switched to less demanding and moderate condition requirements. In the case of characterization, the some techniques such as FTIR spectroscopy require the shaping of MOF powder into pellets, which can impact their surface texture and therefore, there is a need for further studies and improvements. From an application point of view, it is imperative to maintain the stability of MOFs in different solutions and prevent photodegradation. The behavior of MOF in water is an important parameter in defining their chemical stability.[84] For example, MIL-53 has been popular as an adsorbent due to its remarkable water stability.[85] With respect to the photostability of MOFs, it has been observed that terephthalate MOFs are more photostable than trimesate MOFs.[86] Recent development in the field of surface chemistry and nanobioelectronics have made significant contributions to expand MOF knowledge base. A better understanding of structure–property relation, enzyme-MOF interaction, and structural defects should be prioritized. More focus should be given to the application of MOFs in portable electronics for boosting their economic importance in the worldwide industry. More efforts are required for the translation of MOFs and MOF-based materials in implantable medical devices should be made.

## 7 Conclusions

Research and development of novel materials for bioelectrochemical applications have become a major thrust area in the energy sector. The pioneering work on MOFs catalyzed efficient energy conversion and storage, especially in EBFCs and biosensors. MOF attracted a lot of attention in the fabrication of bioelectrodes for superior structure and function. Harnessing the inherent properties of MOF, such as high porosity, surface area, flexibility, and tenability, enables unique prospects in electrobiocatalysis, fuel storage, and biomolecule retention. In this review, MOFs have been highlighted as promising porous nanomaterials for application in EBFCs and biosensors. Various synthetic strategies, including post synthetic modification, have accelerated their applicability in diverse fields. Recent advances in MOF-based bioelectrodes to improve electrochemical performance and the long-term stability of EBFCs have been discussed. Remarkable outcomes have been achieved for the application of MOFs in and as host matrix to support biomolecule immobilization, synergistic biocatalyst to facilitate improved electron transfer, and BRE to increase the sensitivity and selectivity of bioelectrodes. This further enables more opportunities for the fabrication and miniaturization of energy devices. However, research on electroactive MOFs and their application are currently in the nascent phase. Despite encouraging theoretical results, it is of fundamental importance to achieve similar performance of MOF-based bioelectrodes in practical application. Also, the dynamic stability of MOFs in changing environments of pH and temperature for long durations is a prime concern. Efforts should be directed towards the facile and robust synthesis of bio-functionalized MOFs. MOF-based composite materials with high conductivity have been designed using metals as guest materials. These composites have surpassed the drawbacks of single component MOFs. However, poor cycle performance and conductivity of MOFs pose some challenges, which can perhaps be overcome by various techniques. Incorporation of ionic molecules or mobile counter ions within the MOF framework *via* post-synthetic modification can used to boost the charge transport.[87] In-depth studies regarding host–guest interactions, biomimetic


© 2022 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

MOFs and recyclability of these systems should be conducted for value-driven industrial scalability.

## Author contributions

Conceptualization – VA, SS and BDM, writing – original draft – VA and SS, writing – review & editing – BDM, supervision-BDM.

## Conflicts of interest

There are no conflicts of interest to declare.

## Acknowledgements

BDM acknowledges the DST – Science & Engineering Research Board (Government of India) for the award of Distinguished Fellowship (SB/DF/011/2019). SS is thankful to the CSIR India for the award of CSIR-RA fellowship. The authors thank the Vice Chancellor, Delhi Technological University, New Delhi for providing the opportunity to work in the lab.

## References

1 D. Shi, X. Yu, W. Fan, V. Wee and D. Zhao, *Coord. Chem. Rev.*, 2021, **437**, 213794.
2 H.-C. Zhou, J. R. Long and O. M. Yaghi, *Chem. Rev.*, 2012, **112**, 673–674.
3 G. Li and Y. Han, *ACS Cent. Sci.*, 2022, **8**, 150–152.
4 L. Jiao, J. Y. R. Seow, W. S. Skinner, Z. U. Wang and H.-L. Jiang, *Mater. Today*, 2019, **27**, 43–68.
5 H. Li, L. Li, R.-B. Lin, W. Zhou, Z. Zhang, S. Xiang and B. Chen, *EnergyChem*, 2019, **1**, 100006.
6 H.-F. Wang, L. Chen, H. Pang, S. Kaskel and Q. Xu, *Chem. Soc. Rev.*, 2020, **49**, 1414–1448.
7 S. Carrasco, *Biosensors*, 2018, **8**, 92.
8 C. Orellana-Tavra, R. J. Marshall, E. F. Baxter, I. A. Lázaro, A. Tao, A. K. Cheetham, R. S. Forgan and D. Fairen-Jimenez, *J. Mater. Chem. B*, 2016, **4**, 7697–7707.
9 H. Kaur, R. Kumar, A. Kumar, V. Krishnan and R. R. Koner, *Dalton Trans.*, 2019, **48**, 915–927.
10 L. Zhang, Z. Su, F. Jiang, L. Yang, J. Qian, Y. Zhou, W. Li and M. Hong, *Nanoscale*, 2014, **6**, 6590–6602.
11 C. Gong, Y. Shen, J. Chen, Y. Song, S. Chen, Y. Song and L. Wang, *Sens. Actuators, B*, 2017, **239**, 890–897.
12 A. Samui, A. R. Chowdhuri, T. K. Mahto and S. K. Sahu, *RSC Adv.*, 2016, **6**, 66385–66393.
13 G. de Combarieu, M. Morcrette, F. Millange, N. Guillou, J. Cabana, C. P. Grey, I. Margiolaki, G. Férey and J. M. Tarascon, *Chem. Mater.*, 2009, **21**, 1602–1611.
14 S. Maiti, A. Pramanik, U. Manju and S. Mahanty, *ACS Appl. Mater. Interfaces*, 2015, **7**, 16357–16363.
15 P. Sengodu, C. Bongu, M. Perumal and M. Paramasivam, *J. Alloys Compd.*, 2017, **714**, 603–609.
16 J. Yang, P. Xiong, C. Zheng, H. Qiu and M. Wei, *J. Mater. Chem. A*, 2014, **2**, 16640–16644.
17 N. Stock and S. Biswas, *Chem. Rev.*, 2011, **112**, 933–969.
18 F. Y. Yi, R. Zhang, H. Wang, L. F. Chen, L. Han, H. L. Jiang and Q. Xu, *Small Methods*, 2017, **1**, 1700187.
19 Y. Kobayashi, B. Jacobs, M. D. Allendorf and J. R. Long, *Chem. Mater.*, 2010, **22**, 4120–4122.
20 M. Kalaj and S. M. Cohen, *ACS Cent. Sci.*, 2020, **6**, 1046–1057.
21 C. H. Choi, S. H. Park and S. I. Woo, *ACS Nano*, 2012, **6**, 7084–7091.
22 S. Zhao, H. Yin, L. Du, L. He, K. Zhao, L. Chang, G. Yin, H. Zhao, S. Liu and Z. Tang, *ACS Nano*, 2014, **8**, 12660–12668.
23 A. E. Baumann, D. A. Burns, B. Liu and V. S. Thoi, *Commun. Chem.*, 2019, **2**(1), 1–14.
24 J. F. Olorunyomi, S. Teng Geh, R. A. Caruso and C. M. Doherty, *Mater. Horiz.*, 2021, **8**(9), 2387–2419.
25 M. G. Campbell, D. Sheberla, S. F. Liu, T. M. Swager and M. Dincă, *Angew. Chem.*, 2015, **127**, 4423–4426.
26 L. Sun, C. H. Hendon, M. A. Minier, A. Walsh and M. Dincă, *J. Am. Chem. Soc.*, 2015, **137**, 6164–6167.
27 T. C. Narayan, T. Miyakai, S. Seki and M. Dinca, *J. Am. Chem. Soc.*, 2012, **134**, 12932–12935.
28 C. Doonan, R. Riccò, K. Liang, D. Bradshaw and P. Falcaro, *Acc. Chem. Res.*, 2017, **50**, 1423–1432.
29 H. An, M. Li, J. Gao, Z. Zhang, S. Ma and Y. Chen, *Coord. Chem. Rev.*, 2019, **384**, 90–106.
30 W. Ma, Q. Jiang, P. Yu, L. Yang and L. Mao, *Anal. Chem.*, 2013, **85**, 7550–7557.
31 G. A. Udourioh, M. M. Solomon and E. I. Epelle, *Cell. Mol. Bioeng.*, 2021, **14**, 535–553.
32 P. Ling, J. Lei, L. Zhang and H. Ju, *Anal. Chem.*, 2015, **87**, 3957–3963.
33 P. Kumar, A. Deep, A. K. Paul and L. M. Bharadwaj, *J. Porous Mater.*, 2014, **21**, 99–104.
34 R. Ricco, P. Wied, B. Nidetzky, H. Amenitsch and P. Falcaro, *Chem. Commun.*, 2020, **56**, 5775–5778.
35 Y. Chen, V. Lykourinou, C. Vetromile, T. Hoang, L.-J. Ming, R. W. Larsen and S. Ma, *J. Am. Chem. Soc.*, 2012, **134**, 13188–13191.
36 G. Cheng, W. Li, L. Ha, X. Han, S. Hao, Y. Wan, Z. Wang, F. Dong, X. Zou, Y. Mao and S.-Y. Zheng, *J. Am. Chem. Soc.*, 2018, **140**, 7282–7291.
37 C. Wang, H. Sun, J. Luan, Q. Jiang, S. Tadepalli, J. J. Morrissey, E. D. Kharasch and S. Singamaneni, *Chem. Mater.*, 2018, **30**, 1291–1300.
38 Y. Zhang, F. Wang, E. Ju, Z. Liu, Z. Chen, J. Ren and X. Qu, *Adv. Funct. Mater.*, 2016, **26**, 6454–6461.
39 N. Jaiswal, C. M. Pandey, S. Solanki, I. Tiwari and B. D. Malhotra, *Microchim. Acta*, 2020, **187**, 1–8.
40 A. A. Ansari and B. D. Malhotra, *Coord. Chem. Rev.*, 2022, **452**, 214282.
41 X. Li, D. Li, Y. Zhang, P. Lv, Q. Feng and Q. Wei, *Nano Energy*, 2020, **68**, 104308.
42 X. Li, Q. Feng, K. Lu, J. Huang, Y. Zhang, Y. Hou, H. Qiao, D. Li and Q. Wei, *Biosens. Bioelectron.*, 2021, **171**, 112690.
43 W. Chen, W. Yang, Y. Lu, W. Zhu and X. Chen, *Anal. Methods*, 2017, **9**, 3213–3220.

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

44 X. Liu, W. Chen, M. Lian, X. Chen, Y. Lu and W. Yang, *J. Electroanal. Chem.*, 2019, **833**, 505–511.
45 C. Gu, L. Bai, L. Pu, P. Gai and F. Li, *Biosens. Bioelectron.*, 2021, **176**, 112907.
46 G. Yuan, L. Wang, D. Mao, F. Wang and J. Zhang, *Microchim. Acta*, 2017, **184**, 3121–3130.
47 M. Rasmussen, S. Abdellaoui and S. D. Minteer, *Biosens. Bioelectron.*, 2016, **76**, 91–102.
48 S. Nara, R. Kandpal, V. Jaiswal, S. Augustine, S. Wahie, J. G. Sharma, R. Takeuchi, S. Takenaka and B. D. Malhotra, *Biosens. Bioelectron.*, 2020, **165**, 112323.
49 A. T. Yahiro, S. M. Lee and D. O. Kimble, *Biochim. Biophys. Acta, Spec. Sect. Biophys. Subj.*, 1964, **88**, 375–383.
50 M. Mon, R. Bruno, S. Sanz-Navarro, C. Negro, J. Ferrando-Soria, L. Bartella, L. Di Donna, M. Prejanò, T. Marino, A. Leyva-Pérez and D. Armentano, *Nat. Commun.*, 2020, **11**(1), 1–9.
51 C. Tudisco, G. Zolubas, B. Seoane, H. R. Zafarani, M. Kazemzad, J. Gascon, P. L. Hagedoorn and L. Rassaei, *RSC Adv.*, 2016, **6**, 108051–108055.
52 L. Zhang, X. Wang, R. Wang and M. Hong, *Chem. Mater.*, 2015, **27**, 7610–7618.
53 S. Dang, Q.-L. Zhu and Q. Xu, *Nat. Rev. Mater.*, 2017, **3**(1), 1–14.
54 B. Y. Xia, Y. Yan, N. Li, H. B. Wu, X. W. Lou and X. Wang, *Nat. Energy*, 2016, **1**(1), 1–8.
55 X. Wang, J. Zhou, H. Fu, W. Li, X. Fan, G. Xin, J. Zheng and X. Li, *J. Mater. Chem. A*, 2014, **2**, 14064–14070.
56 Y.-Z. Chen, C. Wang, Z.-Y. Wu, Y. Xiong, Q. Xu, S.-H. Yu and H.-L. Jiang, *Adv. Mater.*, 2015, **27**, 5010–5016.
57 J. Xi, Y. Xia, Y. Xu, J. Xiao and S. Wang, *Chem. Commun.*, 2015, **51**, 10479–10482.
58 M. Jahan, Z. Liu and K. P. Loh, *Adv. Funct. Mater.*, 2013, **23**, 5363–5372.
59 J. Wei, Y. Hu, Y. Liang, B. Kong, J. Zhang, J. Song, Q. Bao, G. P. Simon, S. P. Jiang and H. Wang, *Adv. Funct. Mater.*, 2015, **25**, 5768–5777.
60 L. Ge, Y. Yang, L. Wang, W. Zhou, R. De Marco, Z. Chen, J. Zou and Z. Zhu, *Carbon*, 2015, **82**, 417–424.
61 V. Lykourinou, Y. Chen, X.-S. Wang, L. Meng, T. Hoang, L.-J. Ming, R. L. Musselman and S. Ma, *J. Am. Chem. Soc.*, 2011, **133**, 10382–10385.
62 A. Sharma, J. Lim, S. Jeong, S. Won, J. Seong, S. Lee, Y. S. Kim, S. B. Baek and M. S. Lah, *Angew. Chem., Int. Ed.*, 2021, **60**, 14334–14338.
63 H. Kitagawa, *Nat. Chem.*, 2009, **1**, 689–690.
64 D.-W. Lim and H. Kitagawa, *Chem. Rev.*, 2020, **120**, 8416–8467.
65 J. A. Hurd, R. Vaidhyanathan, V. Thangadurai, C. I. Ratcliffe, I. L. Moudrakovski and G. K. H. Shimizu, *Nat. Chem.*, 2009, **1**, 705–710.
66 S. Bureekaew, S. Horike, M. Higuchi, M. Mizuno, T. Kawamura, D. Tanaka, N. Yanai and S. Kitagawa, *Nat. Mater.*, 2009, **8**, 831–836.
67 D. Umeyama, S. Horike, M. Inukai, Y. Hijikata and S. Kitagawa, *Angew. Chem., Int. Ed.*, 2011, **50**, 11706–11709.
68 H. Wang, Y. Zhao, Z. Shao, W. Xu, Q. Wu, X. Ding and H. Hou, *ACS Appl. Mater. Interfaces*, 2021, **13**, 7485–7497.
69 S. Wang, M. Wahiduzzaman, L. Davis, A. Tissot, W. Shepard, J. Marrot, C. Martineau-Corcos, D. Hamdane, G. Maurin, S. Devautour-Vinot and C. Serre, *Nat. Commun.*, 2018, **9**, 4937.
70 Y. Guo, Z. Jiang, W. Ying, L. Chen, Y. Liu, X. Wang, Z.-J. Jiang, B. Chen and X. Peng, *Adv. Mater.*, 2017, **30**, 1705155.
71 S. P. Shet, S. Shanmuga Priya, K. Sudhakar and M. Tahir, *Int. J. Hydrogen Energy*, 2021, **46**, 11782–11803.
72 U. Stoeck, I. Senkovska, V. Bon, S. Krause and S. Kaskel, *Chem. Commun.*, 2015, **51**, 1046–1049.
73 A. J. Pommer, S. Cal, A. H. Keeble, D. Walker, S. J. Evans, U. C. Kühlmann, A. Cooper, B. A. Connolly, A. M. Hemmings, G. R. Moore, R. James and C. Kleanthous, *J. Mol. Biol.*, 2001, **314**, 735–749.
74 S.-i. Noro, S. Kitagawa, M. Kondo and K. Seki, *Angew. Chem., Int. Ed.*, 2000, **39**, 2081–2084.
75 M. P. Suh, H. J. Park, T. K. Prasad and D.-W. Lim, *Chem. Rev.*, 2011, **112**, 782–835.
76 N. L. Rosi, J. Eckert, M. Eddaoudi, D. T. Vodak, J. Kim, M. O'Keeffe and O. M. Yaghi, *Science*, 2003, **300**, 1127–1129.
77 J. L. C. Rowsell and O. M. Yaghi, *Angew. Chem., Int. Ed.*, 2005, **44**, 4670–4679.
78 B. Chen, M. Eddaoudi, S. T. Hyde, M. O'Keeffe and O. M. Yaghi, *Science*, 2001, **291**, 1021–1023.
79 H. K. Chae, D. Y. Siberio-Pérez, J. Kim, Y. Go, M. Eddaoudi, A. J. Matzger, M. O'Keeffe and O. M. Yaghi, *Nature*, 2004, **427**, 523–527.
80 B. Chen, M. Eddaoudi, T. M. Reineke, J. W. Kampf, M. O'Keeffe and O. M. Yaghi, *J. Am. Chem. Soc.*, 2000, **122**, 11559–11560.
81 A. P. Côté and G. K. H. Shimizu, *Chem.–Eur. J.*, 2003, **9**, 5361–5370.
82 P. Gai, C. Gu, X. Kong and F. Li, *iScience*, 2020, **23**, 101133.
83 W. P. R. Deleu, I. Stassen, D. Jonckheere, R. Ameloot and D. E. De Vos, *J. Mater. Chem. A*, 2016, **4**, 9519–9525.
84 N. C. Burtch, H. Jasuja and K. S. Walton, *Chem. Rev.*, 2014, **114**, 10575–10612.
85 X. Qian, B. Yadian, R. Wu, Y. Long, K. Zhou, B. Zhu and Y. Huang, *Int. J. Hydrogen Energy*, 2013, **38**, 16710–16715.
86 N. K. Gupta, J. Bae, S. Kim and K. S. Kim, *RSC Adv.*, 2021, **11**, 8951–8962.
87 S. Kim, B. Joarder, J. A. Hurd, J. Zhang, K. W. Dawson, B. S. Gelfand, N. E. Wong and G. K. H. Shimizu, *J. Am. Chem. Soc.*, 2018, **140**, 1077–1082.
88 E. Zhang, Y. Xie, S. Ci, J. Jia and Z. Wen, *Biosens. Bioelectron.*, 2016, **81**, 46–53.
89 N. S. Lopa, M. M. Rahman, F. Ahmed, S. Chandra Sutradhar, T. Ryu and W. Kim, *Electrochim. Acta*, 2018, **274**, 49–56.
90 S. Zhou, L. Jiang, J. Zhang, P. Zhao, M. Yang, D. Huo, X. Luo, C. Shen and C. Hou, *Microchim. Acta*, 2021, **188**, 160.
91 Y. Song, M. Xu, C. Gong, Y. Shen, L. Wang, Y. Xie and L. Wang, *Sens. Actuators, B*, 2018, **257**, 792–799.


© 2022 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 20 July 2022. Downloaded on 8/1/2022 5:11:57 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

92 X. Wang, X. Lu, L. Wu and J. Chen, *Biosens. Bioelectron.*, 2015, **65**, 295–301.
93 H. Wang, Y. Jian, Q. Kong, H. Liu, F. Lan, L. Liang, S. Ge and J. Yu, *Sens. Actuators, B*, 2018, **257**, 561–569.
94 Y. Yu, C. Yu, Y. Niu, J. Chen, Y. Zhao, Y. Zhang, R. Gao and J. He, *Biosens. Bioelectron.*, 2018, **101**, 297–303.
95 T. Q. N. Tran, G. Das and H. H. Yoon, *Sens. Actuators, B*, 2017, **243**, 78–83.
96 F. Wang, X. Chen, L. Chen, J. Yang and Q. Wang, *Mater. Sci. Eng., C*, 2019, **96**, 41–50.
97 P. Zhang, H. Huang, N. Wang, H. Li, D. Shen and H. Ma, *Microchim. Acta*, 2017, **184**, 4037–4045.
98 C. Li, R. Wu, J. Zou, T. Zhang, S. Zhang, Z. Zhang, X. Hu, Y. Yan and X. Ling, *Biosens. Bioelectron.*, 2018, **116**, 81–88.
99 D. Arif, Z. Hussain, M. Sohail, M. A. Liaqat, M. A. Khan and T. Noor, *Front. Chem.*, 2020, **8**, 573510.
100 W. Meng, Y. Wen, L. Dai, Z. He and L. Wang, *Sens. Actuators, B*, 2018, **260**, 852–860.
101 Y. Yin, C. Gao, Q. Xiao, G. Lin, Z. Lin, Z. Cai and H. Yang, *ACS Appl. Mater. Interfaces*, 2016, **8**, 29052–29061.
102 Z. Kang, K. Jiao, R. Peng, Z. Hu and S. Jiao, *RSC Adv.*, 2017, **7**, 11872–11879.
103 F. Zhang, X. Wu, J. Gao, Y. Chen, Y. Gui, L. Zhang, W. Gan and Q. Yuan, *Electrochim. Acta*, 2020, **340**, 135958.
104 T. Yimamumaimaiti, X. Lu, J.-R. Zhang, L. Wang and J.-J. Zhu, *ACS Appl. Mater. Interfaces*, 2020, **12**, 41429–41436.
105 S. Cui, S. Gu, Y. Ding, J. Zhang, Z. Zhang and Z. Hu, *Talanta*, 2018, **178**, 788–795.
106 M. Rafti, W. A. Marmisollé and O. Azzaroni, *Adv. Mater. Interfaces*, 2016, **3**, 1600047.
107 S. Patra, S. Sene, C. Mousty, C. Serre, A. Chaussé, L. Legrand and N. Steunou, *ACS Appl. Mater. Interfaces*, 2016, **8**, 20012–20022.
108 J. Mao, L. Yang, P. Yu, X. Wei and L. Mao, *Electrochem. Comm.*, 2012, **19**, 29–31.

THE JOURNAL OF DIGITAL FORENSICS, SECURITY AND LAW

Journal of Digital Forensics, Security and Law

Volume 17 Article 2

7-12-2022

# Assessment of 3D mesh watermarking techniques

Neha Sharma
*Delhi Technological University*, nehashrm013@gmail.com

Jeebananda Panda
*Delhi Technological University*, jpanda@dce.ac.in

Follow this and additional works at: https://commons.erau.edu/jdfsl

Part of the Information Security Commons

## Recommended Citation

Sharma, Neha and Panda, Jeebananda (2022) "Assessment of 3D mesh watermarking techniques," *Journal of Digital Forensics, Security and Law*: Vol. 17 , Article 2.
Available at: https://commons.erau.edu/jdfsl/vol17/iss2/2

This Article is brought to you for free and open access by the Journals at Scholarly Commons. It has been accepted for inclusion in Journal of Digital Forensics, Security and Law by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu.

(c)ADFSL

# ASSESSMENT OF 3D MESH WATERMARKING TECHNIQUES

Neha Sharma, Jeebananda Panda
Delhi Technological University
Department of Electronics and Communication Engineering
New Delhi, Delhi 110042, India
nehashrm013@gmail.com
jpanda@dce.ac.in

## ABSTRACT

With the increasing usage of three-dimensional meshes in Computer-Aided Design (CAD), medical imaging, and entertainment fields like virtual reality, etc., the authentication problems and awareness of intellectual property protection have risen over the last decade. Numerous watermarking schemes have been suggested to protect ownership and prevent the threat of data piracy. This paper begins with the potential difficulties that arose when dealing with three-dimension entities in comparison to two-dimensional entities and also lists possible algorithms suggested hitherto and their comprehensive analysis. Attacks, also play a crucial role in deciding on a watermarking algorithm so an attack based analysis is also presented to analyze the resilience of watermarking algorithms under several attacks. In the end, some evaluation measures and potential solutions are brooded over to design robust and oblivious watermarking schemes in the future.

**Keywords**: 3D Mesh, Watermarking, vertices, review

## 1. INTRODUCTION

Due to the advent of state-of-the-art technologies for storage, dissemination, and transmission of data (audio, video, images, etc.) over the last few decades, the threat of tampering, unauthorized duplication, and unrestrained propagation of data is on the rise. Intellectual property rights protection demands security, rightful ownership, and prohibition of unauthorized access to data. Digital watermarking seems a potential solution for this when compared to classical methods of data protection as former one secures data even after its transmission phase is over and accessibility to the data is procured (Borah & Borah, 2018).

The origin of the name "Watermark" dates back to the time when its existence was only seen on paper. In order to create a watermark, the width of the paper was modified when the paper was in wet condition and thus a shadowed or a lighter pattern was developed in the watermarked paper. Therefore, this process gets its name as a watermark. With the upsurge of digital content, this term came to be used with digital practices. The coinage of the term "Digital Watermark" was done by Andrew Tirkel and Charles Osborne in December 1992. Digital watermark shares

© 2022 JDFSL 

a similar objective of physical watermark i.e. to verify the authenticity and integrity of the content as well as its ownership. But unlike a physical copy, a digital copy of data is identical to the original data so a digital watermark can only mark the data and cannot control its access. So it is a passive tool for content protection (Van Schyndel et al., 1994).

In digital watermarking, the embedding of some crucial information relating to either the owner or the content itself is done in such a way that its presence can't be avoided even after several manipulations of the digital media whether intentional or unintentional. The general watermarking system as shown in Figure 1 involves the host media which is to be watermarked undergoing an embedding process first. After being watermarked, media is distributed through channels wherein it bears all genuine and malicious attacks before reaching the final destination. At the receiver also, there has to be some mechanism to detect or retrieve the watermark to authenticate the content or verify ownership respectively (Potdar et al., 2005).

Similar to watermarking. there exist several other popular techniques like cryptography and steganography. Cryptography is usually involved in maintaining the secrecy of the information using a secret key and thus prohibits its reading by any illegitimate user. On the other hand, steganography makes use of cover media to hide information in such a way that it is hardly noticeable to the human eye. But both these techniques have no business in dealing with copyright issues and authentication of data; their prime concern is the protection of information content (Saini & Shrivastava, 2014).

We may classify digital watermarking into different categories based on the type of watermark, its robustness, application, host data and other features. The complete taxonomy of digital watermarking is shown in Table 1.

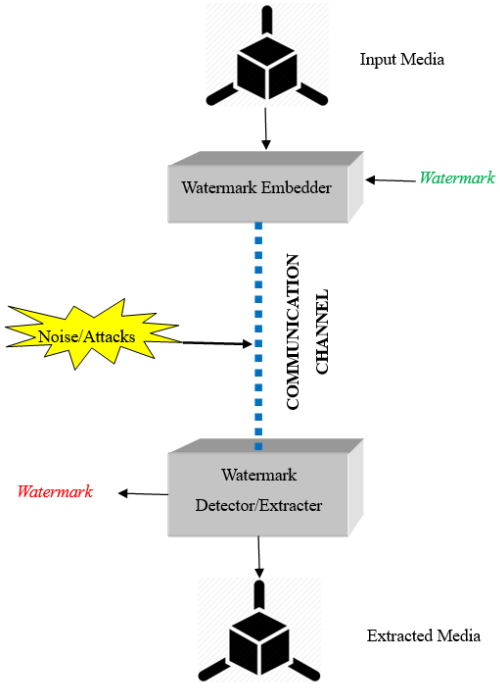



Figure 1. The digital watermarking system

On the basis of the type of watermark to be embedded, two categories are defined, the first is noise type, other is image type. In noise type, any pseudo, chaotic, random or Gaussian sequence can be embedded in the host media whereas, in image type watermark, any logo, stamp or binary image can be inserted in the media as a watermark.

Depending on the ability of the watermark to remain unaffected by modifications in the embedding media, three categories of watermark are defined, namely fragile, semi-fragile and robust. If watermarking is stubborn enough to withstand all modifications whether intentional or unintentional and still be recovered, then watermarking algorithm may be designated as a robust one. Usually, robust algorithms are used for copyright purposes so that it protects embedded copyright information to prevent its illegitimate usage. On the other hand, fragile techniques are used to detect even the slightest modification in the watermark to assess the amount


© 2022 JDFSL

Table 1. Complete Taxonomy of Watermarking System

| S.No | Basis | Categories |
|---|---|---|
| 1 | Type of watermark | Noise, Image |
| 2 | Robustness | Fragile, Semi-fragile, Robust |
| 3 | Domain of watermarking | Spatial, Spectral, Invariant |
| 4 | Imperceptibility | Visible, Invisible, Dual Watermark |
| 5 | Embedding Media | Image, Text, Audio, Video, 3D mesh |
| 6 | Extraction method | Blind, Semi-blind, non-blind |
| 7 | Application | Copyright, Tampering, Anti-counterfeit, Hidden annotation |
| 8 | Key usage | Asymmetric, Symmetric |
| 9 | Information | Informative, Non-informative |

of tampering done. Some of the pioneering work such as R. Ohbuchi et al. (2004) realised that semi-fragile serves the characteristics of both, it is able to resist some degree of perturbations such as due to compression while still being fragile.

We may also choose the domain of embedding watermark such as spatial, spectral or invariant. If watermarking technique directly plays with the host media in the original domain, it is called spatial, and if frequency domain coefficients are altered, it is called spectral. Liu and Chen (2010) identified that it is desired to choose an invariant transform domain that provides robustness to withstand attacks..

We have seen logos of TV channels, stamps in the paper, etc which is conspicuous enough to assert ownership. Such overt labelling of channels, papers, or any other media comes under the category of visible watermarking. But at times, covert protection of ownership details is required without causing a perceptible change in the host media. This category of watermarking comes under an invisible one.

The extraction method of the watermark also varies as per the requirement of original data at the receiver's side. The non-blind technique requires original data for watermark retrieval and the process is much simple here while the blind technique forbids the usage of original data at the receiver's side which makes it complicated due to lack of prior information.

There are different applications of watermarking anti-counterfeiting, tampering, copyright and annotation hiding to name a few. Different applications demand different characteristics of the watermark, for instance, annotations used in confidential files need to be hidden from illegitimate users, therefore the watermark here should be invisible, imperceptible as well as robust to safeguard them.

Recipients of watermarked data are provided with a key to access the data. If this key is identical to the key used while embedding the watermark, such a technique is called symmetric watermarking. On the other hand, if both the keys are different, such a technique is called asymmetric watermarking. Watermark may contain some information indicating the ownership details or it may also be non-informative wherein the receiver's task is just to ascertain the presence or absence of a watermark(Saini & Shrivastava, 2014).

Watermark can be embedded in different types of data for instance text, audio, video 3D graphics etc. and thus is classified according to the domain of embedded media.

© 2022 JDFSL

Of late, extensive usage of three-dimensional mesh in the medical, industrial, and entertainment industries can be observed that has propelled the need for its authentication check to curb the ill effects of its malicious usage. However, a relatively small body of literature exists for them when compared to their two-dimensional counterpart like images. This can be attributed to several potential challenges faced while designing algorithms for these models that are discussed as follows:

- There is a lack of unique arrangement for 3D mesh models, unlike images that are statically represented by rows and columns. This rearrangement leads to synchronization problems during the extraction phase.

- Number of vertices available in a 3D mesh is relatively less when compared to images where millions of pixels are available. So comparatively very less data is available for watermark insertion in the 3D mesh. Hence this poses another challenge in making watermarking distortion-free as well as robust under various attacks.

- Several distortion and distortion-less attacks also pose a threat to watermarked model by destroying the watermark. Similarity transforms, smoothing, noise addition, etc. are some of the distortion attacks which cause visual alteration of the model. Interestingly, some of the prominent works (Ai et al., 2009; J.-W. Cho et al., 2006) highlighted vertex reordering as one of the examples of a distortion-less attack that does not cause any perceptible change in the model.

This paper gives a comprehensive review of the watermarking techniques proposed so far in the last two decades. The organization of the paper is as follows: Section 2 briefly introduces relevant preliminaries for three-dimensional mesh comprehension, Section 3 reviews the techniques proposed for 3D mesh watermarking categorized into different subsections based on insertion domain or type of algorithm used. Moreover, since resilience to attacks plays a vital role in designing a watermarking algorithm, Section 4 gives an attack-based analysis of them. Some evaluation measures to compare watermarking are discussed in Section 5 and lastly, section 6 concludes the review on a promising note.

# 2. PRELIMINARIES

The representation of 3D mesh usually takes the form of polygonal or triangular meshes in computer graphics or Computer-Aided Design(CAD). Several preliminaries are required to be constituted before venturing into the domain of 3D mesh watermarking. A 3D mesh is constituted by three different elements: vertices, edges, and faces. An example of a 3D mesh model is shown in Figure 2. Coordinates of vertices in a Cartesian plane provide geometrical information about the model while edges connecting vertices to form polygonal faces provide the connectivity details of a 3D mesh.

1-ring neighbourhood of a vertex $v_i$ constitutes all the vertices that are adjacent to vertex $v_i$ in a neighbourhood. The degree of vertex $v_i$ is defined as the total number of vertices that are directly connected to it in a 1-ring neighbourhood. It is also referred to as the valence of a vertex. An example of vertices of valences 5 and 6 is shown in Figure 3. The valency of a vertex can be studied from the works by Borah and Borah (2018); El Zein et al. (2017); Ghali (2016); Soliman et al. (2013); Motwani and Harris Jr (2009) to get in-depth analysis of the 3D mesh attributes.


© 2022 JDFSL

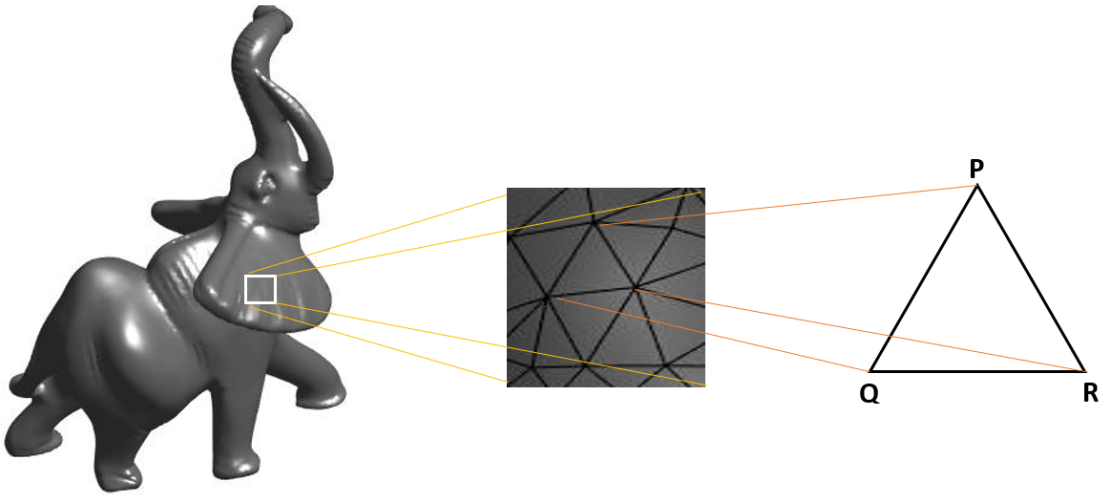



Figure 2. An example of a 3D mesh gives a closer look at the building blocks of mesh namely vertices, edges and faces. In a zoomed-in portion of the triangular mesh having vertices P, Q and R, the triangle PQR is one of the faces formed by joining edges PQ, QR and RP respectively.

# 3. LITERATURE REVIEW

Several past pieces of research on 3D mesh watermarking have been examined in this study. A specific searching standard utilized for a selection of peer-reviewed journals in this work is based on the following keywords: 3D mesh watermarking, digital watermarking, copyright protection, robust watermarking, etc. Prominent electronic research databases like Elsevier, Springer, IEEE, IET, etc. have been adopted in this survey.

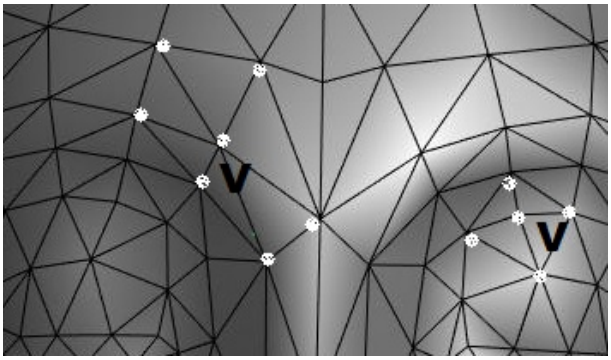



Figure 3. Example showing vertices of degree-6 and degree-5

## 3.1 Watermarking techniques based on geometrical transformations

Most of the initial watermarking techniques involved changing the coordinates of vertices of the 3D mesh model in accordance with the watermark directly. Such techniques generally come under the category of fragile watermarking. Yu et al. (2003) modified the length of vertices from the centre of gravity of the mesh to insert a watermark. Vertices are divided into groups and each group becomes a carrier of a 1-bit watermark as per the modified distance from the gravity centre. But this technique requires registration and re-sampling as a pre-processing step at the extraction site for robustness as this approach is not resilient to similarity transformations.

One of the most important entities of the 3D mesh model for watermark insertion is facets. R. Ohbuchi et al. (1998) proposed a blind watermarking approach using interest-

© 2022 JDFSL

ing characteristics of a facet in a 3D mesh-like Triangle Similarity Quadruple and tetrahedral volume ratio. The algorithm basically uses the ratio of the triangle's height and opposite edge length which is resilient to many transformations like rotation, scaling, and translation but not robust enough to other severe forms of attacks.

Such techniques employ modifications in the geometry of the model to insert a watermark which is simple to implement but their robustness quotient is on the lower side. Moreover, they also suffer from synchronization issues so even in absence of attacks watermark values seem perturbed.

## 3.2 Watermarking techniques based on direct spectral analysis

For other entities like image, audio, etc. it has been observed through spectral analysis that affecting low or middle-frequency coefficients does not distort the final visual output much. This characteristic has been exploited much in the case of image watermarking but in the case of 3D mesh, no such analysis has been made so far. So in such algorithms basic principle of watermarking revolves around modifying spectral coefficients. Different kinds of basis functions are utilized to analyze 3D mesh spectrally. In the laplacian basis function, a square dimension matrix of length equal to a number of vertices in a 3D mesh is made based on the connectivity order of the model. Suppose the number of vertices in a model is M then to determine M x 3 spectral coefficients, projections of the three coordinates vectors of all the vertices on M ordered and normalized eigenvectors of the Laplacian matrix are taken. . R. Ohbuchi et al. (2002) made use of this analysis to watermark 3D mesh by modifying low and middle-frequency coefficients. Cayre et al. (2003) improvised a step further by quantizing those frequency coefficients which will have little or no impact on the final watermarked model. However, the computation time in this work is directly in proportion with the mesh complexity due to the diagonalization of the NxN laplacian matrix. Moreover, the requirement of exact connectivity information for this algorithm to work makes it non-blind. Murotani and Sugihara (2003) used a lower dimension matrix to increase performance to computation time ratio.

## 3.3 Watermarking techniques based on Multi-Resolution analysis

The multi-resolution analysis serves as a perfect tool in order to strike the balance between the capacity of the available resources and the complexity order of the mesh. It gives a coarse (low frequencies) to fine (median and high frequencies) representation of the mesh from basic shape to intricate details at varying levels of resolution. This representation serves different purposes required for different applications, for instance, embedding in low resolution (low frequencies) provides robustness as well as imperceptibility whereas embedding in high-resolution details provides for high capacity.

Wavelets are generally used for such multi-resolution analysis where watermark strings can be inserted in wavelet coefficients at different resolution levels. Under the assumption of statistical independence between wavelet coefficient norms and watermarking bits, Uccheddu et al. (2004) proposed an oblivious watermarking scheme using wavelet-based multiresolution analysis. W.-H. Cho et al. (2004); Jian-qiu et al. (2004) proposed a fragile watermarking technique using the same approach for authentication purposes. The later one made use of both spherical domain wavelet coefficients and coarsest representation to insert a watermark. To make


© 2022 JDFSL

the above approach robust and resilient to common mesh operations like rotation, translation, scaling, etc., and malicious attacks, several schemes were introduced. Yin et al. (2001) used Guskov's multiresolution signal processing method and a Burt-Adelson pyramid is built for mesh for watermark insertion. Praun et al. (1999) extended the spread spectrum approach by giving a solution for the dearth of natural parameterization for decomposition based on frequency.

Multiresolution techniques much in resonance with direct spectral analysis also lack sufficient robustness. Moreover, to ensure robustness, the requirement of registration and resampling at the extraction site by most of the techniques renders them non-blind or at least semi-blind.

## 3.4 Watermarking techniques based on Parameterization

Another interesting approach for watermarking meshes apart from direct spectral analysis and multi-resolution analysis is Parameterization. This approach facilitates the application of existing 2D algorithms on meshes by converting 3D mesh into 2D entities. A semi-blind method was proposed by L. Li et al. (2004) which modified harmonic coefficients of mesh as per the watermark after transforming the original mesh into a spherical parameterization domain.

## 3.5 Watermarking techniques based on Vertex norm distribution

J.-W. Cho et al. (2006) proposed a robust watermarking approach to insert a watermark in 3D mesh without comprising the imperceptibility requirement of the watermark by making use of statistical parameters of vertex norm distribution. In this technique, firstly Cartesian coordinates are transformed into spherical coordinates and only vertex norm values are considered for watermark while the other two spherical coordinates are kept intact so there is less visible distortion on the model. This method though robust but fails in the case of very small-sized models or CAD models which have too flat surfaces. Hu et al. (2009) used constrained optimization to minimize the least square difference between the original model and the watermarked one and proved to be better than Cho's method in resisting Gaussian noise. But this method has its own limitations as with the increasing size of the model, the computational complexity of quadratic programming problems also increases. Levenberg– Marquardt optimization method was also suggested Bors and Luo (2012) for reducing surface distortion to a minimum as well as preserving the original model. This method enhanced robustness against various common mesh operations. Methods like improved vertex grouping were suggested by works like S. Li et al. (2017) to ensure transparency while improving robustness.

## 3.6 Watermarking techniques based on the intelligent selection of watermark carriers

In the last few years, an intelligent watermark embedding technique has been prevalent that involves the selection of watermark carriers suitable for carrying watermark. Their suitability is assessed based on three crucial tenants of digital watermarking namely imperceptibility, capacity, and robustness of the watermarking process. The most common criteria employed for watermark selection in 3D models is their curvature. El Zein et al. (2017) made use of the Fuzzy C-means clustering technique to find vertices that will cause the least perceptible distortion to the 3D model after carrying watermark bits but the computational complexity of Fuzzy C-

© 2022 JDFSL 

means slowed down the process. Soliman et al. (2013) employed another clustering technique namely Self Organising Maps(SOM) to select vertices after calculating the feature vector of the 1-ring neighbourhood of a vertex. Although this method results in better robustness but both the approaches discussed in this method were either non-blind or semi-blind causing additional payload in the watermark extraction process. S. Li et al. (2017) rejected marginal bins in the histogram of eigenvalues of the vertex by 10% and employed the rest of the bins as watermark carriers, however even this method could not improve robustness much in cases of flat-surfaced models like CAD models and small-sized models having fewer vertices. Sharma and Panda (2020) introduced a viable approach wherein computational overheads are reduced by employing histogram binning of feature vectors based on their surface inclination while not compromising the robustness of the watermarking technique as it involves statistical insertion in chosen watermark carriers.

# 4. ATTACK-BASED ANALYSIS

Attacks also form an inevitable factor when designing any watermarking technique. Without showing resilience to different types of attacks, watermarking algorithms may not be successful; therefore robustness towards these attacks becomes an important parameter for the evaluation of any watermarking technique.

## 4.1 Geometric Attacks

As the name itself suggests, these kinds of attacks reflect changes in the geometrical part of the model by perturbing the vertex's positions. Some common mesh operations like rotation, translation, and scaling together known as Similarity transformation fall into this category. Not just robust even fragile watermark is expected to withstand such operations. To make watermarking technique immune to these attacks, one of the suggested solutions in many papers reviewed above is to choose a space that is invariant to such changes in vertices positions. In fact, such space can be created by shifting the origin to the mesh's centre of gravity, uniform scaling bounded by a unit cube/sphere, and rotating towards principal axes of rotation; thus making it translation, scaling, and rotation invariant. Some of the features do change after watermark insertion but if some feature or characteristics of the embedding domain is known, successful extraction is possible but at the cost of making the algorithm more or less semi-blind.

Another solution suggested in the papers above was to use such ratios for instance between edge length and height of triangle of mesh which does not vary with watermark insertion or attacks as suggested by R. Ohbuchi et al. (1998) TSQ and TVR. Several signal processing attacks like random noise addition, smoothing, etc. also come under this category. Such operations can be dealt with nicely in spectral-domain with either direct frequency analysis pr multi-resolution analysis where it is found that low and medium frequency part is more resilient to such attacks and thus can be used for insertion of watermark as discussed in previous sections.

## 4.2 Connectivity Attacks

A common type of operation that comes under this category is cropping. Although it can be regarded as a geometric attack its consequences can be far worse as some part of the watermark string may be lost and become unrecoverable after an attack. A potential solution to this problem is a repetition of the watermark string in different patches of the mesh model so that the watermark string is recoverable even after the attack. Spatial techniques discussed before


© 2022 JDFSL

are less immune to these attacks due to geometric changes caused by cropping as well as loss of synchronization caused due to it. As far as spectral-domain-based techniques are considered whether direct frequency analysis based or multi-resolution analysis based, information about the vertices order or mesh connectivity or both are required at times, making these techniques also less immune to connectivity attacks.

Some other potential solutions in order to deal with connectivity attacks could be finding a domain for watermarking which is invariant to connectivity information as well as robust to connectivity attacks.

### 4.3 File format attacks

File attack includes the rearrangement of the vertices or faces in the mesh model. Format attack includes conversion of 3d model file type which may alter the structure of mesh and thus causing a change in the order of mesh vertices and faces. Watermarking schemes need to be synchronization independent to deal with such attacks. The process used for synchronization should not depend on the vertices and face order of the mesh.

## 5. EVALUATION PARAMETERS FOR ALGORITHM ASSESSMENT

A watermarked model must endure all attacks while preserving the model as well as the watermark. To show such a characteristic a watermarking scheme should possess robustness, imperceptibility as well as capacity which are often in contradiction with each other. For example, employing repetition of the watermark in different patches ensure robustness but decreases capacity. The visual quality of the watermark is also at risk if greater transformations are employed in the model to ensure robustness.

A benchmarking system is required for the evaluation of robust watermarking schemes to analyze the robustness of these schemes to different attacks and to assess their visual quality degradation after watermarking. Wang et al. (2010) provides a software tool for the same and also discusses parameters like MRMS (Maximum root mean square error) to measure the visual difference between the original model and the watermarked model. The endurance of the watermark inserted can be measured through correlation as shown by J.-W. Cho et al. (2006). Furthermore, the imperceptibility of the watermark can be evaluated through VSNR as in El Zein et al. (2017). An overall comparison of 3D watermarking techniques has been presented in Table 2.

© 2022 JDFSL

Table 2. Comparative Analysis of 3D mesh watermarking techniques

| **Watermarking techniques based on** | **Techniques proposed in** | **Algorithm features** | **Attack-centric investigation** | | |
|---|---|---|---|---|---|
| | | | **Geometric Attack** | **Connectivity Attack** | **File format Attack** |
| **Geometrical transformations** | Yu et al. (2003) Ohbuchi et al. (2004) | Simple to implement Not robust enough, Also suffer synchronisation issues. | Invariant | Variant | Variant |
| **Direct spectral analysis** | Ohbuchi et al. (2002), Cayre et al. (2003) Murotani et al. (2003) | Robust and imperceptible Increased complexity and computation time. | Invariant | Variant | Variant |
| **Multi-Resolution Analysis** | Uccheddu et al. (2004) Cho et al. (2005) Jin et al.,2004) Yin et al. (2001) Praun et al. (1999) | Increased robustness and capacity. Requires resampling and registration | Invariant | Variant | Variant |
| **Parameterization** | Li et al. (2004) | Allows application of existing 2D algorithms on 3D mesh Semi-blind Approach. | Invariant | Semi-invariant | Variant |
| **Vertex norm distribution** | Cho et al. (2007), Bors et al. (2012) | Transparency as well as robustness increased. However, transparency is difficult to maintain in case of small sized models and complexity increases in case of large-sized models. | Invariant | Semi-invariant | Invariant |
| **Intelligent watermark selection** | Zein et al. (2016) Solimon et al. (2013) Li et al. (2017) Sharma et al. (2020) | Only selected carriers are chosen for watermarking thus causing minimum perceptible distortion in the model. Less prone to error in watermark extraction process as fewer vertices are involved. | Invariant | Mostly invariant | Invariant |

© 2022 JDFSL

# 6. DISCUSSIONS AND POTENTIAL SOLUTIONS

## 6.1 Capacity, Robustness and Imperceptibility Trade-off

If a watermarking algorithm is designed to be robust under various attacks, then it is observed that the quality of watermarked mesh usually degrades. This visual degradation mortification of the model after watermarking decreases the imperceptibility of the model. Similarly, one of the methods to increase the robustness of the watermark is redundant insertion so that even after several malicious attacks whether intended or not, a significant portion of watermarking bits are preserved. However, such a redundant insertion deteriorates the capacity of the model to carry a watermark as presented in significant works by Sharma and Panda (2020); Wang et al. (2010). The potential solution as suggested by contemporary literature lies in finding an insertion domain or a feature or shape descriptor in the 3D mesh which is invariant to common mesh operations and attacks. The embedding feature or domain is required to be chosen intelligently to ensure the imperceptibility of the watermarked mesh.

## 6.2 Computational Complexity and Execution Time Trade-off

Although non-blind watermarking methods have proven to be more robust than their blind counterparts, the requirement of the original model at the extraction site leads to an excessive payload. Hence the future research proposals are inclined towards a blind framework of watermarking to avoid the expenses and hassle of sending the original model along with the watermarked model. However, the blind watermarking retrieval process is more complex than the non-blind or semi-blind methods due to the absence of the original model (Praun et al., 1999). Some of the relevant works such as by J.-W. Cho et al. (2006), assessed that blind approaches cannot simply rely on comparison with the original mesh to retrieve the watermark rather they require some other information be its the length of the watermark, location of the watermark, etc. which serves as the key for watermark extraction purposes. These approaches require better comprehension of the mesh model characteristics therefore the need for algorithms of complex nature comes into play. The higher the complexity, the higher is the correlation between the true watermark and extracted watermark. Nonetheless, higher complexity also leads to an increase in execution time which is not a desirable characteristic of watermarking process. A potential solution lies in reaching a compromise between the fast execution time of the algorithm and the high correlation between the original and extracted watermark such that optimum efficiency is gained in the watermarking process.

# 7. CONCLUSION

This paper discusses some of the breakthrough research contributions in the 3D mesh watermarking domain in the last two decades. However, owing to many difficulties in constructing an algorithm for watermark insertion in a 3D model as discussed in this paper, the research in this area can still be considered in its infancy stage. Several issues are hampering the utilization of watermarking in 3d mesh at full scale. There is still uncertainty regarding what possible fusion of a variety of attacks, a model may have to undergo in myriad instances. Finding a feasible approach to tackle them without compromising the watermark and the model, pose a big

© 2022 JDFSL 

challenge. Future directions of research could focus on studying hybrid attacks rather than individual attacks and their solutions. Furthermore, there is huge research scope in discovering a new insertion domain, transform domain or a novel approach that will facilitate the optimum requirement of robustness, imperceptibility, and capacity concurrently.

# REFERENCES


Ai, Q., Liu, Q., Zhou, Z., Yang, L., & Xie, S. (2009). A new digital watermarking scheme for 3d triangular mesh models. *Signal Processing*, *89*(11), 2159–2170.

Borah, S., & Borah, B. (2018). Watermarking techniques for three dimensional (3d) mesh authentication in spatial domain. *3D Research*, *9*(3), 1–22.

Bors, A. G., & Luo, M. (2012). Optimized 3d watermarking for minimal surface distortion. *IEEE Transactions on Image Processing*, *22*(5), 1822–1835.

Cayre, F., Rondao-Alface, P., Schmitt, F., Macq, B., & Maıtre, H. (2003). Application of spectral decomposition to compression and watermarking of 3d triangle mesh geometry. *Signal Processing: Image Communication*, *18*(4), 309–319.

Cho, J.-W., Prost, R., & Jung, H.-Y. (2006). An oblivious watermarking for 3-d polygonal meshes using distribution of vertex norms. *IEEE Transactions on Signal Processing*, *55*(1), 142–155.

Cho, W.-H., Lee, M.-E., Lim, H., & Park, S.-Y. (2004). Watermarking technique for authentication of 3-d polygonal meshes. In *International workshop on digital watermarking* (pp. 259–270).

El Zein, O. M., El Bakrawy, L. M., & Ghali, N. I. (2017). A robust 3d mesh watermarking algorithm utilizing fuzzy c-means clustering. *Future Computing and Informatics Journal*, *2*(2), 148–156.

Ghali, N. I. (2016). A non-blind robust watermarking approach for 3d mesh models.

Hu, R., Rondao-Alface, P., & Macq, B. (2009). Constrained optimisation of 3d polygonal mesh watermarking by quadratic programming. In *2009 ieee international conference on acoustics, speech and signal processing* (pp. 1501–1504).

Jian-qiu, J., Min-ya, D., Hu-jun, B., & Qun-sheng, P. (2004). Watermarking on 3d mesh based on spherical wavelet transform. *Journal of Zhejiang University-Science A*, *5*(3), 251–258.

Li, L., Zhang, D., Pan, Z., Shi, J., Zhou, K., & Ye, K. (2004). Watermarking 3d mesh by spherical parameterization. *Computers & Graphics*, *28*(6), 981–989.

Li, S., Ni, R., & Zhao, Y. (2017). A 3d mesh watermarking based on improved vertex grouping and piecewise mapping function. *J Inf Hiding Multimedia Signal Process January*, *8*(1), 97e108.

Liu, C.-C., & Chen, J.-Y. (2010). A watermarking scheme for 3d models using haar discrete wavelet transform. In *2010 international symposium on computer, communication, control and automation (3ca)* (Vol. 1, pp. 244–247).

Motwani, R. C., & Harris Jr, F. C. (2009). Robust 3d watermarking using vertex smoothness measure. In *Ipcv* (pp. 287–293).

Murotani, K., & Sugihara, K. (2003). Watermarking 3d polygonal meshes using the singular spectrum analysis. In *Mathematics of surfaces* (pp. 85–98). Springer.

Ohbuchi, R., Masuda, H., & Aono, M.



© 2022 JDFSL

(1998). Watermarking three-dimensional polygonal models through geometric and topological modifications. *IEEE Journal on selected areas in communications*, *16*(4), 551–560.

Ohbuchi, . R., Mukaiyama, . A., & Takahashi, . S. (2002). A frequency-domain approach to watermarking 3d shapes. In *Computer graphics forum* (Vol. 21, pp. 373–382).

Ohbuchi, R., Mukaiyama, A., & Takahashi, S. (2004). Watermarking a 3d shape model defined as a point set. In *2004 international conference on cyberworlds* (pp. 392–399).

Potdar, V. M., Han, S., & Chang, E. (2005). A survey of digital image watermarking techniques. In *Indin'05. 2005 3rd ieee international conference on industrial informatics, 2005.* (pp. 709–716).

Praun, E., Hoppe, H., & Finkelstein, A. (1999). Robust mesh watermarking. In *Proceedings of the 26th annual conference on computer graphics and interactive techniques* (pp. 49–56).

Saini, L. K., & Shrivastava, V. (2014). A survey of digital watermarking techniques and its applications. *arXiv preprint arXiv:1407.4735*.

Sharma, N., & Panda, J. (2020). Statistical watermarking approach for 3d mesh using local curvature estimation. *IET Information Security*, *14*(6), 745–753.

Soliman, M. M., Hassanien, A. E., & Onsi, H. M. (2013). Robust watermarking approach for 3d triangular mesh using self organization map. In *2013 8th international conference on computer engineering & systems (icces)* (pp. 99–104).

Uccheddu, F., Corsini, M., & Barni, M. (2004). Wavelet-based blind watermarking of 3d models. In *Proceedings of the 2004 workshop on multimedia and security* (pp. 143–154).

Van Schyndel, R. G., Tirkel, A. Z., & Osborne, C. F. (1994). A digital watermark. In *Proceedings of 1st international conference on image processing* (Vol. 2, pp. 86–90).

Wang, K., Lavoué, G., Denis, F., Baskurt, A., & He, X. (2010). A benchmark for 3d mesh watermarking. In *2010 shape modeling international conference* (pp. 231–235).

Yin, K., Pan, Z., Shi, J., & Zhang, D. (2001). Robust mesh watermarking based on multiresolution processing. *Computers & graphics*, *25*(3), 409–420.

Yu, Z., Ip, H. H., & Kwok, L. (2003). A robust watermarking scheme for 3d triangular mesh models. *Pattern recognition*, *36*(11), 2603–2614.

© 2022 JDFSL 

*macromol*

MDPI

*Review*

# Assessment of Bio-Based Polyurethanes: Perspective on Applications and Bio-Degradation

**Raminder Kaur [1,*], Pooja Singh [2], Surya Tanwar [2], Gunjan Varshney [2] and Sarla Yadav [2]**

[1] Department of Polymer Science and Chemical Technology, Delhi Technological University (Formerly Delhi College of Engineering), Delhi 110042, India
[2] Department of Applied Chemistry and Polymer Technology, Delhi Technological University (Formerly Delhi College of Engineering), Delhi 110042, India; poojasingh233dew@gmail.com (P.S.); surya3tanwar@gmail.com (S.T.); varshneygunjan111@gmail.com (G.V.); sarlakhenta@gmail.com (S.Y.)
* Correspondence: rkwaliadtu@gmail.com

**Abstract:** Among numerous synthetic macromolecules, polyurethane in its different forms has proven its sheer dominance and established a reputation as a reliable and trusted material due to its proficiency in terms of superior properties, which include: high mechanical strength and abrasion resistance, good durability, good adhesion, good thermal stability, excellent chemical and weathering resistance. Synthetic polyurethane materials are non-biodegradable, poisonous, and use petrochemical-based raw materials, which are now depleting, leading to a surge in polyurethane production costs. Bio-based polyurethanes (PU) have been synthesized by researchers in recent decades and have mostly overtaken petrochemical-based PU in terms of challenges such as solid pollution, economic effectiveness, and availability of raw materials. Enormous kinds of available bio-renewable sources as predecessors for the production of polyols and isocyanates have been explored for the development of "greener" PU materials; these bio-based polyurethanes have significant potential to be used as future PU products, with a partial or total replacement of petroleum-based polyurethanes, due to increasing concern about the environment, their relatively low cost and biodegradability. This critical review concentrates on the possibilities of renewable sources to be used for polyurethane production and gives a clear perspective on the journey, utilization, and recent advancements in the field of different bio-based polyurethane polymers that have arisen over the last decade.

**Keywords:** polyurethanes; non-isocyanate polyurethanes; adhesives; coatings; recycling; biodegradation

check for updates

**Citation:** Kaur, R.; Singh, P.; Tanwar, S.; Varshney, G.; Yadav, S. Assessment of Bio-Based Polyurethanes: Perspective on Applications and Bio-Degradation. *Macromol* **2022**, *2*, 284–314. https://doi.org/10.3390/macromol2030019

Academic Editor: Andrea Sorrentino

Received: 26 May 2022
Accepted: 29 June 2022
Published: 4 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Polyurethanes (PU) represent nearly 8% of the total plastics production and stand tall as the 6th most widely used polymer in the world, with total production reaching 18 MTon in 2016 [1] and still progressing. The superior properties of PU, such as flexibility, corrosion, resistance to chemicals, high automatic strength, low-temperature comfortability, adhesiveness, and chemical structure adaptability, etc. account for their considerable popularity among researchers. PU have urethane groups as major repeating units along their main backbone chains, which are derived by reacting hydroxyl (-OH) terminated chemical entities, with isocyanate (-NCO) terminated compounds, as represented by a generalized scheme shown in Figure 1. In addition, some other functional groups may also be present in the end constitutions of the PU chains, including ethers, esters, urea, biuret, and aromatic moieties [2].

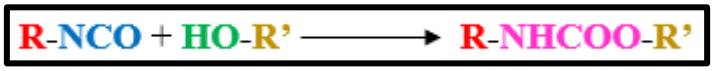
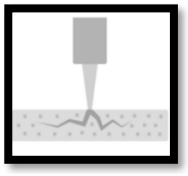



**Figure 1.** Formation of Urethane Linkage.

By altering the raw materials, additives, and the type of manufacturing process, the PUs can be tailor-made with a wide spectrum of product implications, including foams, fibres, adhesives, coatings, elastomers, and sealants, etc., Table 1 depicts the wide range of applications associated with polyurethanes. According to the latest report, the global polyurethane market will reach USD 88 million by 2026, with a compound annual growth rate (CGAR) of 6% [3]. The utilization of Polyurethanes in various spheres of life is represented in Figure 2.

**Table 1.** Wide Range of Applications Associated with Polyurethanes.

| Thermoplastic PU | Flexible PU | Rigid PU | Polyurethane Ionomers | Water-Based PU |
|---|---|---|---|---|
| | | | |  |
| Flexible and elastic with good abrasion, weather, and abrasion resistance. | Block copolymers with phase separation between soft and hard segments that are flexible. | Versatile and energy-saving insulator | Improved thermal and mechanical characteristics and greater dispersion in polar liquids due to increased hydrophilicity. | Coatings and adhesives which use water as solvent are called water-based polyurethane. |
| For example, automotive instruments, footwear, medical devices, film, sheet, and applications with a profile. | For example, cushion materials, carpet underlays, furniture, packaging, biomedical, and industries that use cushion materials. | For example, sound and thermal insulators. Used in commercial and residential appliances. | For example, biomedical devices, shape memory and are biocompatible. | For example, coatings, adhesives, sealants, binders |

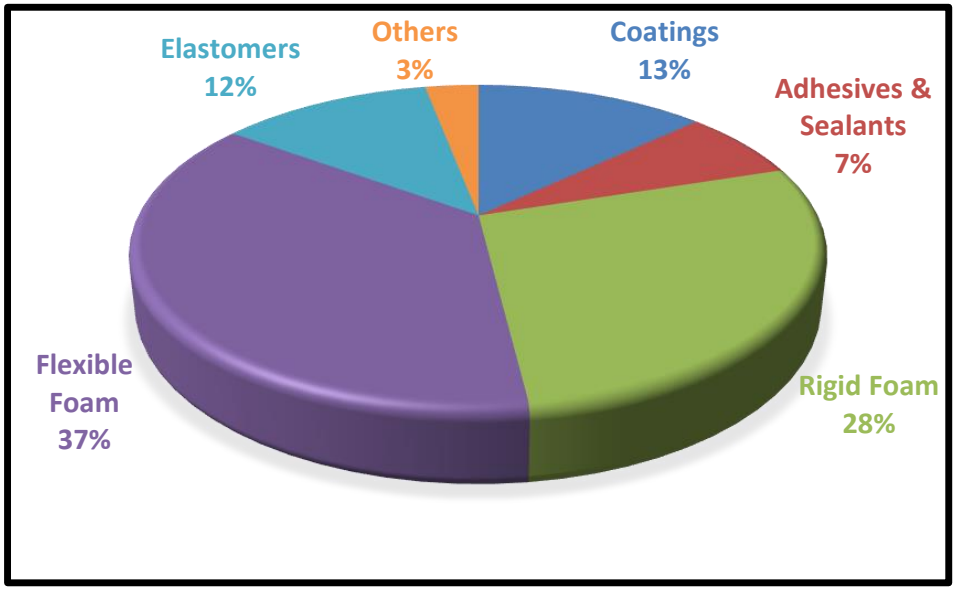



**Figure 2.** Global Consumption of PUs in Different Sectors.

The physical properties of PU, such as superb adhesion, adaptability, superior chemical resistance, and high automotive strength make it a great alternative to the widely used polymers such as formaldehyde and epoxy resins; however, the drawbacks associated with the preparation and usage of polyurethanes poses a serious threat to the human health and environmental hazards, which includes (i) the use of dangerous phosgene gas for the conversion of amines into isocyanates, (ii) usage of CMR (Carcinogenic, Mutagenic,

and Reprotoxic) labelled isocyanates for PU synthesis, namely Methylene Diphenyl Diisocyanate (MDI) and Toluene Diisocyanate (TDI), and the health risks associated with their long-term exposure, such as asthma, dermatitis conjunctivitis and acute poisoning (iii) post-consumption dumping (into landfills and dumping ground) or incineration of PU, leading to either hydrolytic degradation, that generates toxic amines which contaminates air and water bodies or the disintegration to yield hydrogen cyanide (HCN), during burning [4,5].

To address these issues, a new quest has already begun and gained popularity to produce non-isocyanate-based polyurethanes (NIPU), in which the common approach for the PU synthesis involves the reaction between cyclic carbonates and amines, resulting in a class of compounds called Polyhydroxyurethane (PHU) with hydroxyl groups [6,7]. This extensive review provides a deep insight into the recent research and developments related to polyurethanes prepared from bio-based materials being used in different segments, such as adhesives, coatings, elastomers, and foams for a broad range of practical applications. In addition, the new methodologies adopted in the field of NIPU, PU recycling, and PU biodegradation have been highlighted in detail.

### *1.1. Bio-Based Polyurethane*

Synthetic polyurethane materials are non-biodegradable, expensive, poisonous, and have petrochemical-based raw materials, which are now depleting. Bio-based polyurethanes are one of today's most flexible materials, which can play a significant role in fostering environmental goals of promoting waste reduction and sustainability. PU has seen a significant shift from traditional petroleum-based raw feedstock toward diverse renewable alternatives such as fatty acids [8], proteins [9], carbohydrates [10], vegetableoils [11], starch [12], polysaccharides [13], cellulose [14], and a variety of different agricultural products and by-products. Using renewable resources to produce polyurethane reduces adverse influences on the environment, for instance, greenhouse gas emissions. In the recent decade, diverse research groups have made rewarding efforts toward the synthesis of biopolyol-based polyurethanes [15,16]. Polyols behave as soft segments and transmit flexibility and exhibit elastomeric properties in the polyurethanes [17]. Polyols derived from vegetable oils have been the subject of extensive research. Hydroformylation [18], ozonolysis [19], amidation-esterification [20], metathesis [21], and epoxidation-ring opening [22], coupling of thiol-ene with mercaptoethanol [23] have all been used to achieve this, and all have successfully provided polyol with the desired properties; moreover, petrochemical-based isocyanates are replaced by bio-based isocyanate to remove toxicity such as partially bio-based aliphatic isocyanate. These impart water resistance, good mechanical strength, and flexibility to the PU network [24]. Hence, bio-based polyurethane materials were created to solve challenges such as solid pollution, economic effectiveness, and easy availability of raw materials. A comparison of the properties of bio-based PU and petro-based PU is given in Table 2.

**Table 2.** Comparison of the Properties of Bio-Based Polyurethane and Petro-Based Polyurethane [16].

| Property | Bio-Based Polyurethanes | Petro-Based Polyurethanes |
|---|---|---|
| Density (kg $m^{-3}$) | 112–181 | 141–181 |
| Thermal Conductivity ($Wm^{-1}K^{-1}$) | 0.06–0.0540 | 0.0540 |
| Volatile Organic Compound (µg $m^{-3}$) | Not detected at 61–91 °C | Not detected at 61–91 °C |
| Compressive Strength (N $mm^{-2}$) | 111–171 | 58.0 |
| Tensile Strength (N $mm^{-2}$) | 1.4–1.6 | 0.84 |
| Bending strength (N $mm^{-2}$) | 2.7–3.5 | 1.88 |
| Bending Stress (N $mm^{-2}$) | 1.55 | 1.50 |
| Water Absorption (kg $m^{-3}$) | 0.3 | 0.24 |

### 1.1.1. Vegetable Oil-Derived Monomers

Vegetable-oil-based polyols have reportedly been synthesized by ozonolysis, thiol-ene coupling process, photochemical oxidation, hydroformylation, ring-opening events, and epoxidation (Figure 3). Epoxidation is the major chemical pathway for making polyol, followed by epoxy group reactions with ring-opening reagents such as glycerol, alcohol, 1,2-propanediol, water, and acids.

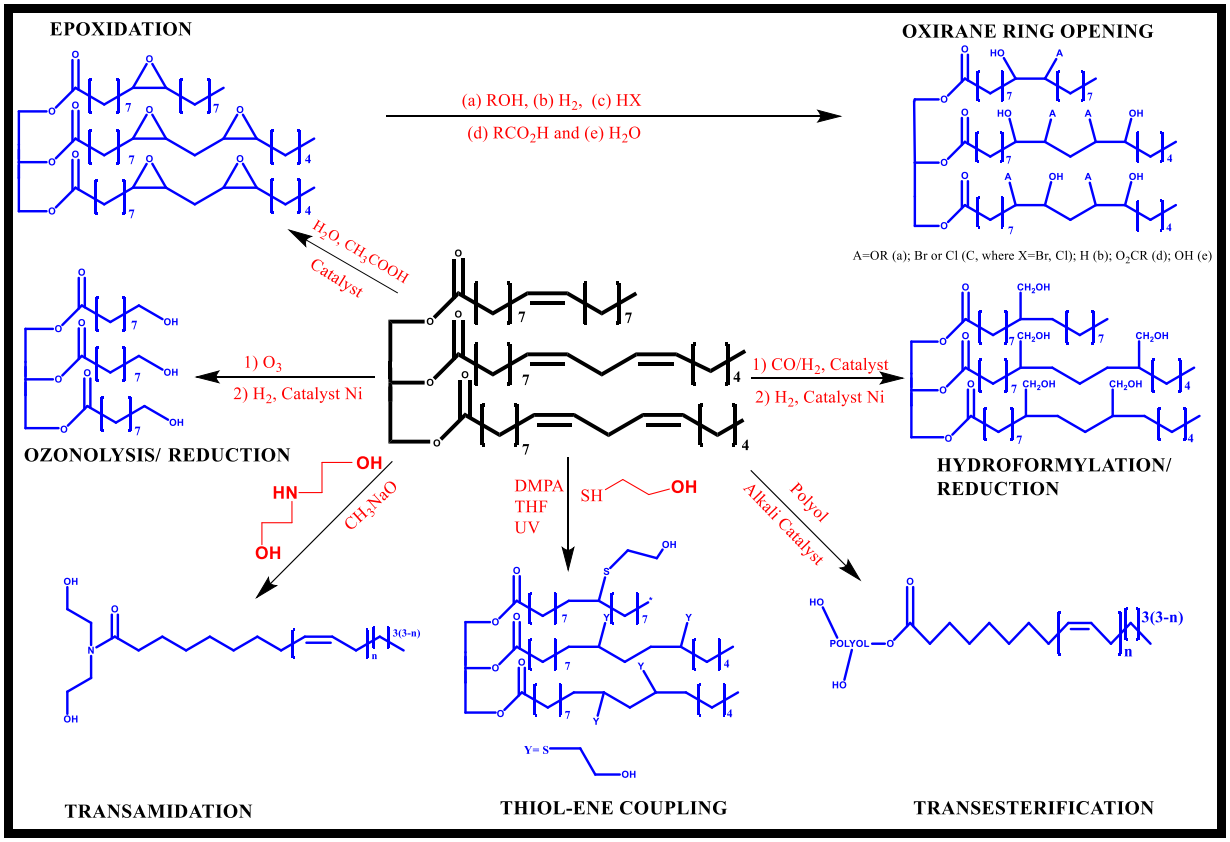



**Figure 3.** Synthetic pathways for the synthesis of polyols based on vegetable oil [25].

Additionally, lactate polyols, prepared from 100% bio-renewable material are polyester polyols that incorporate lactic acid units as a component. Polyurethane obtained from lactate polyols is found to be biodegradable and biocompatible. Two methods are commonly being reported for the synthesis of polyols from polylactic acids: (1) ring-opening addition (lactide is added to hydroxyl groups to open the ring). (2) Different polyols are esterified using lactic acid or Transesterification with lactic acid esters [25].

### 1.1.2. Bio-Based Isocyanates

To boost the renewable content of PUs, partially or potentially bio-based di-, tri- or poly-isocyanates have been utilized. There are also reports on PU films made with two bio-based isocyanates—ethyl ester L-lysine diisocyanate (LLDI) and ethyl ester l-lysine triisocyanate (LLTI)—and one commercial isocyanate, isophorone diisocyanate (IPDI) [25]. Another study documents the preparation of bio-based polyurethane by utilizing isocyanates derived from palm oil, which was reported to exhibit low toxicity, low carbon emission, along with being cost-effective [26].

## 2. Polyurethanes: Types, Synthesis, and Utilities

### *2.1. Polyurethane Adhesives*

Polyurethane (PU) adhesives are known for their remarkable adhesion, weathering resistance, excellent low-temperature performance, flexibility, and rate of curing, which can be speckled according to the need of the manufacturer [27,28]. PU adhesives have advantages over other adhesives being pressure-sensitive, having high tensile strength, compressive modulus, and rip strength, and being transparent at room temperature. Owing to a vast diversity among the raw materials, the properties of the PU adhesives can be accustomed to a vast extent as per the needs of the users. The high-strength PU adhesives can be utilized in both inelastic and permanently elastic applications [29].High penetration of the adhesive into the wood, limited resistance to exfoliation, and undesired gap-filling properties are the main drawbacks of PU adhesives [30]. Most of the commercially important PU adhesives can be categorized as one component or two components of adhesives, as depicted in Figure 4.

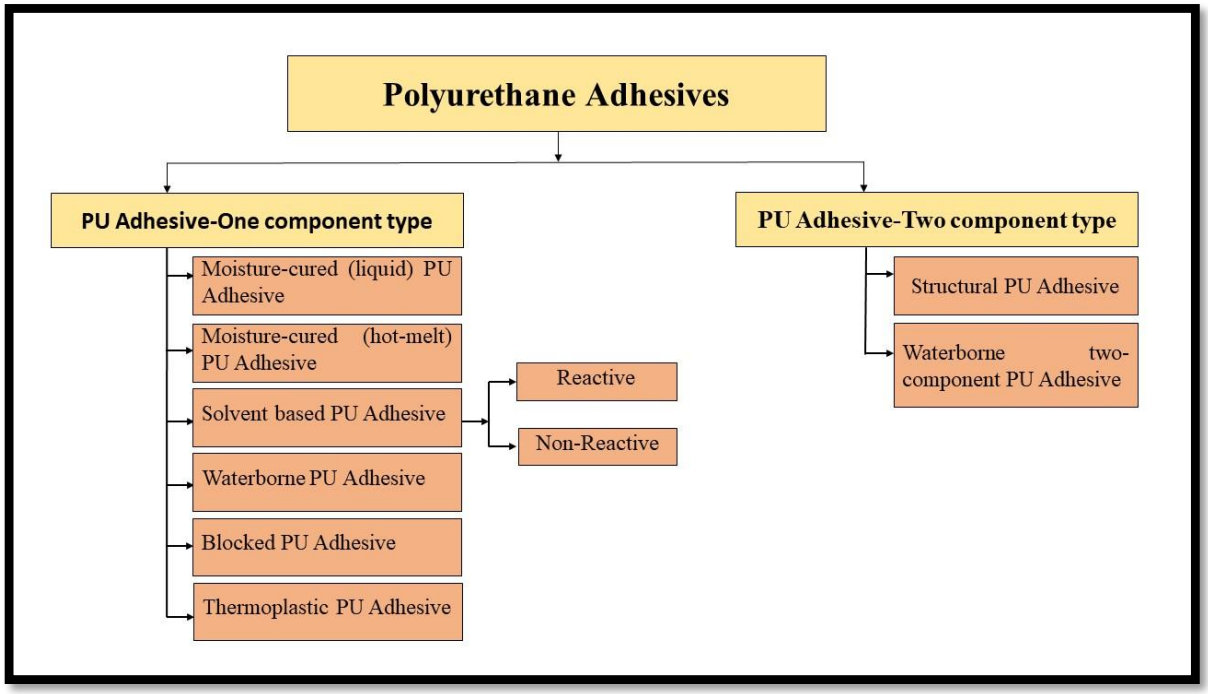



**Figure 4.** Classification of Polyurethane Adhesives Based on Number of Constituents.

**PU Adhesives—One-component Type:** One component adhesives are self-curing adhesives that do not have to be mixed with any other ingredients (curing agent) before being applied to the substrate; they use an isocyanate-terminated pre-polymer that cures naturally in the presence of moisture. After applying to substrates, a mist of water is spurted on the adhesive to speed up the curing process [30].

Other types of one-component PU adhesives include solvent-based PU adhesives, which are further categorised as reactive solvent-borne PU adhesive and non-reactive solvent-borne PU adhesive [31]. The former involves high-molecular-weight oligomers bearing isocyanate functionality and the latter shows no further reaction after application. Solvents commonly used in these adhesives are toluene, ketones, ethyl acetate, butyl acetate, etc.

Additionally, Waterborne polyurethanes are a kind of system composed of a dispersion of binary colloidal particles within an aqueous medium that necessitate the addition of water-compatible stabilising groups (emulsifiers) due to their hydrophobicity [32]. Blocked PU Adhesives, as the term 'blocked' implies, are the one in which one of the reactants

involved is blocked chemically to inhibit the reaction. Commercially available blocked isocyanates are mostly based on TDI, and derivatives of HMDI.

The thermoplastic PU adhesives are linear in structure with a functionality equal to 2 [33]. The desired molecular weight to synthesize thermoplastic PU adhesives is about 100,000 or higher; these are superior compared to hot melt adhesives, in terms of holding strength, attained by crystallization.

**PU Adhesives—Two-Component Type:** Two-component PU adhesives are applied where, a better flexible bond line is desired; these are commonly used for bonding two different kinds of substrate, which exhibits a wide difference in their thermal expansion coefficients. Due to their fast reaction rate, two-component polyurethane adhesives are generally prepared with components supplied in separate compartments. The two-component PU adhesive is broadly categorized as Structural PU adhesives and Waterborne Two-Component PU adhesives.

Structural PU Adhesive provides excellent mechanical strength by forming a chemical cross-link between the reactive components, thus leading to a three-dimensional polymer network, a necessity for structural performance [34]. Waterborne Two-Component PU Adhesive consists of two components, with one of the components as PU dispersions and the other one as a crosslinker [35].

Moreover, the addition of fillers can potentially change the physicochemical and mechanical properties of adhesives. Talc, calcium carbonate, silica, titanium oxide, carbon fibre, clay, and other fillers are commonly used in PU. The effect of nanosized $TiO_2$ filler on various characteristics of PU glue was investigated in a recent study. In PU adhesives, $TiO_2$ has been shown to improve surface adherence, solar reflectance, and crack resistance [36].

In general, synthetic adhesives are more resistant to moisture, chemicals, and biological organisms than natural adhesives. Bio-renewable sources such as vegetable oils, starch, protein, lignin, lactic acid etc. constitute a valuable source for the synthesis of polyols to produce "sustainable, environment friendly'' PU adhesives [37]. Bio-based polyurethane adhesives from various sources are discussed here:

#### 2.1.1. Protein

Protein-based PU adhesives can be prepared from both animal as well as plant sources. The advantages and disadvantages of the different animal and plant protein-based adhesives have been reported in Table 3.

**Table 3.** Different Animal and Plant Protein-based Adhesives [38].

| **Animal-Based Adhesives** | **Advantages** | **Disadvantages** |
|---|---|---|
| Tissue and bones | Cheap, non-toxic, sets quickly, has a high shear strength, and doesn't discolour wood | Has low water and damp resistance and does not produce distinct adhesive lines |
| Albumin and blood | Rapid heat setting, excellent dry shear strength, microorganism resistance, and does not discolour wood | Suitable to glue thin sheets |
| Casein | Resistance to water, damp conditions, elevated temperature, high shear strength | May discolour woods, dissolves only at high pH, expensive, not suitable for exterior use |
| **Plant-Based Adhesives** | **Advantages** | **Disadvantages** |
| Soybean | Non-toxic, has strong dry strength, and is moderately resistant to heat, wet, and damp environments. | Denaturation of proteins in alkaline solution which is caustic and discolour wood, sensitive to microbial degradation, and not suitable for exterior applications |
| Peanut | Non-toxic, hygroscopic, superiority of colour | Low shear strength, produces blisters, small bubble-like voids, and peanut allergies limit its applications. |
| Gluten | Non-toxic, sustainable source of protein | Costly |

### 2.1.2. Vegetable Oils

Vegetable oils are an excellent raw material for the production of PU adhesives. The most important component of vegetable oil is triglycerides, which are glycerol esters with three long chains of fatty acids of varying composition depending on the source of the oil. As triglycerides hydrolyse to form a range of fatty acids and glycerols, polyols, and isocyanates, raw materials for the synthesis of polyurethane can be derived from them. Various vegetable oils that can be utilized for PU synthesis are Castor Oil, Palm Oil, Canola Oil, Soyabean Oil and Jatropha Oil.

#### Castor Oil

Of all the vegetable oils, castor oil is the favourite among the researchers and the most commonly used one because of its renewability, low cost, easy availability, and being non-edible. Castor oil contains ricinoleic acid, which is a monounsaturated fatty acid, bearing a hydroxyl functional group and thus, can be employed for PU synthesis [39]. Different compositions of PU adhesives derived from castor oil have been reported in Table 4. Being non-edible (due to the presence of toxic protein, i.e., ricin), there is no hesitation among the researchers to use it as a potential source of PU synthesis. The mechanical properties of the castor oil-based adhesive are highly dependent on the NCO/OH molar ratio, due to higher or lower cross-linking achieved during the reaction. Studies on solvent-less castor oil-based PU adhesives reported the impact of the NCO/OH molar ratio and the chemical characteristics of the substrates on the adhesion force [40]. As reported, the resulted PU adhesive foam joints displayed peeling strength values of 75% and wood joints showed lap shear strength values 20% higher than that of commercialized solvent-based adhesives.

**Table 4.** Different Compositions of PU Adhesives Derived from Castor Oil [35].

| S. No. | Compositions of PU Adhesive | Method of Preparation | Applications |
|---|---|---|---|
| 1. | Castor oil—pentaerythritol | Trans esterification reaction | Bonding/adhesion of wood to wood and metal to metal |
| 2. | Castor oil—polyester polyols | Trans esterification reaction | Wood adhesive |
| 3. | Castor oil—cellulose acetate | Condensation reaction | Wood-wood, steel-steel adhesion |
| 4. | Castor oil—MDI modified cellulose acetate | Condensation reaction | Wood-wood, stainless steel lubricants |
| 5. | Castor oil-based PU adhesives | Condensation/trans esterification reactions | Wood adhesives, sealants |
| 6. | Castor oil—dicarboxylic acid (maleic acid, fumaric acid, oxalic acid) | Condensation reaction | Wood Adhesives |

#### Palm Oil

Palm oil and palm kernel oil are the two main forms of oil generated from the palm oil tree. Palm oil has been reportedly used by researchers for the production of PU adhesives, which shows higher hydrolytic stability and hydrophobicity. Patel et al. studied the ring-opening reaction of epoxidized palm olefin with phthalic acid, which is utilized to generate PU adhesives with, which contain palm oil-based polyester [41]. Khoon et al. studied polyurethane wood adhesive derived from palm oil-based polyester polyol, where glycerol was used as a crosslinker and DBTL as a catalyst [18]. TDI and MDI were employed as isocyanates in this study. Radojčićetal., explored the ring-opening polymerization reaction of palm oil-based polyol to synthesize polyester PU with low hydroxyl content, with high molecular weight exhibiting a longer gel time [42].

Canola Oil

Canola oil has unsaturated double bonds in its fatty acid chains as well as an ester group, which allows it to be employed in the manufacturing of a wide range of important goods, such as polyols, owing to its high protein content and hydrophobic properties. Canola oil-based PU adhesives have excellent attributes such as equivalent lap shear strength, chemical resistance, and hot water resistance, implying that they have a lot of potential in particleboard manufacture. Li et al. [20] studied the adhesion attributes of sodium bisulfite ($NaHSO_3$)-modified canola protein. Protein was extracted from canola meal through alkali solubilization and acid precipitation methods and was further modified with different concentrations of $NaHSO_3$ (0–15 g/L). With the increase in the concentration of $NaHSO_3$, the purity of canola protein was found to be decreased. The isolated wet protein was used as an adhesive.

Jatropha Oil

Jatropha oil is a non-edible, renewable oil with gum that can be utilized in adhesives, with excellent adhesion and improved mechanical characteristics. Aung et al. reported jatropha oil-based PU adhesive bearing epoxy groups which enhanced adhesion [19]. Gadhave et al. reported jatropha oil-based polyurethane adhesive via epoxidation reaction of hydroxylated polyols with 2,4-toluene diisocyanates [43]. Saalah et al., synthesized a renewable jatropha oil-based polyol by polymerization of jatropha oil with isophorone diisocyanate and dimethylol propionic acid; the product is widely used as a binder for wood and decorative coatings [22].

Soybean Oil

Soybean-based adhesives have mostly been utilized in the production of adhesives as well as wood composites such as particleboard and medium-density fibre board. Dai et al. synthesized Soy oil-based PU adhesives by reacting BIOH polyols X-0002 and IPDI via maintaining NCO/OH ratio 1:4 and examined the variations in physical and mechanical properties with hydroxyl functionalities in soy-based polyols in PU films [37].

2.1.3. Lignin

Lignin is generated as a by-product from the paper and pulp industry. The glass transition temperature, aromatic content of the final polymer, and thermal stability of adhesives can all be improved by including lignin into the polyurethane matrix as a polyol substitute. Due to its aromatic crosslinked structure, lignin is somewhat steady at high temperatures and works as a thermosetting material. Because of its aromatic composition, lignin acts as a hard segment and replaces the soft segment of PU, i.e., polyol, and boosts the mechanical strength, thermal characteristics, and composite material properties of blends, copolymers, and composite materials. Lignin incorporation into PU matrix leads to enhancement in its thermal stability, delamination, and resistance to abrasion [44]. Alinejad et al., prepared lignin-based polyurethane adhesives by polymerization of lignin with toluene -2,4-diisocyanate (TDI) and partial polyethylene glycol (PEG). Partial polyethylene glycol was used but, it was further replaced by lignin to avoid the hard and brittle property. The final product was used as wood adhesives. The amount of lignin incorporated enhanced the glass transition temperature and thermal stability of the adhesive [29].

2.1.4. Starch

Starch is widely being used in many industrial applications, such as the food industry, pharmaceutical industry, paper industry, and printing industry because of its inherent biodegradability, abundance, and renewability [45]. The main sources of starch are wheat, rice, potato, corn, etc. Two natural resources are potato starch and edible or non-edible plant-derived oils that can be employed to prepare PU adhesives. Starch glycosylation followed by oil transesterification to generate polyol can be used to make starch-based PU adhesives.

#### 2.1.5. Polylactic Acid

Thermoplastic polyester polylactic acid is noted for its biodegradability. Lactate polyols are polyester polyols that include lactic acid units as a component and can be prepared from 100% bio-renewable material. Gadhave et al., prepared high functionality polyester polyols containing lactate polyols suitable for rigid cast polyurethane, having high bio-based content and that exhibit biodegradability [43].

#### 2.1.6. Applications of Bio-Based PU Adhesive

Owing to the high stickiness on a variety of substrates such as metals, wood, plastics, glass, etc. These bio-based polyurethane adhesives have a wide range of industrial applications. The applications of bio-based polyurethane adhesive in assorted sectors are described below:

*In Construction:* PU adhesives are widely used in wood flooring, heavy-duty rubber flooring, and parquet bonding due to their great strength and resistance. PU adhesives are suitable for ceramic tiles, mirrors, bath surrounds, and concrete due to their ability to bond wet/frozen timber, as well as their high green strength and shear strength [27].

*In Packaging:* Snack food bags, printed films, and shopping bags commonly use one-component PU adhesives for adherence. In addition, for medical applications, a tailored two-component PU adhesive is used to adhere polyvinyl chloride to aluminium sheets. Hot melt PU adhesives are suited for wrapping applications due to their low viscosity and low application temperature.

*Transportation:* Polyurethane adhesives are used in a variety of industries, including automotive, trucks, and railways. Polyurethane adhesives are highly appreciated due to their short cure time and ability to generate strong flexible bonds [46].

*Electronics:* PU adhesives are good for bonding copper foil, electronic coil fabrication, surface fix conductive adhesion, and bonding die-stamped printed circuit boards due to their properties of curing at room temperature, endurance, and great resistance to thermal shocks and vibrations.

*Footwear:* The shoe industry requires PU adhesive to join waterproof insole materials and in the sole attachment because of its excellent water resistance, substantial heat resistance, flexibility in cold rapid settings, and good initial strength [47].

*Biomedical applications:* Tissue adhesives have gained popularity in recent decades, owing to an increase in the number of surgical procedures performed around the world. Depending on the adhesive purpose, these tissue adhesives are categorised into three basic types: (i) Hemostats, which prevent blood loss by interfering with the coagulation cascade, generally have weak mechanical characteristics and are employed in cases of blood loss owing to sutures tissue injury; (ii) glues that cling strongly to tissues and (iii) Sealants, which create a physical barrier to stop air/gas or blood seeping [48]; they usually provide mid-range tissue adhesion and are utilised in conjunction with sutures [34], and these systems, on the other hand, perform better in a dry atmosphere.

### *2.2. Polyurethane Coatings*

PU coatings are used to enhance mechanical strength and to induce desired barrier properties in the components of aircraft, automotive, container ships, oil tankers, industrial machines, residential refrigerators, etc., [49]. In addition, these organic coatings shield the materials from intrusive environmental conditions such as humidity, biological deterioration, chemical and automated destruction, and radiation and impart colour and gloss [50]. Different types of PU coatings of industrial importance are discussed in the subsequent section.

**High solids polyurethane coatings:** High solids coatings have narrow molecular weight distribution and low solvent emissions; they have a suitable spray viscosity, which in combination with their low molecular weight, is responsible for their drying behaviour, without compromising the performance of the material [51].

**Ultraviolet (UV) curable polyurethane coatings:** PU curing polyurethane coatings have become popular in recent years. Polymerization and thermal curing processes, which involve solvents and catalysts, are used to fix traditional polyurethane coatings [52]. UV-curing technology is distinguished from thermal curing by the 5E abbreviation, which stands for Eco-Friendly, Economical, Efficiency, Enabling, and Energy Saving, making it "clean and green" [53]. Polyurethane acrylate-based UV-curable coatings are excellent in imparting abrasion resistance, impact resistance, scratch resistance, and optical transparency [53].

**Hyperbranched polyurethane coatings:** The three main types of dendritic polymers are dendrimers, hyperbranched polymers, and dendrigrafts. Because of their unique structural properties such as no entanglement and high surface functionalities, hyperbranched polymers and dendrimers emerged as an important class of polymeric materials in the recent years; they also have enhanced cross-link density, excellent reactivity, and reduced viscosity. Hyperbranched polymers, on the other hand, are preferred over dendrimers due to their one-step synthesis and fewer purification processes. Hyperbranched structures outperform linear structures in terms of characteristics [54]. Hyperbranched poly (ester amide) was also shown to be biodegradable, which might be considered a green advantage in advanced surface coating technology [55].

**Water-borne polyurethane coatings:** WPU material has received a lot of attention in the past few decades due to its many advantages, such as good flexibility, stiffness, impact and abrasion resistance, gloss, chemical resistance, lesser flammability, high adhesive strength, durability, low viscosity, easy cleaning and weather-ability, as well as zero or low volatile organic chemical compound emissions [56]. Because the viscosity of a WPU dispersion is independent of the polymer's molecular weight, high solid content WPU dispersions with high-molecular-weight films can be produced with outstanding quality simply by physical drying [57]. A binary colloidal system in which PU particles are continually disseminated in aqueous solutions is known as PU dispersions. Due to ionic groups (PU ionomers) in the linear thermoplastic PU backbone, WPUs are dispersible in water. Polyurethanes, on the other hand, are not water-dispersible due to hydrophobic isocyanates (which also react with water) [58]. A PU ionomer is a copolymer with a PU backbone and repeat units that include pendant acid groups that are neutralised completely or partially to form salts. Internal emulsifiers include cationic-like quaternary ammonium groups, anionic-like carboxylated or sulfonated groups, and non-ionic groups such as polyols containing ethylene oxide end groups. Micelles are formed in water by cationic and anionic PU ionomers, respectively. The types of ionomers, isocyanates, and polyols utilised all have an impact on the performance of the resulting WPU dispersions.

*UV curable WPU coatings:* UV stable coatings cure quickly when exposed to ultraviolet light, resulting in outstanding performance. UV-curable waterborne polyurethane combines the advantages of UV equipment with those of aqueous coatings. UV-curable WPU coatings had been widely investigated because of their flexibility, non-toxic behaviour, chemical resistance, and remarkable mechanical properties. UV-WPU coatings, despite their many advantages, have worse solvent resistance and thermal stability than typical solvent-based polyurethane coatings, and the tensile strength is significantly reduced at temperatures above 80 °C. Due to their lower VOC emissions than solvent-based PU coatings, environmentally friendly waterborne polyurethane (WPU) coatings are widely employed [58]. The UV-cured WPU coatings are a significant class of green and environmentally friendly coatings with excellent mechanical qualities and a quick curing procedure. Due to numerous end groups, compact molecular structure, and reducing chain entanglement, hyperbranched polyurethanes (PUs) have intriguing features such as high solubility, reactivity, and good rheological behaviour.

*Hyperbranched WPU coatings:* Because of their distinct molecular architectures, such as compact molecular nature, numerous end groups, and decreased chain entanglement, hyperbranched polymers differ significantly from their linear counterparts. Branched-chain polymers have the most exciting properties, such as good rheological behaviour, strong

reactivity, and solubility, and so have a wide range of applications [59]. The step-growth polycondensation technique is used to make polycarbonate, polyphenylene, polyesters, poly(ether ketone), polyamides, poly(4-chloromethylstyrene), polyurethanes, and other hyperbranched polymers. The number of generations, which is a statistical measure of the total number of monomers in a hyperbranched polymer, influences the compatibility of hyperbranched WPUs with nanofillers.

*Nanocomposite WPU coatings:* Waterborne surface coating materials with low volatile organic compounds (VOCs) are environmentally safe and, as a result, the most sought-after materials in today's society. Waterborne polyurethane (WPU) coatings offer excellent stiffness and elasticity, but they have much lower mechanical strength and toughness than solvent-based coatings. Nano-sized inorganic fillers have been added to WPU dispersions to create nanostructured coatings, demonstrating that this is a potential method for increasing WPU properties. The manufactured WPU nanocomposites have a lot of potential as an environmentally friendly transparent surface coating material with low VOC content.

### 2.2.1. Vegetable Oil-Based Polyurethane Coatings

Polyurethane coatings based on vegetable oil have proven their worth by exhibiting significant properties such as abrasion and chemical resistance, remarkable toughness, corrosion resistance, low-temperature flexibility, and prospering industrial applications, as well as lowering or eliminating the use of volatile organic compounds (VOCs). The type of PU coating depends on the curing procedure and solvent employed. In addition to dispersing conventional polyurethane using petrochemical solvents, extensive research has been done on water-based PU extracts that use water as a solvent. Dispersions of waterborne PU have an inherent benefit over traditional in that they are environmentally friendly and safe to use [54]. PU treatment can occur at room temperature or below room temperature. UV curing is a type of thermal curing that uses UV radiation to provide the energy needed for the curing operation. To improve the properties of conventional PU coatings to fulfil the industrial norms, extensive investigation has been done on high-solid contents, nanocomposites, and hyperbranched coatings of PU, which are dependent on the respective properties and compatibility of distinct components [49]. Because it takes less time, uses less energy, and has a better curing action than other techniques, this method has proven to be more efficient than others. A sustainable alternative to mitigating failures from degradative corrosion is the appropriate utilization of domestically leftover resources such as vegetable oils (Figure 5).

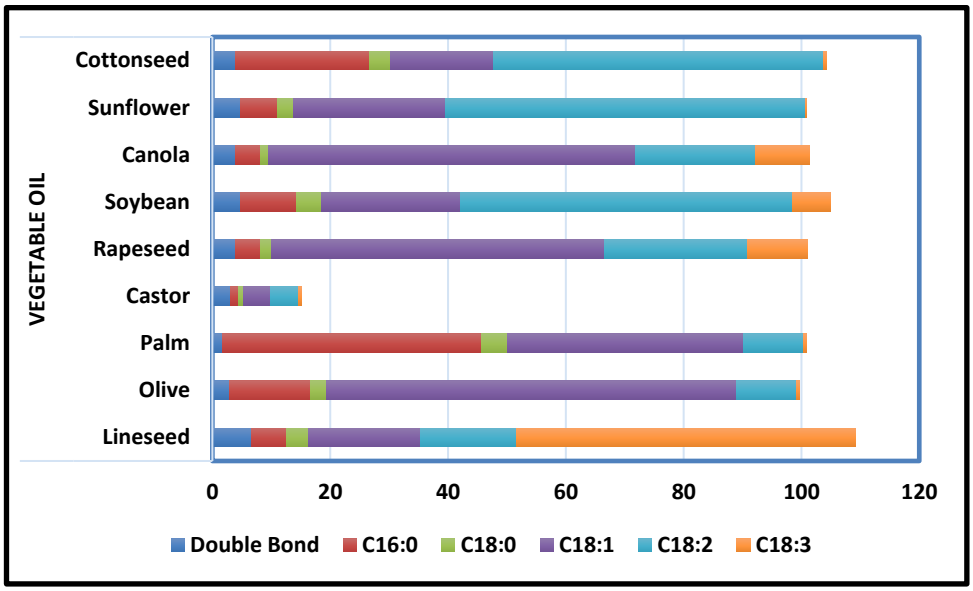



**Figure 5.** Common Vegetable Oils' Degree of Unsaturation and Fatty Acid Composition [55].

2.2.2. PU Coatings Based on Cashew Nut Shell Liquid (CNSL)

In addition to vegetable oils, CNSL, an agricultural by-product of the cashew industry, has been utilized as binders for coatings material [60–62]. Meta substituted phenolic compounds identified in CNSL, a reddish-brown phenolic lipid, include 2-methyl cardol, cardol, cardinal, and anacardic acid. High vacuum distillation of CNSL yielded cardanol, a crucial precursor for a variety of oligomers and polymers [62,63]. For polymer curing, the meta-position of an unsaturated alkyl chain ($C_{15}H_{31-2n}$, n = 0–3) can be used as an autoxidizable site [64]. CNSL (cashew nut shell liquid) is a naturally occurring substituted phenol that can be used in place of phenol, with similar or superior outcomes. It is a low-cost and renewable material that finds use in Insecticidal, fungicidal, anti-termite, and therapeutic applications. In addition, it is also used as an additive in many plastic compositions. Cardanol-based polyols are considerably modified (the most common modification site is the phenolic hydroxyl group), resulting in autoxidizable bio-based polyurethane coatings with isocyanates [64]. All cardanol-based polyols have a 15-carbon side chain at the meta position, which may affect the polymers' thermal and mechanical properties, as all changes are made to these unsaturation sites. Friction materials, cars, surface coatings, adhesives, laminates, rubber compounding, and a range of other applications use CNSL resins. The chemical structure of different cardanol-based polyols (CAP) from aromatic cardanol is depicted in Figure 6.

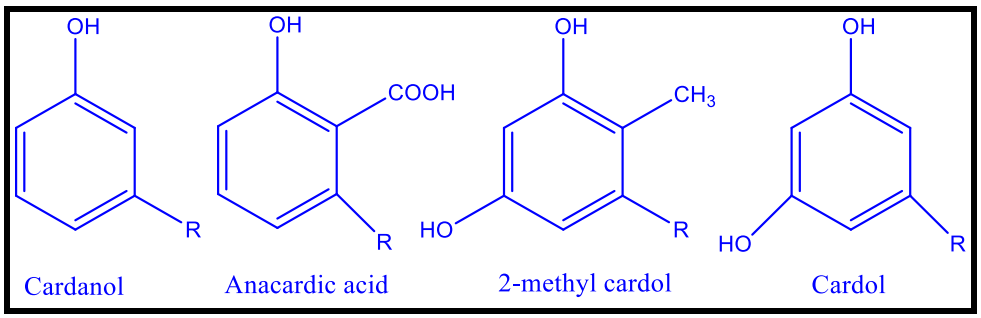



**Figure 6.** Chemical Structures of CNSL's Major Components [63].

Because of the aromatic ring of cardanol, the PU backbone is particularly rigid, and the PU coating has a firm adherence to the metal, resulting in improved stiffness and hardness. Aside from CNLS-based polyol, CNLS-based cyclic carbonate (CC) was efficiently used as a precursor for non-isocyanate polyurethane preparation (NIPUs) [64]. For the synthesis of NIPU, di-glycidyl ether of cardanol (NC-514), a CNSL-derived product and epoxy formulation modification, was employed.

2.2.3. Polyurethane Coatings Based on Terpene

Terpene-based polyols with strong impact strength, water resistance, adhesion, flexibility, and pencil hardness can be crosslinked with poly-isocyanate to produce PU coatings (Figure 7). Terpenoids are good bio-based raw ingredients that have a high value in polymer applications [65].

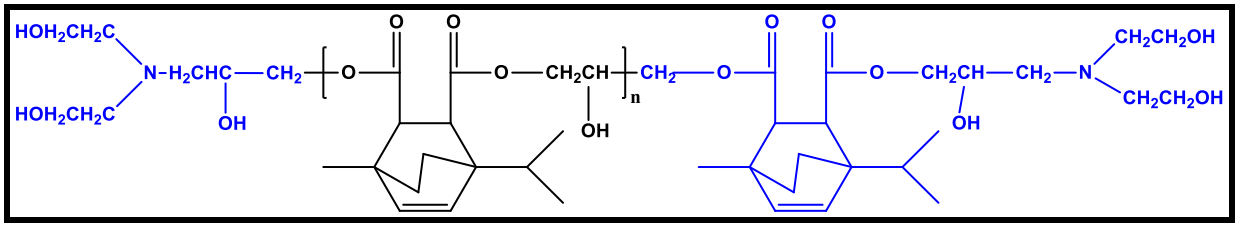



**Figure 7.** Chemical structure of polyols based on terpene [65].

#### 2.2.4. Polyurethane Coatings Based on Eucalyptus Tar

Eucalyptus tar is a complicated combination primarily made up of phenols, particularly guaiacyl and syringyl derivatives [66]. A solid residue known as bio-pitch is created during the distillation of bio-oil, and it makes up a considerable component of the final product (approximately 50%). To make a tar-based varnish and a bio-pitch-based black paint, eucalyptus tar derivatives are used as a source of hydroxyl groups. Eucalyptus tar was distilled to produce heavy oil and bio-pitch, which were then mixed with castor oil to create a renewable polyol [67].

### *2.3. Polyurethane Foams*

Among most polymers, the actual kingpin of the cellular structure kingdom is Polyurethane (PU) foams, which account for 67% of the global PU consumption and represents 50% of the polymer foam market [68]. PU foams hold a good reputation for several applications such as thermal insulation, acoustic insulation, automotive seating, mattresses, etc. Based on their density they are classified as rigid, semi-rigid, and flexible foams.

**Bio-based polyurethane foams:** Bio-based building blocks such as alginate and other carbohydrates have been successfully integrated into PU foam networks via their hydroxyl groups (-OH) for improved biocompatibility and breakdown kinetics.

#### 2.3.1. Vegetable Oil-Based Polyurethane Foams

In recent years, interest in the synthesis of vegetable oil-based polyurethane foams has increased significantly. Lubczak et al. [69] produced rigid polyurethane foams based on polyol derived from hydroxyalkylation of cellulose with glycidol and ethylene carbonate. These foams exhibited several inherent properties equivalent to conventional rigid polyurethane foams including apparent density (60.5 kg/m$^3$), low shrinkage (~3%), and low water absorption percentage of 6.6%. The thermal Conductivity of these foams turned out to be 0.0338 W/m K. Lin et al. [70] produced bio-based polyurethanes utilizing polyols consisting of lipids derived from enzymatic hydrolysis of food wastes, comprising a mixture of rice, noodles, meat, and vegetables, etc. The polyols required for foam synthesis were developed by trans-amidation with diethanolamine and epoxidation followed by epoxy ring-opening with trimethylolpropane. Takemura et al. [71] tried to reduce the reliance on petroleum-based polyol by using four different sets of agricultural waste liquefied polyols including Oil seed Rape Straw (OS), Rice Straw (RS), Wheat Straw (WS), and Corn Stover (CS) respectively and consequently studied the influence of NCO/OH on the foam properties such as density, compressive strength, elastic modulus, morphological structure, storage modulus (E'), glass transition temperature (Tg), water absorption ability, etc. respectively. It was observed that density, compressive strength, and elastic modulus first decrease then held stable, and finally improve with increasing NCO/OH ratio; moreover, the glass transition temperature values shift to the higher temperature range at a high NCO/OH ratio. Chittapun et al. [72] investigated the usage of cellulose extracted in a high yield percentage of ~37.2% from Kai algae composed of *Cladophora* sp. and *Microspora* sp. by the alkalization of macroalgae having incorporated in varying dosages as a reinforcing filler for polyurethane foam composites. The addition of algal cellulose led to open cells and softer foam with reduced cell size along with an increase in apparent density, however, a decline in compressive modulus values was observed. The presence of filler at 1% ($w/v$) concentration has caused increased stiffness and transition temperature but a drop in loss modulus.

Hu et al. [73] successfully developed flame retardant and self-extinguishing flexible polyurethane foam by coating the foam surface through a layer-by-layer assembly of up to 9 bilayers of alginate, alginate, chitosan, and hydroxyapatite particles; they studied the effect of the number of bilayers deposited on the foam surface on the flame-retardant and the mechanical properties of the resultant foams. PU Foams coated with 9 bilayers exhibited self-extinguishing characteristics at 32 s during exposure to a butane torch. The incorporation of 9 bilayers of FR coating to the flexible foam remarkably led to a

decline in both peak heat release rate and smoke production rate by 77.7% and 53.8% respectively. The introduction of hydroxyapatite particles as reinforcement having a rigid surface resulted in an enhancement in tensile strength value by 56% against uncoated foams [74]. Senthil et al. [75] prepared reinforced biodegradable polyurethane foams and strips in a one-shot process by reacting a trifunctional polyol-polycaprolactonetriol as a soft segment and isophorone diisocyanate (IPDI) as a hard segment while incorporating fullerenes as a reinforcing agent in different loadings into the cellular network structure. It has been observed that the density and the hardness values of the fabricated foams tend to increase and vary as a function of $C_{60}$ concentrations. Leszczyńskaa et al. [76] investigated the combined role of renewable resources on the individual properties of cellular foams during the fabrication of rigid polyurethane foams; they utilized different concentrations (10–50 wt.%) of rapeseed oil-based polyol (RO) and egg shell (ES) waste in a fixed amount of 20 php as reinforcing filler for the production of rigid foams. The obtained foams appeared light-yellow in colour with apparent density ranging from 85 to 93 kg/m$^3$. The presence of RO polyol and ES filler induced lowering of glass transition temperature in the final foams owing to the plasticizing effect of fatty acid hydrocarbon chains of vegetable oil-based polyol. The compressive strength values of obtained foams were also found to be decreased with the addition of RO in the formulation recipe. The incorporation of vegetable-based polyol in higher concentrations has also compensated for the rise in foam friability and water absorption characteristics due to the presence of filler powder while improving the dimensional stability of the resultant foams.

#### 2.3.2. Application of Bio-Based Polyurethane Foams

Polyurethane foams are one of the most versatile materials used today in different engineering and household applications, which include:

*Insulation:* Polyurethanes are inexpensive, long-lasting, and reduce carbon emissions [77]. Polyurethanes provide insulation that reduces heat loss in buildings, automobiles and refrigerators, thus saves energy. Polyurethane insulations are popular because of their durability and less maintenance and easy installation [78].

*Furniture and Bedding:* Polyurethanes are widely employed in furniture and beddings [79]. Flexible polyurethane foams are generally used for cushioning purposes, as their density can be adjusted according to the requirement of the users.

*Footwear:* Polyurethanes allow designers to make footwear that should be restful, durable and fit for use [80]. Polyurethanes with good abrasion-resistant properties are used to make shoe soles of high strength.

*Packing Foams:* Packing foams provides protection and cushioning to packaged products [81]. Polyurethane foams are available in convoluted, flat and custom-cut shapes and are often used to pack highly sensitive equipment such as electronics and circuit boards.

*Rafting Boats:* Raft manufacturers use RPUFs to produce inflatable boats and surfboards, as it provides strength, buoyancy and sounds deadening [82].

*Biomedical applications of PU foams:* There are various advantages to using PU foams for wound dressing purposes; they can provide acceptable gas and moisture mobility while having little adhesion to the wound spot due to their tunable porosity and pore diameters [83]. Silver nanoparticles for asiaticoside and antibacterial activity, a recognised herbal wound curing agent, were used to enhance bioactive characteristics. High polysaccharide components contributed to a powdery surface and reduced mechanical characteristics, so foams with middle polysaccharide components were analysed for their experimental efficacy in human patients with distressing dermal burns or abrasion wounds [84]. Outcomes demonstrated non-infected and accelerated wound healing as compared to usual treatments. Silver nanoparticles display cytotoxicity at high amounts, an approach to tackle this concern is by nanoparticle encapsulation with phenolated lignin before its incorporation into the PU foam [85].

### *2.4. Polyurethanes Elastomers*

Polyurethane is an extensively used synthetic elastomer in biomedical applications due to its outstanding hemocompatibility and biocompatibility. Amending its chemical structure can affect its mechanical attributes such as responsiveness, fatigue resistance, elasticity, tolerance, or durability in the body during healing. Additionally, adjustment of chemical groups for PU structure, allows for an alteration of surface and bulk, by introducing biomolecules such as biorecognizable groups or anticoagulants, as well as hydrophilic/hydrophobic balance. Such variations have been developed to improve implant acceptability. As a consequence, traditional solvent-based polyurethanes have become the industry standard for high-performance systems, and are widely employed in medical devices such as tubing, dressings, antimicrobial membranes, catheters, and total artificial hearts, among others [86]. Figure 8 briefly displays the different applications of polyurethane elastomers in the biomedical field.

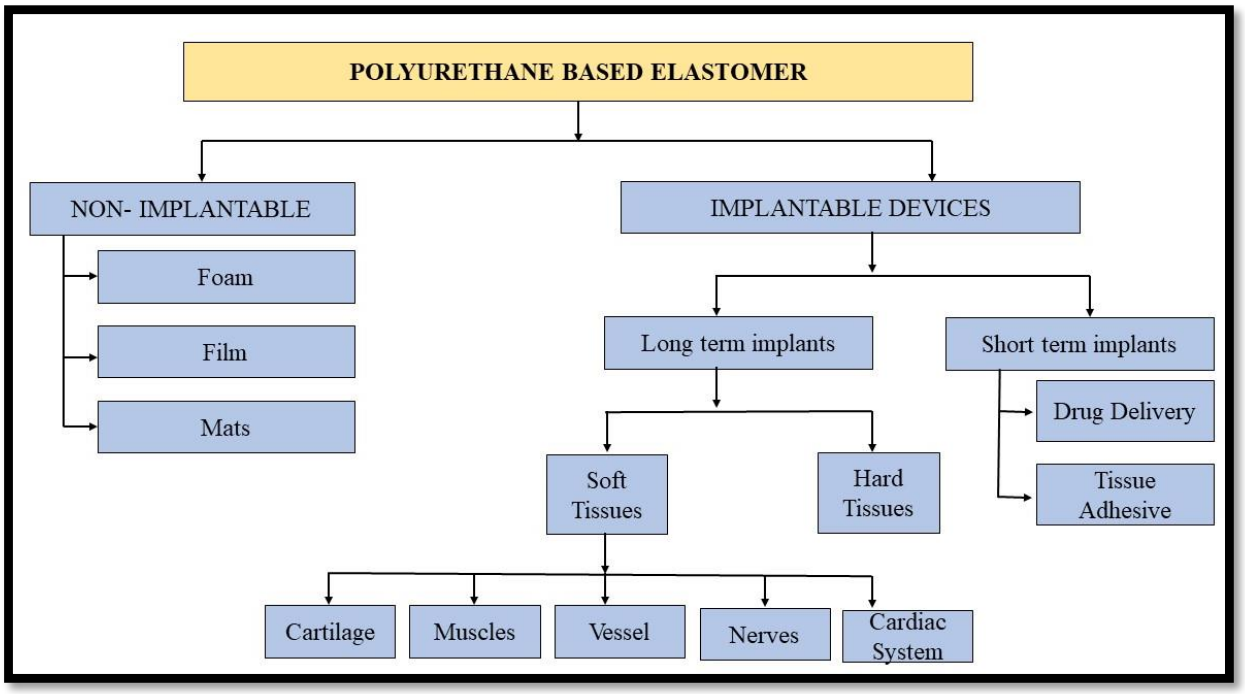



**Figure 8.** Classification of Polyurethane Elastomers in Biomedical Field.

#### 2.4.1. Non-Implantable Devices

Wound Dressing

Skin healing takes place in different time spans, depending on the extent of the case, according to the wound kind, infection degree, origin, depth, and health of the patient (diabetes, cancer). Wound exfoliation (removal of infected tissues), swabbing for infection, washing, and ultimately dressing with an acceptable system are all part of the chronic wound care standard; this stage is especially delicate, as an inappropriate substance might induce wound dryness or over-moisture, which can lead to bacterial infection [87].

Films as Wound Dressing Applications

Bio-based PU wound dressing membranes have been developed from vegetable oils. For example, a ricinoleic methyl ester-based prepolymer and methoxysilane (Si–OMe)-terminated castor oil prepolymer were used to make hybrid PU/siloxanes [88]. The introduction of a synthesised aniline tetramer resulted in significant electroactivity, antioxidant, and antibacterial properties [89].

Fibrous Mat for Wound Dressing Applications

Blending TPUs with biopolymers such as gelatin and/or chitosan before electrospinning has significantly improved the attributes such as biocompatibility, flexibility, and hydrophilicity [90] of wound dressing mats. Another method relies on the creation of a bi-layered mat prepared using gelatin electrospinning and electrospun TPU [91]. With $AgNO_3$, for example, antibacterial TPU-gelatin-chitosan blends have been created, where, the mechanical support was provided by PU, with gelatin and chitosan as supplementary components.

2.4.2. Implantable Devices

Long Term Implants

Soft Tissues

*Cardiac system:* The production of biomimetic cardiac patches with elastomeric qualities while preserving mechanical support for the tissue is the topic of investigation for cardiac system applications. Due to a dearth of effective solutions, the demand for bioactive and biocompatible cardiac patches for improved cell proliferation and adhesion is increasing. Many polycaprolactone (PCL)-based PUs are used in cardiac patches and cardiac scaffolds containing urethane and urea groups for better mechanical attributes, derived from castor oil-based polyols [92]. To promote cell electroactivity and proliferation, oligomeric aniline subunits were added [93]. The generated scaffolds in this case had increased conductivity as the aniline level increased. The integration of manufactured gold nanowires/nanotubes in a castor oil-based PU also led to a fruitful electrical stimulation of cardiomyocyte cells for improved proliferation, adhesion, and subsequent heart tissue regeneration [94].

*Cartilage:* Cartilage tissues need better mechanical strength and modulus. Supramolecular ionic bonds from alginate-based PU elastomers with configurable elevated mechanical performance have been developed for this application [93]. The Young moduli ranged from 14 to 93 MPa. In vivo, the scaffolds degraded slowly, which was equivalent to PCL-based scaffolds. The preparation of water-based PU micelles has also been tried in conjunction with the synthesis of biodegradable PU in the form of porous scaffolds or ionic nanoparticles [95]. Researchers proceeded further by adding chondrogenic induction factors and blending the PU with hyaluronan solution for improved cartilage healing [96].

*Muscles:* For muscular applications, biobased TPU films based on biobased polyesters such as polylactic acid (PLA) [97] or polyhydroxyalkanoates (PHA) derivates have lately been prepared. Novel polyurethanes made from Hexamethylene Diamine Diisocyanate (HMDI), poly (3-hydroxybutyrate-co-4-hydroxybutyrate) (P3/4 H B) diol, and PLA diol were produced in a recent work [88]. Owing to a more regular structure compared to arbitrary block polymerization, porous scaffolds from PULA-alt-3/4 H B exhibit superior hemocompatibility, and shape recovery hydrophilicity, in vivo and in vitro cytocompatibility, which makes it a better candidate for muscle tissue scaffolds [88].

*Nerves:* Treatment options for peripherical nerve injuries is end-to-end interconnection if the space is less than 2–3 cm, and autologous graft for a longer gap. Biobased and biodegradable nerve conduit scaffolds have now been developed to direct the nerve's repair through a tubular graft to replace autologous grafts. Owing to their excellent biocompatibility, tunable mechanical strength, biodegradability, and flexibility, polyurethanes are the most competitive materials. WPU films and a porous polyurethane network were prepared from a reported formulated [95] and employed in various biological applications. The inner, outer surfaces (9 and 42 m, respectively) and cross-section (23 m) of a porous PU scaffold showed pore interconnectivity and unequal pore diameters, which are reportedly found to be more effective than symmetrical pores, at draining wound inflammation waste during the early stages of neuron regeneration [98], thus, resulting in faster nerve regeneration. Porous networks showed superior permeability of a model bovine serum albumin (BSA) solution compared to PU films [99], implying effective mass transport for improved nerve healing.

*Vessels:* Currently, autogenous grafts are the most common procedure for vascular replacement, much as they are for hard tissue repair. Some individuals, however, lack sufficient autogenous tissues, necessitating the use of synthetic vascular grafts. Nonetheless, the development of tiny vessel scaffolds is required for improved compatibility and tissue regeneration [100]. A lysine triisocyanate PEG prepolymer was combined with an earlier synthesised polyester triol composed of glycolide, glycerol, D, L lactide, and caprolactone to create biobased PUs networks [101].

Hard Tissues

Even though the physicochemical and mechanical qualities of scaffolds are acceptable for bone tissue growth, and bone tissue regeneration; using PU scaffolds can have drawbacks such as inadequate cellular adhesion, differentiation, and biomineralization; this could be due to some breakdown products causing large pH changes in the scaffold microenvironment. It has been documented that a pH of around 7.40, which is around the physiological value, promotes osteoblast proliferation and differentiation [102]. Another disadvantage of PUs is their overall lack of bioactivity in specific tissue applications such as bone tissue regeneration. The scaffold prepared from glycerol-modified CO had better mechanical properties than the CO-based scaffold, with an elastic modulus of up to 165.36 MPa, probably due to a reduced number of hanging and unreacted end-groups. Good biocompatibility and osteogenic differentiation were achieved, as well as excellent bone matrix and trabecula regeneration. Another method is to graft bone ECM components to boost cell affinity and, as a result, promote high bone tissue regeneration [103].

Short-Term Implants

**Drug delivery:** Bio-based drug delivery systems have been developed with the goal of enhancing medication release by accelerating PU breakdown. It was accomplished by incorporating PLA [96] or PLGA [25] macromers into the PU backbone. By altering the quantity of PLA or molar mass in the final polyurethane, tailored degradation rates can be obtained [104]. Biodegradable PU with various soft segments are reported as efficient carriers of epirubicin [105]. Another method involves grafting PU onto biopolymers such as chitosan to create injectable hydrogels with controlled drug release, such as tetracycline hydrochloride [106]. The swellingbehaviour of pure Chitosan shows very high swelling (6000% of pure Chitosan), and it can be tuned nicely up to 200% by controlling its degree of substitution in PU or cross-link density through grafting. The first stimulus (such as temperature) normally allows these drug-loaded micelles to enter cancer cells, and the second stimulus (here, enzymatic attack) causes micelles to disassemble and release the drug. Other degradation stimuli have been investigated, with the goal of developing enzymatic intracellular sensitive bio-based NIPU micelles for the release of anticancer drugs [107]. From L-tyrosine amino acid resources, few amphiphilic poly-(ester-urethane)s were synthesized, and their self-assembled nanoparticles were used as drug delivery carriers for cancer therapy [108].

## 3. Non-Isocyanate Polyurethanes (NIPUs)

The isocyanates, as raw material in PU synthesis, are produced from hazardous and toxic phosgene by the process called phosgenation. Phosgene is a subtle toxin as the smell may not be observed and symptoms may be slow to arrive. It is known to be a widely used chemical weapon during the first world war accounting for a significant majority of fatalities. Moreover, the isocyanates themselves are toxic and moisture sensitive and cannot be prepared without sophisticated safety devices and massive investments.

PU foam production involves the generation of gas, usually, carbon dioxide, which may promote the release of isocyanates as vapours or aerosols, whose exposure may result in detrimental health effects such as asthma and skin irritation. The Bhopal Gas Tragedy of December 1984 in India was one of the deadliest industrial accidents in history that resulted in the deaths of thousands of people due to the methyl isocyanate gas leak; this chemical disaster attracted worldwide attention and demonstrated the immediate and hazardous

effects of isocyanates [109]. Isocyanates enter the body primarily through inhalation or skin exposure. The most extensively used isocyanates for polyurethane production; MDI (methylene diisocyanate) and TDI (toluene diisocyanate) are labeled as CMR (carcinogenic, mutagen, reprotoxic) and are categorized as "very harmful" chemicals by the European Community. Perpetual exposure to MDI and TDI vapors leads to irritation in the skin, eyes, and respiratory tract. Additionally, during the storage and formation processes of moisture curing polyurethanes, the reaction between the isocyanate and water is troublesome, as it may cause moisture curing. In such systems, carefulseparation of these materials from the water becomes exceptionally essential to avoid an irreversible reaction between the polyisocyanate component and water, forming $CO_2$ and urea, thus, resulting in an unusable, hardened product [109] (Figure 9).

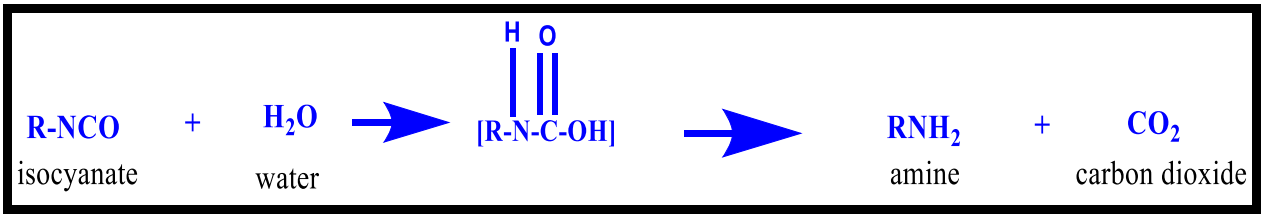



**Figure 9.** Reaction of Isocyanate with water [109].

Thus, modern researchers have recognized the need to eliminate the use of isocyanates and are looking for ways to employ environment-friendly alternatives. One such alternative is the synthesis of Non-isocyanate polyurethanes (NIPU), a new type of PU that is safer and more promising than conventional PUs, and is becoming a viable substitute for conventional PUs.

One example of the isocyanate-free routes of PU production is the synthesis of hydroxyurethane in the presence of cyclic carbonate and amines [110], which is regarded as the best alternative method to synthesize Non-isocyanate polyurethanes (NIPU) as shown in Figure 10. There are two isomers formed in the reaction, one containing a secondary hydroxyl group, the other having a methylol group, each containing urethane, and a hydroxyl group. These hydroxyl entities can develop intermolecular hydrogen bonds with the urethane group [109].

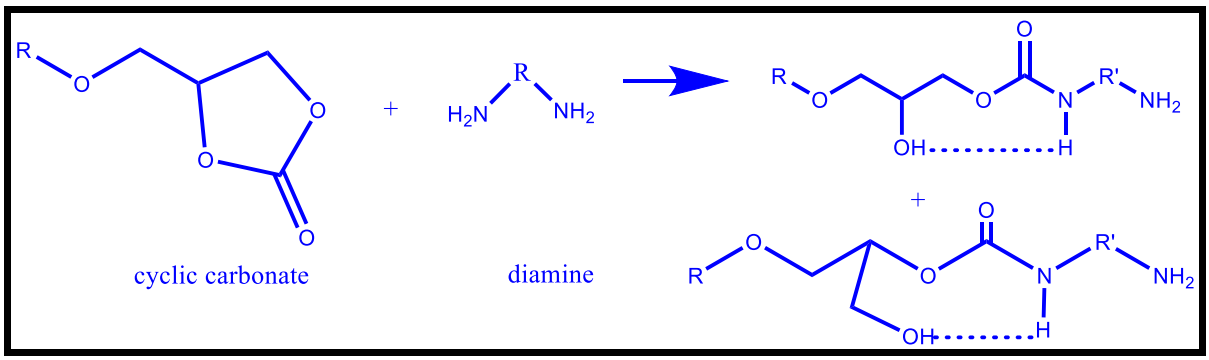



**Figure 10.** Reaction of Cyclic Carbonates and Amines [110].

Another important method of NIPU synthesis is Aminolysis, which is mainly a two-step procedure involving, (i). the development of the cyclic carbonate oligomers and then (ii) treating it with polyamines or amines resulting in the production of NIPU; as depicted in Figure 11. Step 1 includes the development of cyclic carbonate, i.e., one of the most efficient ways of producing cyclic carbonate, by introducing the $CO_2$ within the epoxide group using a suitable catalyst, followed by Step 2, i.e., Treating this cyclic carbonate oligomer

with a primary amine to provide NIPU, via an addition reaction after ring-opening. Thus, if the cyclocarbonate oligomer has only cyclocarbonate, then it is called NIPU. Moreover, if something more than cyclocarbonate is present, i.e., hydroxyl groups or epoxy groups, etc., then it is termed HNIPU, which has different properties than that acquired by NIPU [109]. This is the most broadly used technique for the synthesis of isocyanate-free polyurethanes; the procedure comprises the ring opening of a cyclic carbonate via polyamine, leading to the formation of hydroxypolyurethanes.

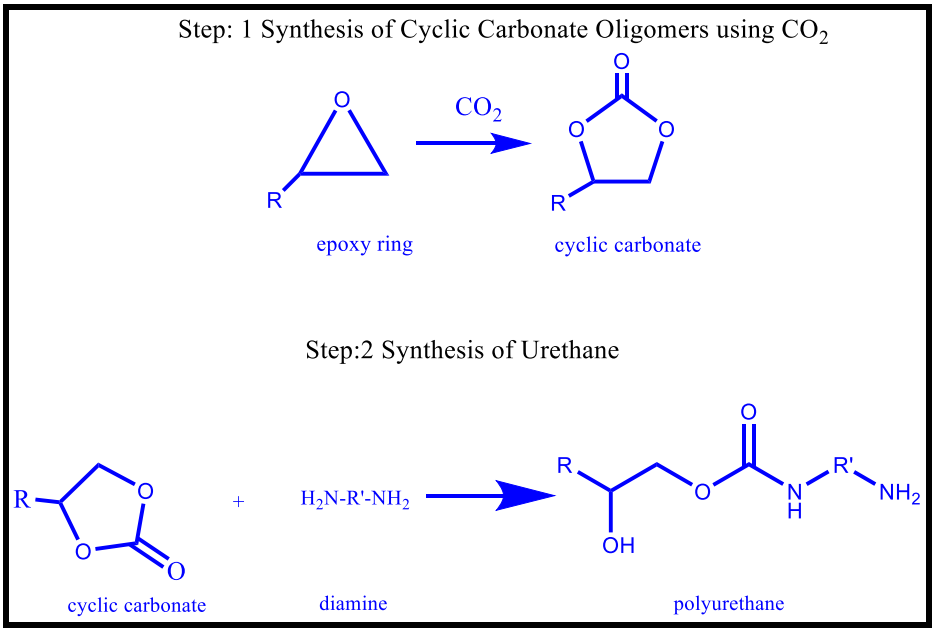



**Figure 11.** Aminolysis for the synthesis of NIPU [109].

**Merits of NIPU over PU:** PU finds remarkable applications in paints, coatings, adhesives, and inks; for this, they are generally prepared in organic solvents. Consequently, large concentrations of hazardous air pollutants (HAPs) and volatile organic compounds (VOCs) evaporate in the atmosphere. Exposure to VOCs can cause health effects such as dizziness, headaches, cancer, irritation, etc. NIPU has many advantages over conventional polyurethanes. Primarily, the elimination of the use of isocyanates. Besides this, the cyclic carbonates are derived from the natural sources available through epoxy resins and $CO_2$. Carbon dioxide is a cheap starting material and its utilization for PUs production would also definitely help to trim down the Global Warming, by reducing $CO_2$, which is a Greenhouse gas. Thus, the two problems can be solved at once- firstly the shortage of oil resources, and secondly, the greenhouse effect. So overall it stands by the laws for environmental protection [111].

*Applications of NIPU*

**NIPU Coatings and Adhesives:** The applications of NIPU include especially chemical-resistant coatings and adhesives. In addition to the numerous applications of PUR in the production of coatings and in polymer industries. In addition to thermosetting coatings and UV-curable coatings, NIPU systems are capable of being used for monolithic flooring. Bio-sourced NIPU coatings and adhesives from mono(glucose) and disaccharides (sucrose) have been reported. It was also observed that the glucose-based NIPU displayed good surface coating and thermosetting joint adhesives [110].

Fire-resistant hybrid coatings made from inorganic material and PHUs have also been reported to have better properties, such as chemical resistance or fire resistance, or processability [112]. NIPU technology can also be employed in gasoline-stable and hydrolysis-stable sealants for safeguarding electronic devices and their associated machin-

ery in rocket construction and aircraft and for civil engineering applications; glues with high adhesion strength and longevity. Lately, a UV curable PU resin synthesized using NIPU chemistry using cyclic carbonate of bisphenol-based epoxy resin and polyether amine was used for hydrophilic textile treatment. NIPU developed from plant sources has gained vast popularity recently. Moritz Bahr et al., studied linseed and soy-oil-based NIPU, synthesized by curing the carbonated linseed and carbonated soybean oil with different diamines such as ethylene diamine; they revealed a new-route to prepare the crosslinked as well as linear terpene-based non-isocyanate poly(hydroxypolyurethanes) (NIPU) and prepolymers obtained from novel cyclic limonene dicarbonate [113]. Bio-based NIPUs synthesized from carbonated sunflower oil have been studied by Simanta Doley and Swapan K. Dolui [110]. Carbonated vegetable oil was derived from a solvent-free reaction mixture of carbon dioxide with epoxidized sunflower oil. Javani et al. [114] documented that aromatic amine can be utilized as hardeners for various NIPU systems. P-xylenediamine, m-xylenediamine, and aliphatic/cycloaliphatic diamines were used by them for curing the reaction of cyclic carbonated soyabean oil. Xylene diamine led to increased tensile strength, with an inferior elongation in comparison to aliphatic diamine [114].

NIPU-based products can also be employed as sealants, owing to their brilliant adhesive properties. Dendro-aminosilane hardeners can also be used in NIPU sealant production, which enhances their thermal characteristics. High adhesion to glass and steel and high impact resistance can also be acquired based on organosilanepentacyclo-carbonate [115].

## 4. Recycling and Disposing of Polyurethanes

Polyurethane is the world's sixth most widely used polymer, with a yearly production of 18 million tonnes, equating to each day's production of PU speciality of more than 1 million $m^3$. As a coordinate effect of its commercial victory, a plethora of waste is produced as a direct result of its commercial achievement and its management is an environmental and pressing concern. Land filling was once the quick fix to the obstacle; however, because of the long life cycle of polyurethanes (>10 years), limited knowledge is available about their activity in landfills. One of the problems confronting the current world is the recycling of any type of plastic in order to transform waste into profitable items. Furthermore, if the recycling procedure is green within itself, that would be an incredible accomplishment; this fact, combined with new environmental standards, makes the development of environmentally viable recycling processes critical.

Primary and secondary recycling are the two types of physical recycling processes, that are commonly used. These methods do not alter the structural features of the polymer. Polymeric waste is converted into powder, flakes, or granules for utilization in new material production in physical processes. The following are the most important physical recycling procedures for polyurethane:

1. Rebonding: Flexible PU foam is cut into smaller pieces and it is used in the manufacture of sports mats and carpet lay [115].
2. Regrind or Powdering: Powdered PU waste is blended with one of the virgin reagents (usually with the polyol up to 30% wt) to build new PU goods [115].
3. Compression Molding: Powdered PU waste is exposed to high heat and pressures in the mold. It can enable upto approximately 100% recycled substance to be obtained [115].

These physical procedures are beneficial with polymers that are thermoplastic, but due to the thermostable nature of thermoset PU, they are ineffective with them. The process adopted, from the initial efforts of re-purposing polyurethane (PU) leftovers, as a cushion filler to one of the most recent chemical processes involving recycling agents which are green and sustainable. Chemical recycling enables the production of fundamental hydrocarbon entities called monomers (tertiary recycling) which could be utilized as production resources in the petrochemical industry and chemical synthesis materials. Using this method, it is conceivable to attain items with additional value.

### *4.1. Hydrolysis*

The earliest chemical approach for recycling polyurethane waste, particularly flexible foams, was hydrolysis. It is the result of a reaction between polyurethane waste and water, which can be steam or liquid [116]. Amine intermediates, carbon compounds, and Polyols, are amidst the end products (Figure 12).

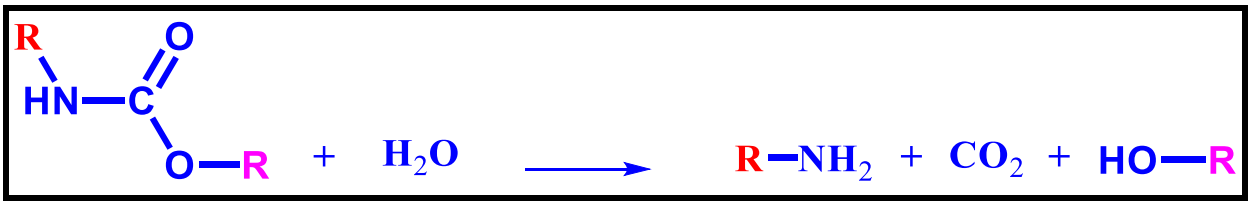



**Figure 12.** Hydrolysis of Polyurethane [117].

It is essential to note that the hydrolysis technologies have merely been ramped up to pilot plant scale, and they've never made it to the commercial level because of the large energy and pressure input demand in the reactor [117].

### *4.2. Aminolysis*

Aminolysis is a sort of transesterification reaction during which the amine group replaces the ester group attached to the carbonyl carbon of the carbamate linkage of urethane [115], as shown in Figure 13. Products containing polyurethanes are generally crushed or dissolved in a specific solvent, such as cyclic ether or a nitrogen-containing chlorinated hydrocarbon solvent; they are then aminolyzed with a molecule that contains at least one $–NH_2$ group; this reaction takes place at a temperature of 80–190 °C, and it is widely practiced with catalysts such asaluminium hydroxide, sodium methoxide, and sodium hydroxide. Polyfunctional alcohols and amines can be produced according to the diamine or amino alcohol employed; they can be used to synthesize polycarbonates, melamine resins, polyurethanes, polyesters, epoxy resins, etc. Two or more of the resulting functional phenols can also be used to recover the associated isocyanates; this method has also not gained any commercial success yet.

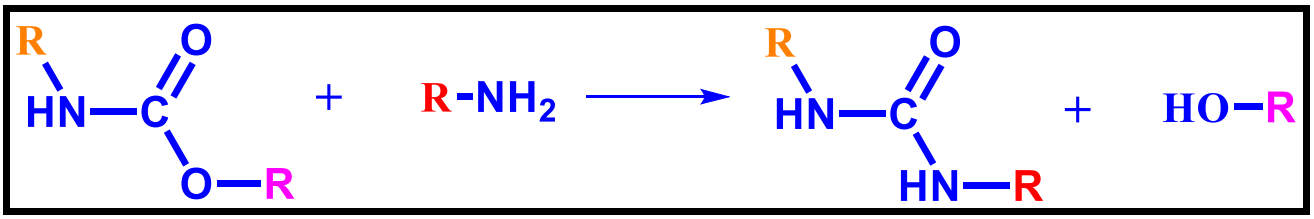



**Figure 13.** Aminolysis of Polyurethanes [115].

### *4.3. Phosphorolysis*

Phosphorolysis is a process comparable to hydrolysis where esters of phosphoric acids or phosphonic display similar behaviour as water with the production of phosphate (Figure 14); this method allowed a combination of phosphorus-containing oligo-urethanes to be created. Because of the phosphate functionality of the recovered product, phosphorolysis can provide glycolyzates that can be employed in the manufacture of polyurethanes with flame retardant characteristics; this approach has the privilege of being able to create novel polyurethanes with enhanced adhesive characteristics, flame retardancy (comparable to industrial goods with larger concentrations of flame retardants), and UV resistance [117].

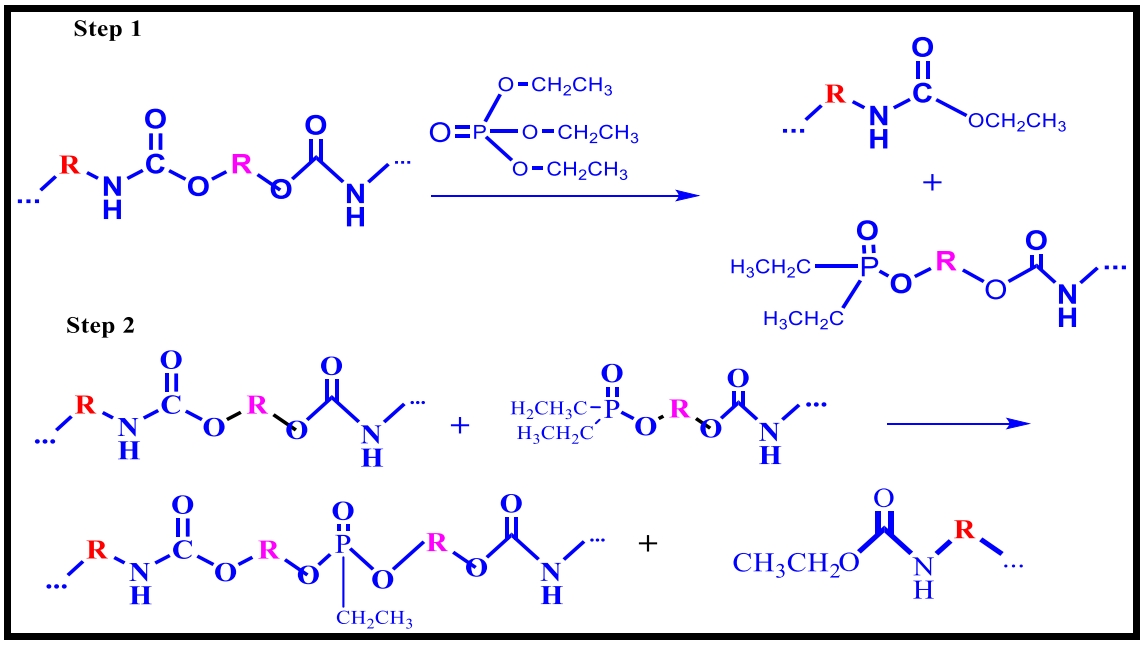



**Figure 14.** Phosphorolysis of Polyurethane [117].

*4.4. Glycolysis*

The most extensively utilised chemical recycling procedure for PU is glycolysis. It comprises a transesterification reaction wherein a glycol's hydroxyl group replaces the ester group attached to the carbonylic carbon of the urethane [118]. The primary reaction is as shown in Figure 15.

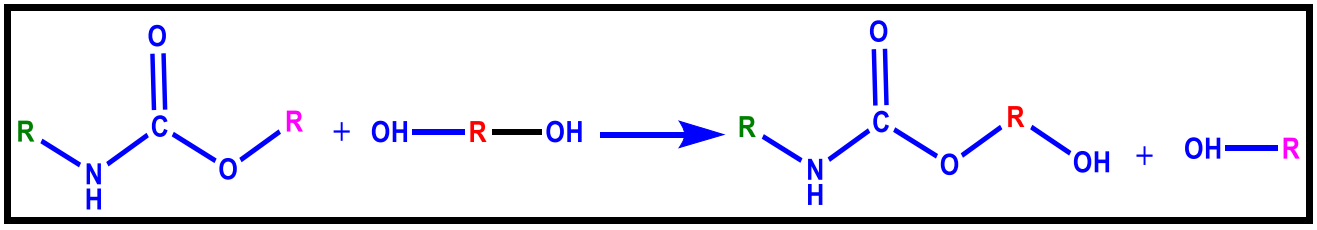


**Figure 15.** Glycolysis of Polyurethanes [118].

This reaction creates polyols with properties that can be tweaked to a few degrees and are potentially similar to the original substance; they can be used to manufacture polyurethanes. Glycolysis can go in two different directions. The first allows polyols to be recovered and used to make flexible PU foam. Split-phase glycolysis (SPG), on the other hand, produces flexible or rigid polyols. SPG is built on diethylene glycol (DEG) glycolysis processing of the bottom and top layers separately. The upper layer, which primarily consists of flexible polyols, is washed with DEG, while the lower layer is allowed to react with propylene oxide to produce rigid polyols. The biggest impediment of glycolysis is the disparity in process parameters between flexible and rigid foams, which forces a waste separation. Due to the great susceptibility of the process to the presence of contaminants that commonly exist in consumer trash, this approach is also substantially more successful when employed in post-production waste [118]. Table 5 reviews the glycolysis of various polyurethane materials. The following agents contribute significantly toward the glycolysis of polyurethane materials:

1. Catalysts: The catalysts engaged are amines such as diethanolamine and ethanolamine; alkaline acetates such as sodium, potassium, or lithium acetate; hydroxides such

as potassium or sodium hydroxides; metallic octoates such as stannous or lithium octoates and organometallic compounds such as titanium butoxide [115].

2. Cleavage Agents: DPG, DEG, MPG, and MEG, and are some of the most commonly utilised cleavage agents, with DEG being the most common; however, crude glycerol is now the most assuring [115].
3. Glycolysis agent: PU ratios: In transesterification reaction, the most frequent glycolysis agent: PU ratios used. It can be seen that the glycolysis agent to PU ratio is often more than 1, in order to ensure optimal split-phase glycolysis and, as an outcome, a good quality glycolyzate that may be used in the generation of a fresh PU product [115].
4. Temperature: the best temperature for carrying out the glycolysis reaction is between 160 and 250 °C, because cooler temperatures result in a slow recovery course and a low polyol proportion, whereas higher temperatures result in excessive evaporation of cleavage agent and a rise in the extension of the secondary reactions [115].

**Table 5.** Glycolysis of Various Polyurethane Materials.

| PU Type | Elastomer | Coating | Rigid Foam | Flexible Foam |
|---|---|---|---|---|
| Catalyst | DEA, EA, LiAc | KAc, NaOH, | BaO, DEA, EA, NaOH, KOH, NaAc, Dibutyltindilaurate, ZnAc2 | KAc, $ZnAc_2$ DEA, TiBut, SnOct, LiOct |
| Cleavage agent | Diethylene glycol Monopropylene glycol Monoethylene glycolPoly(ethylene glycol) Triethylene glycol | Monoethylene glycol | Dipropylene glycol, Tetraethylene glycol, Diethylene glycol | Diethylene glycol, Monoethylene glycol, Dipropylene glycol, Monopropylene glycol, Poly(ethylene glycol), Triethylene glycol, Pentane 1,5 diol, Glycerol, Butane 1,4 diol, Polyethylene glycol 400, 1,6-hexane diol, Crude glycerol |
| Glycol: PU (wt.%) | 2:1 | 1:10–1:2 | 1:1–2:1 | 1:10–2:1 |
| Temperature range (°C) | 160–190 | 190–240 | 170–200 | 160–250 |
| Potential Market for recovered products | Adhesives | Coatings | Rigid foams | Flexible foams, Rigid foams, Adhesives, Elastomers |

### *4.5. Gasification*

Gasification is the partial oxidation of carbonate compounds which is a highly exothermic process, resulting in the formation of Syngas (a combination of mostly hydrogen and carbon monoxide) and ash. One of the most notable benefits of gasification is that it eliminates the requirement for waste segregation. Polyurethanes that have been combined with some other raw materials can also be employed in the course of action. The profitability of this method, nevertheless, completely depends on the probable use of syngas asraw material and a source of energy for the production of acetic acid, ammonia, methanol, carbohydrates, and other chemicals [119].

### *4.6. Pyrolysis*

Pyrolysis is the anaerobic, high-pressure thermal disintegration of complex polymeric chains into simpler molecules [120]. According to a study, polyurethane pyrolysis occurs in at least two stages, which corresponds to the sequential thermal decomposition of the isocyanate and polyol parts of the polyurethane. Over half of the polymer weight is lost in the first stage, which runs at temperatures lying in the range of 100–300 degrees Celsius. The next stage of degradation begins between 300 and 800 degrees Celsius. The amount of ash left over after pyrolysis is even less than 3% of the original polyurethane [121]. Pyrolysis has undeniable advantages, including the little quantity of waste produced and the ability

to use the resultant chemicals in other petrochemical operations. Unfortunately, due to the failure to define the ratio of individual items and the complexity of getting products with the appropriate qualities, this is not always practicable. The resultant gas has a high calorific value and can be used to heat and power gas engines. Furthermore, gas products carry hazardous chemicals such as aniline, benzene, and hydrogen cyanide [117].

### *4.7. Hydrogenation*

Pyrolysis and gasification are both processes that are analogous to hydrogenation. Gas and oil are formed as a result of this process. The use of high-pressure hydrogen rather than inert gas is the primary distinction between this technique and pyrolysis. This strategy relies on addressing two key issues: the composition of the obtained gas and oil, as well as the cost of converting them into functional products that may be utilised as substrates and energy materials in chemical methods. Only gasification and glycolysis have been executed on a wide scale among chemical processes of polyurethane recycling, while others are still in the research phase. Furthermore, just glycolysis results in raw material recovery [122].

## 5. Biological Degradation of Polyurethane

The decomposition of organic compounds by various living organisms or their enzymes is referred to as biodegradation. It leads to the shortening of polymeric chains and the eradication of some of their components leading to the decrease in its molecular mass, and under the appropriate conditions, it can also lead to total mineralization of deteriorated material. Since the process does not demand elevated temperatures or sophisticated chemicals, biodegradation is generally more environmentally pleasant than chemical degradation. It can also be used for the decomposition of post-consumer waste. Bacterial biodegradation, fungal biodegradation, and enzymatic degradation are the three types of polyurethane modifications. Polyester polyurethanes are substantially more biologically degradable than polyether polyurethanes. The use of fungal strains in research yields more promising outcomes for polyether ones. Mechanical cracking of more obstinate polyether urethanes, which are commonly used in foams, can be linked to mycelium accessing material pores and causing mechanical cracking. On the other side, microorganisms are used in the majority of PU coating degradation studies. It is possible that this is due to bacteria's ability to develop biofilm on the sleek coated surface.

### *5.1. Degradation by Fungi*

Fungi are microbes that are recognised to be the primary cause of PU degradation in experimentally tested species. The majority of studies focus on soil microorganisms involved in polyurethane breakdown [123,124]. Irrespective of isolation circumstances, soil source, or polyurethane employed by scientists, the majority of isolated species that exhibit promising results in PU breakdown fall into one of a few categories such as Aspergillus, Cladosporium, Chaetomium, Penicillium, and Trichoderma; however, most of the fungi responsible for degrading polyurethane are filamentous, with only a few yeasts yielding positive results [125]; this disparity could be due to abiotic outcomes of fungal biodegradation. Growing filaments enter the material, causing the pore size to expand, perhaps leading to material fracture; however, the mechanism of this course, as well as the role of abiotic and biotic components of fungal degradation, remain unknown. All types of polyurethanes, including coatings, foams, polyether, polyester, and thermoplastic PU, have been found to be biodegradable by fungus strains and communities [117].

### *5.2. Degradation by Bacteria*

Though fungi are responsible for the majority of the encouraging outcomes, some studies point to bacteria as possible polyurethane degrading microorganisms. *Acinetobactergerneri*, *Bacillus subtilis*, *Comamonasacidovorans*, and Pseudomonas chlororaphis were found as microbes capable of degrading polyester polyurethane Impranil DLN in research [117].

*5.3. Degradation by Enzymes*

The fundamental benefit of enzymatic degradation over microbial degradation is the ability to manage bond breakage, which results in the synthesis of building blocks that can be recycled or used as a substrate in the manufacture of other materials. The key problem in this approach is the wide variety of raw materials utilised in polyurethane synthesis, which necessitates treating each type of PU separately.

Figure 16 summarises the degradation of different polyurethanes by various enzymes.

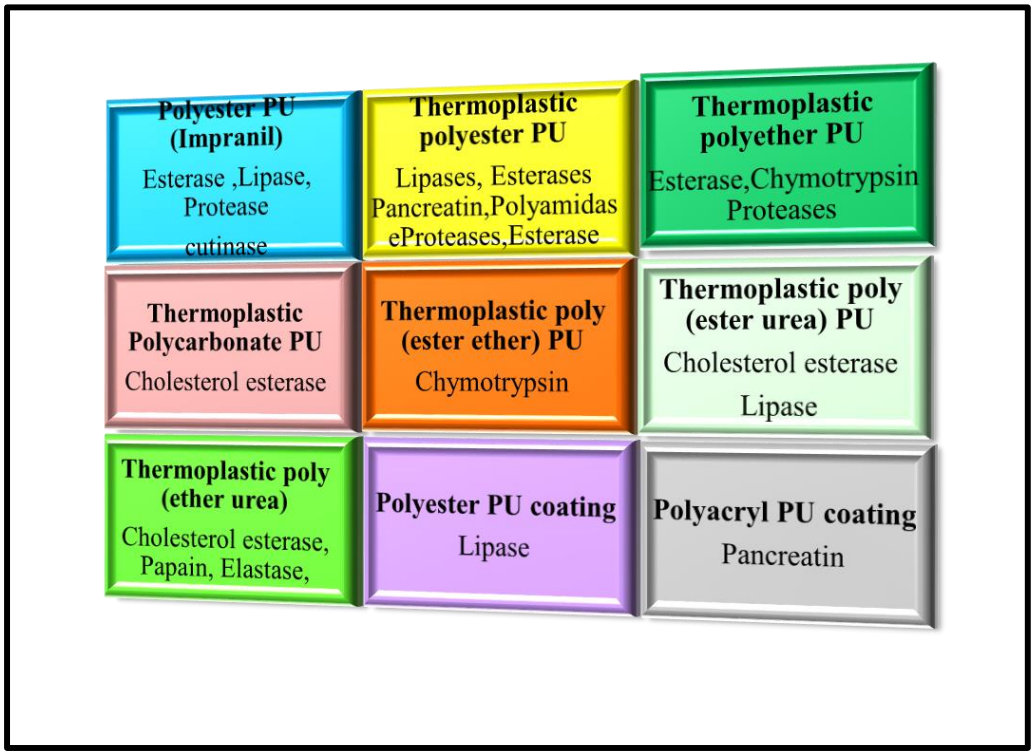



**Figure 16.** Degradation of Different Polyurethanes by Various Enzymes.

The hydrolysis of the polyester component of polyester-based polyurethanes by esterases (Figure 17) is responsible for the majority of favourable biodegradation outcomes. According to certain research, esterases are also accountable for the hydrolysis of urethane linkages, resulting in alcohol chain-ends and carbamic acid; however, because acid is unstable and dissociates quickly, the release of carbon dioxide and amine is a more likely event. Furthermore, because most polyurethane degradation assays use polyester types, they do not distinguish between urethane hydrolysis and ester bond hydrolysis. Researchers studying urease degradation of polyurethane encounter a similar issue, but the majority of their observations are favourable as long as the polyurethane contains urea linkages. Those linkages could be a piece of the soft section or the result of water foaming. Those in long, stretchy chains, on the other hand, are much simpler to disintegrate. The generation of alcohol, amine, and the release of carbon dioxide, proved to be extremely fruitful for urethane bond hydrolysis by peptidases and amidases [122].

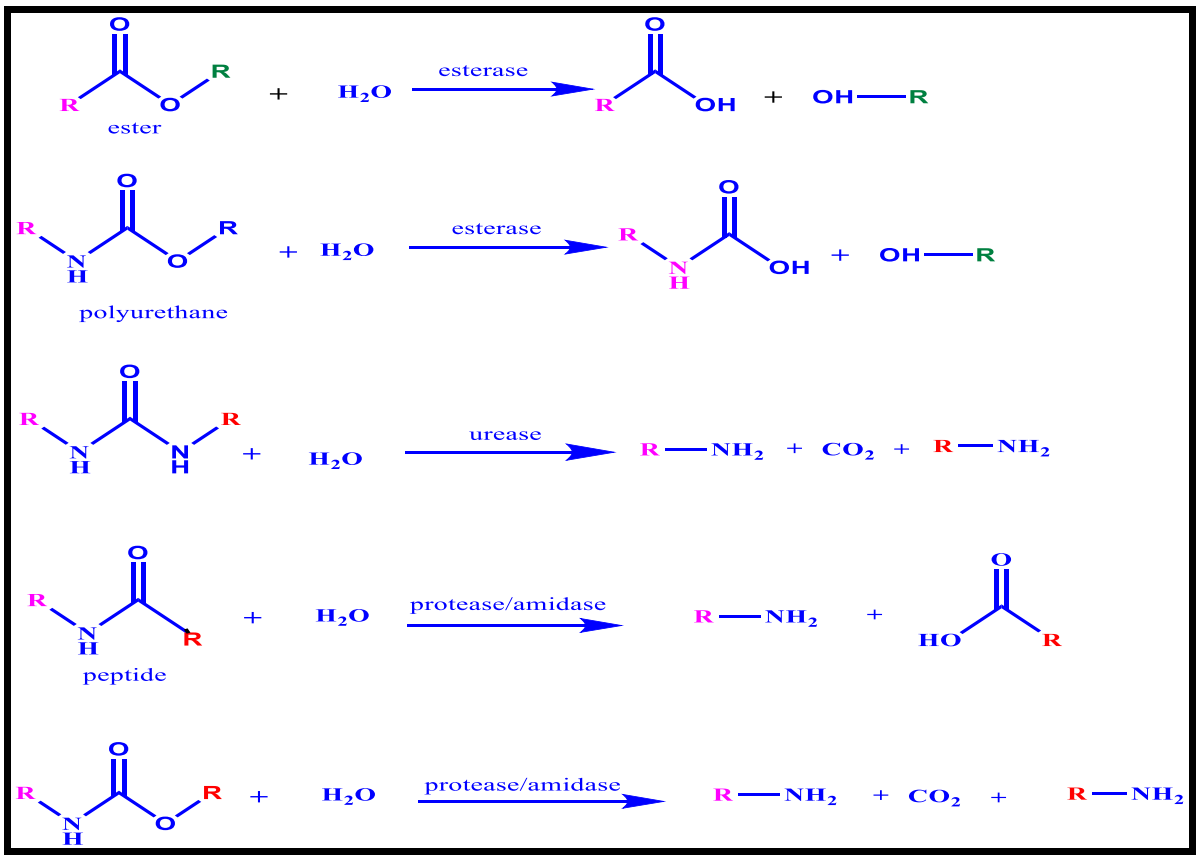



**Figure 17.** Enzymatic Degradation of Polyurethane.

## 6. Conclusions

Attempts have been devoted in the recent past to manufacture green PU materials from renewable resources to satisfy the specified qualities for social, technological, and economic criteria. Different compound strategies for the integration of PU have been offered to study the use of renewable resources rather than petroleum base requirements in both research and industry. The bio-based PU's promise as a long-term contender for upgrading bio-based derivatives via different processes. The bio-based polyurethanes have significant potential to be used as future PU products, with a partial or total replacement of petroleum-based polyurethanes; however, the literature studies have revealed a few problems, such as, the low bio-based content, lack of comprehension of individual formulations and a lack of insight into catalyst toxicity and degradation products. Nonetheless, the overall energy/cost imbalance is an open difficulty associated with these reaction processes; this gave rise to the sustainable alternative for the conventional polyurethanes i.e., Non-Isocyanate polyurethanes. The NIPUs are proving to be very efficient in a diverse range of applications owing to their hydrolytic stability, moisture insensitivity, and green pathway. To deal with the problem of post-consumption PU wastage, different recycling techniques are employed and reviewed in detail. Special emphasis has been given to the biodegradation of polyurethane wastes. Biological degradation requires a moderate temperature and does not necessitate the use of toxic chemicals, but there is still a long journey to go before it can be employed on a large basis. Biodegradation, in particular, is the most attractive due to a large number of available options and accessible alterations. By employing suitably altered building blocks while preparing, PU can be made 100% degradable; this approach could be a realistic alternative to deal with the accumulation of polyurethane waste. In nutshell, the path of polyurethane synthesis progresses from hazardous to non-toxic, and finally to sustainable.

**Author Contributions:** Conceptualization, R.K.; Methodology, P.S.; Validation, R.K.; Formal Analysis, S.T.; Investigation, P.S.; Resources, S.Y.; Data Curation, S.T.; Writing—original Draft Preparation, P.S. and G.V.; Writing—Review and Editing, G.V. and S.Y.; Supervision, R.K.; Project Administration, R.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to thank Delhi Technological University for providing technical and financial assistance.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Deng, Y.; Dewil, R.; Appels, L.; Ansart, R.; Baeyens, J.; Kang, Q. Reviewing the thermo-chemical recycling of waste polyurethane foam. *J. Environ. Manag.* **2020**, *278*, 111527. [CrossRef]
2. Akindoyo, J.O.; Beg, M.D.H.; Ghazali, S.; Islam, M.R.; Jeyaratnam, N.; Yuvaraj, A.R. Polyurethane types, synthesis and applications-a review. *RSC Adv.* **2016**, *6*, 114453–114482. [CrossRef]
3. Cornille, A.; Guillet, C.; Benyahya, S.; Negrell, C.; Boutevin, B.; Caillol, S. Room temperature flexible isocyanate-free polyurethane foams. *Eur. Polym. J.* **2016**, *84*, 873–888. [CrossRef]
4. Thébault, M.; Pizzi, A.; Essawy, H.A.; Barhoum, A.; Van Assche, G. Isocyanate free condensed tannin-based polyurethanes. *Eur. Polym. J.* **2015**, *67*, 513–526. [CrossRef]
5. Kulkarni, R.D.; Paraskar, P.M. Synthesis of Isostearic Acid/Dimer Fatty Acid-Based Polyesteramide Polyol for the Development of Green Polyurethane Coatings. *J. Polym. Environ.* **2021**, *29*, 54–70.
6. Cheng, Z.; Li, Q.; Yan, Z.; Liao, G.; Zhang, B.; Yu, Y.; Yi, C.; Xu, Z. Design and synthesis of novel aminosiloxane crosslinked linseed oil-based waterborne polyurethane composites and its physicochemical properties. *Prog. Org. Coat.* **2018**, *127*, 194–201. [CrossRef]
7. Kovács, E.; Turczel, G.; Szabó, L.; Varga, R.; Tóth, I.; Anastas, P.T.; Tuba, R. Synthesis of 1,6-Hexandiol, Polyurethane Monomer Derivatives via Isomerization Metathesis of Methyl Linolenate. *ACS Sustain. Chem. Eng.* **2017**, *5*, 11215–11220. [CrossRef]
8. Liang, H.; Feng, Y.; Lu, J.; Liu, L.; Yang, Z.; Luo, Y.; Zhang, Y.; Zhang, C. Bio-based cationic waterborne polyurethanes dispersions prepared from different vegetable oils. *Ind. Crops Prod.* **2018**, *122*, 448–455. [CrossRef]
9. Gomez-Jimenez-Aberasturi, O.; Gomez, J.R. New approaches to producing polyols from biomass. *J. Chem. Technol. Biotechnol.* **2017**, *92*, 705–711. [CrossRef]
10. Wang, Y.; Deng, L.; Fan, Y. Preparation of Soy-Based Adhesive Enhanced by Waterborne Polyurethane: Optimization by Response Surface Methodology. *Adv. Mater. Sci. Eng.* **2018**, *2018*, 9253670. [CrossRef]
11. Zuber, M.; Shah, S.A.A.; Jamil, T.; Asghar, M.I. Performance behavior of modified cellulosic fabrics using polyurethane acrylate copolymer. *Int. J. Biol. Macromol.* **2014**, *67*, 254–259. [CrossRef]
12. Sharmin, E.; Zafar, F.; Akram, D.; Alam, M.; Ahmad, S. Recent advances in vegetable oils based environment friendly coatings: A review. *Ind. Crop. Prod.* **2015**, *76*, 215–229. [CrossRef]
13. Vanbésien, T.; Monflier, E.; Hapiot, F. Hydroformylation of vegetable oils: More than 50 years of technical innovation, successful research, and development. *Eur. J. Lipid Sci. Technol.* **2015**, *118*, 26–35. [CrossRef]
14. Paraskar, P.M.; Prabhudesai, M.S.; Hatkar, V.M.; Kulkarni, R.D. Vegetable oil based polyurethane coatings—A sustainable approach: A review. *Prog. Org. Coat.* **2021**, *156*, 106267. [CrossRef]
15. Peyrton, J.; Chambaretaud, C.; Sarbu, A.; Avérous, L. Biobased Polyurethane Foams Based on New Polyol Architectures from Microalgae Oil. *ACS Sustain. Chem. Eng.* **2020**, *8*, 12187–12196. [CrossRef]
16. Devi, K.; Nurul, P.P.; Ain, H.; Tuan Noor Maznee, T.I.; MohdNorhisham, S.; Srihanum, A.; Norhayati, M.N.; Yeong, S.K.; Hazimah, A. *Green Polyurethane for Ornamental Products*; MPOB Information Series, MPOB TT No. 522; MPOB: Bandar Baru Bangi, Malaysia, 2012.
17. De Souza, F.M.; Kahol, P.K.; Gupta, R.K. Introduction to Polyurethane Chemistry. *ACS Symp. Ser.* **2021**, *1380*, 1–24.
18. KhoonPoh, A.; Choy Sin, L.; Foon, C.; Hock, C. Polyurethane wood adhesive from palm oil-based polyester polyol. *J. Adhes. Sci. Technol.* **2014**, *28*, 1020–1033. [CrossRef]
19. Ang, K.P.; Lee, C.S.; Cheng, S.F.; Chuah, C.H. Synthesis of palm oil-based polyester polyol for polyurethane adhesive production. *J. Appl. Polym. Sci.* **2013**, *131*, 6. [CrossRef]
20. Li, N.; Qi, G.; Sun, X.S.; Stamm, M.J.; Wang, D. Physicochemical Properties and Adhesion Performance of Canola Protein Modified with Sodium Bisulfite. *J. Am. Oil Chem. Soc.* **2011**, *89*, 897–908. [CrossRef]
21. Aung, M.M.; Yaakob, Z.; Kamarudin, S.; Abdullah, L.C. Synthesis and characterization of Jatropha (Jatropha curcas L.) oil-based polyurethane wood adhesive. *Ind. Crops Prod.* **2014**, *60*, 177–185. [CrossRef]

22. Saalah, S.; Abdullah, L.; Aung, M.; Salleh, M.; Biak, D.A.; Basri, M.; Jusoh, E.; Mamat, S.; Al Edrus, S.O. Chemical and Thermo-Mechanical Properties of Waterborne Polyurethane Dispersion Derived from Jatropha Oil. *Polymers* **2021**, *13*, 795. [CrossRef]
23. Wang, C.-S.; Yang, L.-T.; Ni, B.-L.; Shi, G. Polyurethane networks from different soy-based polyols by the ring opening of epoxidized soybean oil with methanol, glycol, and 1,2-propanediol. *J. Appl. Polym. Sci.* **2009**, *114*, 125–131. [CrossRef]
24. Valero, M.; Gonzalez, A. Polyurethane adhesive system from castor oil modified by a transesterification reaction. *J. Elastomers Plast.* **2012**, *44*, 433–442. [CrossRef]
25. Wang, Z.; Yu, L.; Ding, M.; Tan, H.; Li, J.; Fu, Q. Preparation and rapid degradation of nontoxic biodegradable polyurethanes based on poly(lactic acid)-poly(ethylene glycol)-poly(lactic acid) and l-lysine diisocyanate. *Polym. Cem.* **2011**, *2*, 601–607. [CrossRef]
26. Das, S.; Pandey, P.; Mohanty, S.; Nayak, S.K. Evaluation of biodegradability of green polyurethane/nanosilica composite synthesized from transesterified castor oil and palm oil based isocyanate. *Int. Biodeterior. Biodegrad.* **2017**, *117*, 278–288. [CrossRef]
27. Rodriguez, R.; Pérez, B.; Florez, S. Effect of Different Nanoparticles on Mechanical Properties and Curing Behavior of Thermoset Polyurethane Adhesives. *J. Adhes.* **2014**, *90*, 848–859. [CrossRef]
28. Mirabedini, S.; Mohseni, M.; PazokiFard, S.; Esfandeh, M. Effect of $TiO_2$ on the mechanical and adhesion properties of RTV silicone elastomer coatings. *Colloids Surf. A Physicochem. Eng. Asp.* **2008**, *317*, 80–86. [CrossRef]
29. Alinejad, M.; Henry, C.; Nikafshar, S.; Gondaliya, A.; Bagheri, S.; Chen, N.; Singh, S.K.; Hodge, D.B.; Nejad, M. Lignin-Based Polyurethanes: Opportunities for Bio-Based Foams, Elastomers, Coatings and Adhesives. *Polymers* **2019**, *11*, 1202. [CrossRef]
30. Magalhães, S.; Alves, L.; Medronho, B.; Fonseca, A.C.; Romano, A.; Coelho, J.F.J.; Norgren, M. Brief Overview on Bio-Based Adhesives and Sealants. *Polymers* **2019**, *11*, 1685. [CrossRef]
31. Mapari, S.; Mestry, S.; Mhaske, S.T. Developments in pressure-sensitive adhesives: A review. *Polym. Bull.* **2021**, *78*, 4075–4108. [CrossRef]
32. Liu, X.; Hong, W.; Chen, X. Continuous Production of Water-Borne Polyurethanes: A Review. *Polymers* **2020**, *12*, 2875. [CrossRef]
33. Fuensanta, M.; Martin-Martínez, J.M. Thermoplastic polyurethane pressure sensitive adhesives made with mixtures of polypropylene glycols of different molecular weights. *Int. J. Adhes. Adhes.* **2019**, *88*, 81–90. [CrossRef]
34. Mehdizadeh, M.; Yang, J. Design strategies and applications of tissue bioadhesives. *Macromol. Biosci.* **2013**, *13*, 271–288. [CrossRef]
35. Zhou, X.; Tu, W.; Hu, J. Preparation and Characterization of Two-component Waterborne Polyurethane Comprised of Water-soluble Acrylic Resin and HDI Biuret. *Chin. J. Chem. Eng.* **2006**, *14*, 99–104. [CrossRef]
36. Malik, M.; Kaur, R. Mechanical and Thermal Properties of Castor Oil–Based Polyurethane Adhesive: Effect of $TiO_2$ Filler. *Adv. Polym. Technol.* **2016**, *37*, 24–30. [CrossRef]
37. Dai, J.; Liu, X.; Ma, S.; Wang, J.; Shen, X.; You, S.; Zhu, J. Soybean oil-based UV-curable coatings strengthened by crosslink agent derived from itaconic acid together with 2-hydroxyethyl methacrylate phosphate. *Prog. Org. Coat.* **2016**, *97*, 210–215. [CrossRef]
38. Raydan, N.D.V.; Leroyer, L.; Charrier, B.; Robles, E. Recent Advances on the Development of Protein-Based Adhesives for Wood Composite Materials—A Review. *Molecules* **2021**, *26*, 7617. [CrossRef]
39. Mubofu, E.B. Castor oil as a potential renewable resource for the production of functional materials. *Sustain. Chem. Processes* **2016**, *4*, 11. [CrossRef]
40. Silva, B.B.R.; Santana, R.M.C.; Forte, M.C. A solventless castor oil-based PU adhesive for wood and foam substrates. *Int. J. Adhes. Adhes.* **2010**, *30*, 559–565. [CrossRef]
41. Patel, M.R.; Shukla, J.M.; Patel, N.K.; Patel, K.H. Biomaterial based novel polyurethane adhesives for wood to wood and metal to metal bonding. *Mater. Res.* **2009**, *12*, 385–393. [CrossRef]
42. Radojčić, D.; Ionescu, M.; Petrović, Z.S. Novel potentially biodegradable polyurethanes from bio-based polyols. *Contemp. Mater.* **2013**, *4*, 9–21. [CrossRef]
43. Gadhave, R.V.; Mahanwar, P.A.; Gadekar, P.T. Bio-Renewable Sources for Synthesis of Eco-Friendly Polyurethane Adhesives—Review. *Open J. Polym. Chem.* **2017**, *7*, 57–75. [CrossRef]
44. Börcsök, Z.; Pásztory, Z. The role of lignin in wood working processes using elevated temperatures: An abbreviated literature survey. *Eur. J. Wood Wood Prod.* **2021**, *79*, 511–526. [CrossRef]
45. Jiang, T.; Duan, Q.; Zhu, J.; Liu, H.; Yu, L. Starch-based biodegradable materials: Challenges and opportunities. *Adv. Ind. Eng. Polym. Res.* **2019**, *3*, 8–18. [CrossRef]
46. Sudoł, E.; Kozikowska, E. Mechanical Properties of Polyurethane Adhesive Bonds in a Mineral Wool-Based External Thermal Insulation Composite System for Timber Frame Buildings. *Materials* **2021**, *14*, 2527. [CrossRef]
47. Calpena, E.O.; Francisca, A.A.; Sánchez, M.A.M. Adhesives in the Footwear Industry: A Critical Review. *Rev. Adhes. Adhes.* **2019**, *7*, 69–91. [CrossRef]
48. Bao, Z.; Gao, M.; Sun, Y.; Nian, R.; Xian, M. The recent progress of tissue adhesives in design strategies, adhesive mechanism and applications. *Mater. Sci. Eng. C* **2020**, *111*, 110796. [CrossRef]
49. Liu, L.; Lu, J.; Zhang, Y.; Liang, H.; Liang, D.; Jiang, J.; Lu, Q.; Quirino, R.L.; Zhang, C. Thermosetting polyurethanes prepared with the aid of a fully bio-based emulsifier with high bio-content, high solid content, and superior mechanical properties. *Green Chem.* **2018**, *21*, 526–537. [CrossRef]
50. Chattopadhyay, D.; Kothapalli, R.V. Structural engineering of polyurethane coatings for high performance applications. *Prog. Polym. Sci.* **2007**, *32*, 352–418. [CrossRef]
51. Sawpan, M.A. Polyurethanes from vegetable oils and applications: A review. *J. Polym. Res.* **2018**, *25*, 184. [CrossRef]

52. Hu, Y.; Zhu, G.; Zhang, J.; Huang, J.; Yu, X.; Shang, Q.; An, R.; Liu, C.; Hu, L.; Zhou, Y. Rubber Seed Oil-Based UV-Curable Polyurethane Acrylate Resins for Digital Light Processing (DLP) 3D Printing. *Molecules* **2021**, *26*, 5455. [CrossRef] [PubMed]
53. Patil, A.M.; Jirimali, H.D.; Gite, V.V.; Jagtap, R.N. Synthesis and performance of bio-based hyperbranched polyol in polyurethane coatings. *Prog. Org. Coat.* **2020**, *149*, 105895. [CrossRef]
54. Jeon, I.-Y.; Noh, H.-J.; Baek, J.-B. Hyperbranched Macromolecules: From Synthesis to Applications. *Molecules* **2018**, *23*, 657. [CrossRef]
55. Noreen, A.; Zia, K.M.; Zuber, M.; Tabasum, S.; Saif, M.J. Recent trends in environmentally friendly water-borne polyurethane coatings: A review. *Korean J. Chem. Eng.* **2015**, *33*, 388–400. [CrossRef]
56. Anıl, D.; Berksun, E.; Durmuş-Sayar, A.; Sevinis-Özbulut, E.B.; Ünal, S. Chapter 11—Recent advances in waterborne polyurethanes and their nanoparticle-containing dispersions. In *Handbook of Waterborne Coatings*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 249–302.
57. Zhang, Y.; Zhang, W.; Wang, X.; Dong, Q.; Zeng, X.; Quirino, R.L.; Lu, Q.; Wang, Q.; Zhang, C. Waterborne polyurethanes from castor oil-based polyols for next generation of environmentally-friendly hair-styling agents. *Prog. Org. Coat.* **2020**, *142*, 105588. [CrossRef]
58. Agnol, L.D.; Dias, F.T.G.; Ornaghi, H.L.; Sangermano, M.; Bianchi, O. UV-curable waterborne polyurethane coatings: A state-of-the-art and recent advances review. *Prog. Org. Coat.* **2021**, *154*, 106156. [CrossRef]
59. Ghosh, T.; Karak, N. Cashew nut shell liquid terminated self-healable polyurethane as an effective anticorrosive coating with biodegradable attribute. *Prog. Org. Coat.* **2019**, *139*, 105472. [CrossRef]
60. Kathalewar, M.; Sabnis, A.; D'Mello, D. Isocyanate free polyurethanes from new CNSL based bis-cyclic carbonate and its application in coatings. *Eur. Polym. J.* **2014**, *57*, 99–108. [CrossRef]
61. Lubi, M.C.; Thachil, E.T. Cashew nut shell liquid (CNSL)—A versatile monomer for polymer synthesis. *Des. Monomers Polym.* **2000**, *3*, 123–153. [CrossRef]
62. Kinaci, E.; Can, E.; La Scala, J.J.; Palmese, G.R. Epoxidation of Cardanol's Terminal Double Bond. *Polymers* **2020**, *12*, 2104. [CrossRef]
63. Asif, A.H.; Mahajan, M.S.; Sreeharsha, N.; Gite, V.V.; Al-Dhubiab, B.E.; Kaliyadan, F.; Nanjappa, S.H.; Meravanige, G.; Aleyadhy, D.M. Enhancement of Anticorrosive Performance of Cardanol Based Polyurethane Coatings by Incorporating Magnetic Hydroxyapatite Nanoparticles. *Materials* **2022**, *15*, 2308. [CrossRef] [PubMed]
64. Stachak, P.; Łukaszewska, I.; Hebda, E.; Pielichowski, K. Recent Advances in Fabrication of Non-Isocyanate Polyurethane-Based Composite Materials. *Materials* **2021**, *14*, 3497. [CrossRef] [PubMed]
65. Liu, G.; Wu, G.; Jin, C.; Kong, Z. Preparation and antimicrobial activity of terpene-based polyurethane coatings with carbamate group-containing quaternary ammonium salts. *Prog. Org. Coat.* **2015**, *80*, 150–155. [CrossRef]
66. Llevot, A.; Meier, M. Perspective: Green polyurethane synthesis for coating applications. *Polym. Int.* **2019**, *68*, 826–831. [CrossRef]
67. Olcay, H.; Kocak, E.D.; Yıldız, Z. Sustainability in Polyurethane Synthesis and Bio-based Polyurethanes BT. In *Sustainability in the Textile and Apparel Industries: Sourcing Synthetic and Novel Alternative Raw Materials*; Springer International Publishing: Cham, Switzerland, 2020; pp. 139–156.
68. Lee, S.T.; Park, C.B.; Ramesh, N.S. Polymeric foams. In *Polymers*, 1st ed.; CRC/Taylor & Francis: Boca Raton, FL, USA, 2006; Volume 6.
69. Szpiłyk, M.; Lubczak, R.; Lubczak, J. The biodegradable cellulose-derived polyol and polyurethane foam. *Polym. Test.* **2021**, *100*, 107250. [CrossRef]
70. Lin, C.S.K.; Kirpluks, M.; Priya, A.; Kaur, G. Conversion of food waste-derived lipid to bio-based polyurethane foam. *Case Stud. Chem. Environ. Eng.* **2021**, *4*, 100131. [CrossRef]
71. Zhang, J.; Hori, N.; Takemura, A. Influence of NCO/OH ratio on preparation of four agricultural wastes liquefied polyols based polyurethane foams. *Polym. Degrad. Stab.* **2020**, *179*, 109256. [CrossRef]
72. Jonjaroen, V.; Ummartyotin, S.; Chittapun, S. Algal cellulose as a reinforcement in rigid polyurethane foam. *Algal Res.* **2020**, *51*, 102057. [CrossRef]
73. Nabipour, H.; Wang, X.; Song, L.; Hu, Y. A fully bio-based coating made from alginate, chitosan and hydroxyapatite for protecting flexible polyurethane foam from fire. *Carbohydr. Polym.* **2020**, *246*, 116641. [CrossRef]
74. Thangavelu, S.A.G.; Mukherjee, M.; Layana, K.; Kumar, C.D.; Sulthana, Y.R.; Kumar, R.R.; Ananthan, A.; Muthulakshmi, V.; Mandal, A.B. Biodegradable polyurethanes foam and foam fullerenes nanocomposite strips by one-shot moulding: Physicochemical and mechanical properties. *Mater. Sci. Semicond. Process.* **2020**, *112*, 105018. [CrossRef]
75. Moon, J.; Kwak, S.B.; Lee, J.Y.; Kim, D.; Ha, J.U.; Oh, J.S. Recycling of bio-polyurethane foam using high power ultrasound. *Polymer* **2019**, *186*, 122072. [CrossRef]
76. Leszczyńska, M.; Ryszkowska, J.; Szczepkowski, L.; Kurańska, M.; Prociak, A.; Leszczyński, M.K.; Gloc, M.; Antos-Bielska, M.; Mizera, K. Cooperative effect of rapeseed oil-based polyol and eggshells on the structure and properties of rigid polyurethane foams. *Polym. Test.* **2020**, *90*, 106696. [CrossRef]
77. Mirski, R.; Dukarska, D.; Walkiewicz, J.; Derkowski, A. Waste Wood Particles from Primary Wood Processing as a Filler of Insulation PUR Foams. *Materials* **2021**, *14*, 4781. [CrossRef] [PubMed]
78. Agrawal, A.; Kaur, R.; Walia, R.S. PU foam derived from renewable sources: Perspective on properties enhancement: An overview. *Eur. Polym. J.* **2017**, *95*, 255–274. [CrossRef]

79. Kaur, R.; Kumar, M. Function of silicon oil in the castor oil based rigid polyurethane foams. *J. Polym. Eng.* **2013**, *33*, 875–880. [CrossRef]
80. Kumar, M.; Kaur, R. Effect of Different Formulations of MDI on Rigid Polyurethane Foams based on Castor Oil. *Int. J. Sci. Res. Rev.* **2013**, *2*, 29–42.
81. Agrawal, A.; Kaur, R.; Walia, R.S. Investigation on flammability of rigid polyurethane foam-mineral fillers composite. *Fire Mater.* **2019**, *43*, 917–927. [CrossRef]
82. Kumar, M.; Kaur, R. Glass fiber reinforced rigid polyurethane foam: Synthesis and characterization. *e-Polymers* **2017**, *17*, 517–521. [CrossRef]
83. Agrawal, A.; Kaur, R.; Walia, R. Development of vegetable oil-based conducting rigid PU foam. *e-Polymers* **2019**, *19*, 411–420. [CrossRef]
84. Agrawal, A.; Kaur, R.; Walia, R.S. Flame retardancy of ceramic-based rigid polyurethane foam composites. *J. Appl. Polym. Sci.* **2019**, *136*, 48250. [CrossRef]
85. Kaur, R.; Kumar, M. Addition of anti-flaming agents in castor oil based rigid polyurethane foams: Studies on mechanical and flammable behaviour. *Mater. Res. Express* **2020**, *7*, 015333. [CrossRef]
86. Shin, E.J.; Choi, S.M. Advances in Waterborne Polyurethane-Based Biomaterials for Biomedical Applications. *Adv. Exp. Med. Biology.* **2018**, *1077*, 251–283.
87. Dreifke, M.B.; Jayasuriya, A.A.; Jayasuriya, A.C. Current wound healing procedures and potential care. *Mater. Sci. Eng. C* **2015**, *48*, 651–662. [CrossRef] [PubMed]
88. Wendels, S.; Avérous, L. Biobased polyurethanes for biomedical applications. *Bioact. Mater.* **2021**, *6*, 1083–1106. [CrossRef]
89. Baheiraei, N.; Gharibi, R.; Yeganeh, H.; Miragoli, M.; Salvarani, N.; Di Pasquale, E.; Condorelli, G. Electroactive polyurethane/siloxane derived from castor oil as a versatile cardiac patch, part I: Synthesis, characterization, and myoblast proliferation and differentiation. *J. Biomed. Mater. Res. Part A* **2016**, *104*, 775–787. [CrossRef]
90. Tan, L.; Hu, J.; Huang, H.; Han, J.; Hu, H. Study of multi-functional electrospun composite nanofibrous mats for smart wound healing. *Int. J. Biol. Macromol.* **2015**, *79*, 469–476. [CrossRef]
91. Tan, L.; Hu, J.; Zhao, H.-F. Design of bilayered nanofibrous mats for wound dressing using an electrospinning technique. *Mater. Lett.* **2015**, *156*, 46–49. [CrossRef]
92. Baheiraei, N.; Yeganeh, H.; Ai, J.; Gharibi, R.; Ebrahimibarough, S.; Azami, M.; Vahdat, S.; Baharvand, H. Preparation of a porous conductive scaffold from aniline pentamer-modified polyurethane/PCL blend for cardiac tissue engineering. *J. Biomed. Mater. Res. Part A* **2015**, *103*, 3179–3187. [CrossRef]
93. Guo, B.; Finne-Wistrand, A.; Albertsson, A.C. Facile Synthesis of Degradable and Electrically Conductive Polysaccharide Hydrogels. *Biomacromolecules* **2011**, *12*, 2601–2609. [CrossRef]
94. Ganji, Y.; Kasra, M.; Salahshour, K.S.; Bagheri, H.M. Synthesis and characterization of gold nanotube/nanowire–polyurethane composite based on castor oil and polyethylene glycol. *Mater. Sci. Eng. C* **2014**, *42*, 341–349. [CrossRef]
95. Tsai, M.-C.; Hung, K.-C.; Hung, S.-C.; Hsu, S.-H. Evaluation of biodegradable elastic scaffolds made of anionic polyurethane for cartilage tissue engineering. *Colloids Surf. B Biointerfaces* **2015**, *125*, 34–44. [CrossRef] [PubMed]
96. Hung, K.-C.; Tseng, C.-S.; Dai, L.-G.; Hsu, S.-H. Water-based polyurethane 3D printed scaffolds with controlled release function for customized cartilage tissue engineering. *Biomaterials* **2016**, *83*, 156–168. [CrossRef] [PubMed]
97. Chen, J.; Dong, R.; Ge, J.; Guo, B.; Ma, P.X. Biocompatible, Biodegradable, and Electroactive Polyurethane-Urea Elastomers with Tunable Hydrophilicity for Skeletal Muscle Tissue Engineering. *ACS Appl. Mater. Interfaces* **2015**, *7*, 28273–28285. [CrossRef]
98. Chang, C.-J.; Hsu, S.-H. The effect of high outflow permeability in asymmetric poly(dl-lactic acid-co-glycolic acid) conduits for peripheral nerve regeneration. *Biomaterials* **2006**, *27*, 1035–1042. [CrossRef]
99. Hsu, S.-H.; Chang, W.-C.; Yen, C.-T. Novel flexible nerve conduits made of water-based biodegradable polyurethane for peripheral nerve regeneration. *J. Biomed. Mater. Res. A* **2017**, *105*, 1383–1392. [CrossRef]
100. Rocco, K.A.; Maxfield, M.W.; Best, C.A.; Dean, E.W.; Breuer, C.K. In vivo applications of electrospun tissue-engineered vascular grafts: A review. *Tissue Eng. Part B Rev.* **2014**, *20*, 628–640. [CrossRef]
101. Adolph, E.J.; Guo, R.; Pollins, A.C.; Zienkiewicz, K.J.; Cardwell, N.L.; Davidson, J.M.; Guelcher, S.A.; Nanney, L.B. Injected biodegradable polyurethane scaffolds support tissue infiltration and delay wound contraction in a porcine excisional model. *J. Biomed. Mater. Res. Part B Appl. Biomater.* **2016**, *104*, 1679–1690. [CrossRef]
102. Liu, W.; Wang, T.; Yang, C.; Darvell, B.W.; Wu, J.; Lin, K.; Chang, J.; Pan, H.; Lu, W. Alkaline biodegradable implants for osteoporotic bone defects–importance of microenvironment pH. *Osteoporos. Int.* **2016**, *27*, 93–104. [CrossRef]
103. González Paz, R.J.; Lligadas, G.; Ronda, J.C.; Galià, M.; Ferreira, A.M.; Boccafoschi, F.; Ciardelli, G.; Cadiz, V. Enhancement of fatty acid-based polyurethanes cytocompatibility by non-covalent anchoring of chondroitin sulfate. *Macromol. Biosci.* **2012**, *12*, 1697–1705. [CrossRef]
104. Blakney, A.K.; Simonovsky, F.I.; Suydam, I.T.; Ratner, B.D.; Woodrow, K.A. Rapidly Biodegrading PLGA-Polyurethane Fibers for Sustained Release of Physicochemically Diverse Drugs. *ACS Biomater. Sci. Eng.* **2016**, *2*, 1595–1607. [CrossRef]
105. Żółtowska, K.; Piotrowska, U.; Oledzka, E.; Kuras, M.; Zgadzaj, A.; Sobczak, M. Biodegradable Poly(ester-urethane) Carriers Exhibiting Controlled Release of Epirubicin. *Pharm. Res.* **2017**, *34*, 780–792. [CrossRef] [PubMed]
106. Mahanta, A.K.; Mittal, V.; Singh, N.; Dash, D.; Malik, S.; Kumar, M.; Maiti, P. Polyurethane-Grafted Chitosan as New Biomaterials for Controlled Drug Delivery. *Macromolecules* **2015**, *48*, 2654–2666. [CrossRef]

107. Yu, C.; Tan, X.; Xu, Z.; Zhu, G.; Teng, W.; Zhao, Q.; Liang, Z.; Wu, Z.; Xiong, D. Smart drug carrier based on polyurethane material for enhanced and controlled DOX release triggered by redox stimulus. *React. Funct. Polym.* **2020**, *148*, 104507. [CrossRef]
108. Aluri, R.; Jayakannan, M. Development of l-Tyrosine-Based Enzyme-Responsive Amphiphilic Poly(ester-urethane) Nanocarriers for Multiple Drug Delivery to Cancer Cells. *Biomacromolecules* **2017**, *18*, 189–200. [CrossRef]
109. Kathalewar, M.S.; Joshi, P.B.; Sabnis, A.S.; Malshe, V.C. Non-isocyanate polyurethanes: From chemistry to applications. *RSC Adv.* **2013**, *3*, 4110–4129. [CrossRef]
110. Doley, S.; Dolui, S.K. Solvent and catalyst-free synthesis of sunflower oil based polyurethane through non-isocyanate route and its coatings properties. *Eur. Polym. J.* **2018**, *102*, 161–168. [CrossRef]
111. Malik, M.; Kaur, R. Synthesis of NIPU by the carbonation of canola oil using highly efficient 5,10,15-tris(pentafluorophenyl)corrolato-manganese(III) complex as novel catalyst. *Polym. Adv. Technol.* **2018**, *29*, 1078–1085. [CrossRef]
112. Carré, C.; Ecochard, Y.; Caillol, S.; Avérous, L. From the Synthesis of Biobased Cyclic Carbonate to Polyhydroxyurethanes: A Promising Route towards Renewable Non-Isocyanate Polyurethanes. *ChemSusChem* **2019**, *12*, 3410–3430. [CrossRef]
113. Rokicki, G.; Parzuchowski, P.G.; Mazurek, M. Non-isocyanate polyurethanes: Synthesis, properties, and applications. *Polym. Adv. Technol.* **2015**, *26*, 707–761. [CrossRef]
114. Javni, I.; Hong, D.P.; Petrović, Z.S. Polyurethanes from soybean oil, aromatic, and cycloaliphatic diamines by nonisocyanate route. *J. Appl. Polym. Sci.* **2013**, *128*, 566–571. [CrossRef]
115. Simón, D.; Borreguero, A.M.; de Lucas, A.; Rodríguez, J.F. Recycling of polyurethanes from laboratory to industry, a journey towards the sustainability. *Waste Manag.* **2018**, *76*, 147–171. [CrossRef] [PubMed]
116. Motokucho, S.; Nakayama, Y.; Morikawa, H.; Nakatani, H. Environment-friendly chemical recycling of aliphatic polyurethanes by hydrolysis in a $CO_2$ -water system. *J. Appl. Polym. Sci.* **2017**, *135*, 14–17. [CrossRef]
117. Kemona, A.; Piotrowska, M. Polyurethane Recycling and Disposal: Methods and Prospects. *Polymers* **2020**, *12*, 1752. [CrossRef]
118. Simón, D.; Borreguero, A.M.; de Lucas, A.; Rodríguez, J.F. Glycolysis of viscoelastic flexible polyurethane foam wastes. *Polym. Degrad. Stab.* **2015**, *116*, 23–35. [CrossRef]
119. Guo, X.; Zhang, W.; Wang, L.; Hao, J. Comparative study of nitrogen migration among the products from catalytic pyrolysis and gasification of waste rigid polyurethane foam. *J. Anal. Appl. Pyrolysis* **2016**, *120*, 144–153. [CrossRef]
120. AnuarSharuddin, S.D.; Abnisa, F.; Wan Daud, W.M.A.; Aroua, M.K. A review on pyrolysis of plastic wastes. *Energy Convers. Manag.* **2016**, *115*, 308–326. [CrossRef]
121. Jomaa, G.; Goblet, P.; Coquelet, C.; Morlot, V. Kinetic modeling of polyurethane pyrolysis using non-isothermal thermogravimetric analysis. *Thermochim. Acta* **2015**, *612*, 10–18. [CrossRef]
122. Yang, W.; Dong, Q.; Liu, S.; Xie, H.; Liu, L.; Li, J. Recycling and Disposal Methods for Polyurethane Foam Wastes. *Procedia Environ. Sci.* **2012**, *16*, 167–175. [CrossRef]
123. Osman, M.; Satti, S.M.; Luqman, A.; Hasan, F.; Shah, Z.; Shah, A.A. Degradation of Polyester Polyurethane by Aspergillus sp. Strain S45 Isolated from Soil. *J. Polym. Environ.* **2018**, *26*, 301–310. [CrossRef]
124. Khan, S.; Nadir, S.; Shah, Z.U.; Shah, A.A.; Karunarathna, S.C.; Xu, J.; Khan, A.; Munir, S.; Hasan, F. Biodegradation of polyester polyurethane by Aspergillus tubingensis. *Environ. Pollut.* **2017**, *225*, 469–480. [CrossRef]
125. Magnin, A.; Pollet, E.; Phalip, V.; Avérous, L. Evaluation of biological degradation of polyurethanes. *Biotechnol. Adv.* **2020**, *39*, 107457. [CrossRef]

Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Computers and Electrical Engineering 102 (2022) 108224

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

# Automatic illness prediction system through speech☆

Husam Ali Abdulmohsin [a,*], Belal Al-Khateeb [b], Samer Sami Hasan [a], Rinky Dwivedi [c]

[a] Computer Science Department, College of Science, University of Baghdad, Baghdad, Iraq
[b] College of Computer Science and information Technology, University of Anbar, Anbar, Iraq
[c] Maharaja Surajmal Institute of Technology, New Delhi, India

ARTICLE INFO

*Keywords:*
Medical speech transcription and intent dataset
Automatic disease prediction
Mel frequency Cepstral coefficient
Spectral centroid variability
Forward-backward filter
Neural network

ABSTRACT

Due to the COVID-19 epidemic and the curfew caused by it, many people have sought to find an ADPS on the internet in the last few years. This hints to a new age of medical treatment, all the more so if the number of internet users continues to expand. As a result, automatic illness prediction online applications have attracted the interest of a large number of researchers worldwide. This work aims to develop and implement an automated illness prediction system based on speech. The system will be designed to forecast the sort of ailment a patient is suffering from based on his voice, but this was not feasible during the trial, therefore the diseases were divided into three categories (painful, light pain and psychological pain), and then the diagnose process were implemented accordingly. The medical dataset named "speech, transcription, and intent" served as the baseline for this study. The smoothness, MFCC, and SCV properties were used in this work, which demonstrated their high representation to human being medical situations. The noise reduction forward-backward filter was used to eliminate noise from wave files captured online in order to account for the high level of noise seen in the deployed dataset. For this study, a hybrid feature selection method was created and built that combined the output of a genetic algorithm (GA) with the inputs of a NN algorithm. Classification was performed using SVM, neural network, and GMM. The greatest results obtained were 94.55% illness classification accuracy in terms of SVM. The results showed that diagnosing illness through speech is a difficult process, especially when diagnosing each type of illness separately, but when grouping the different illness types into groups, depending on the amount of pain and the psychological situation of the patient, the results were much higher.

## 1. Introduction

Researchers and companies have been attracted to the automatic disease prediction system (ADPS) in recent years because it enables rapid access to medical treatment, avoids traffic and long-distance travel, and is especially beneficial to senior citizens, regardless of the new era the world is entering with the spread of COVID-19 and the periodic curfews implemented by many countries

*Abbreviations:* ADPS, Automated Disease Prediction System; CPU, Central Processing Unit; GA, Genetic Algorithm; GB, Giga Byte; GMM, Gaussian Mixture Model; MFCC, Mel Frequency Cepstral Co-efficient; NN, Neural Network; RAM, Random Access Memory; RSM, Response Service Methodology; SCV, Spectral Centroid Variability; SVM, Support Vector Machine.

☆ This work is implemented in University of Baghdad, College of science, Computer science department labs, Baghdad, Iraq, not funded.

* Corresponding author.
*E-mail address:* husam.a@sc.uobaghdad.edu.iq (H.A. Abdulmohsin).

https://doi.org/10.1016/j.compeleceng.2022.108224
Received 30 January 2022; Received in revised form 29 June 2022; Accepted 8 July 2022
Available online 21 July 2022
0045-7906/© 2022 Elsevier Ltd. All rights reserved.

to prevent the virus from spreading.

It is critical to understand the human speech production process in order to deal with ADPSs professionally. Phonetics is the study of the sounds produced by human speech. Speech is produced by pushing air from the lungs to the larynx (respiration), where the vocal cords may be open to allow for air flow or may vibrate, producing sound (phonation). The articulators in the mouth and nose, who are responsible for articulation, will affect the lungs' airflow.

Acoustic and articulatory phonetics are two perspectives of spoken sounds. Acoustic phonetics is a subfield of linguistics, but it is also a branch of physics, focusing on the acoustic and physical aspects of sound waves. Articulatory phonetics is the method through which human bodies are employed to produce spoken sounds [1]. Three processes are required to generate speech. To begin, an energy source is required. Anything that makes sound requires an energy source, and in the case of human speaking sounds, the energy source is the air coming from the lungs. The second mechanism is the sound source, which is the vocal cords of the larynx (or vocal folds). The articulators are the third mechanism, and they are used to filter or shape the sound. The oral cavity (mouth space), the nasal cavity (space within and behind the nose), the jaws, teeth, lips, and, most importantly, the tongues are all articulators [2].

Acoustic voiced sounds are produced when air from the lungs passes through the vocal cords, causing the vocal cords to vibrate and produce periodic signals. Due to the unique physical qualities of each human being's vocal cords, distinct frequencies are produced during voice production, which reflects the unique attributes of each human being's acoustics [3]. The epiglottis is the second component of the human speech production system. Each individual's epiglottis has a unique bulk and flexibility [4,5]. This also results in the addition of new elements to speech, regardless of the tong position or the mouth cavity, which both result in various pronunciations for the same words when spoken by different persons. The varied properties of the epiglottis also vary across genders, with female formants being more frequent than male formants and the spectrum of voiced sounds often decreasing in loudness with increasing frequency. All of these acoustic phenomena are a result of speech production. As a result, it is feasible to identify gender-specific elements in acoustic speech signals [6].

According to the source of creation, speech characteristics can be classified as acoustic or articulation features. If a human being is weary or in pain, his articulators will reflect his medical condition, and his sound will be created differently, since speaking requires a great deal of energy, and the first thing that will be impacted in his voice is his fatigue or discomfort [7–9]. The amount of verbal and gestural communication generated is influenced by the severity of acute pain. When pain intensity rises, so does spontaneous verbal communication [10]. Furthermore, research demonstrates that when pain is more acute, not only the use of nonverbal pain behaviors, but also the usage of co-speech gestures, rises in the realm of visible physical communication [11]. These gestures are rich in semantic content and offer crucial information about a sufferer's pain experience. In contrast to pain behaviors, demonstrating that when pain is more acute, people use multimodal communication resources to give greater information about their suffering. [12–14]. This is regarded evidence that any human health condition may be translated through his voice, but the most critical aspect is to select the most pertinent characteristics that will translate the sickness, infection, discomfort, or any other health issue.

Other factors that increase the challenge of health detection in general are the following:

- Certain psychological situations among different individuals, under certain circumstances, can affect the patient's judgment of his health situation. Sickness can be faked.
- To discern health automatically from vocal expressions, we must delete superfluous and dangerous information such as organ noise and human reactions that might result in abnormal voices.
- This work has made several contributions, that overcome the limitations of the state-of-art researches, which are listed below:
- Gender and speaker independency, a new features selection method enabled the system to function independently with respect to gender and speaker, and this is a huge contribution to health detection.
- Accuracy compared to literature survey; high classification accuracy results are achieved through the recognition strength of the features extracted from the feature extraction method deployed. Notably, when applied to the medical speech, transcription, and intent dataset, the method achieves 94.55% classification accuracy, and this is the highest accuracy result gained compared to the state-of-art works.
- Generality, the same system setup was used to detect health for all languages, genders and ages represented in the dataset, and no tweaking was done for each category of people. This demonstrates that the characteristics derived in this work are highly capable of detecting illness and that they function across languages, recording frequencies, genders, and ages and are speaker independent.
- Age and language independent, all the results given in this work were gained by recognizing all limits of ages, all languages, and all genders in the datasets utilized without removing any category from the dataset or merging multiple categories into one group to avoid misclassification. This was a good challenge because all three datasets contained conflicting emotions, including happy and anger, neutral and calm, and disgust and surprise, which are difficult to distinguish.

The related works will be described in Section 2 of this article. The statistics and features of the benchmark datasets used in this research, as well as the frame-work for the proposed technique, will be discussed in depth in Section 3. Section 4 will present and debate the work's highest categorization findings. Section 5 will summarize the findings from this effort.

## 2. Related Work

Numerous studies have established the presence of distinct acoustic and physiological characteristics in human voices that can be utilized to identify patient symptoms, but they have not yet achieved the needed classification accuracy. Several scholars have investigated ADPSs using a variety of various methodologies and processes.

One of the processes upon which ADPS's rely is on clinical reports to identify the patient's highest expected condition [15]. Another mechanism by which healthcare ADPSs operate is through the question-and-answer approach, which relies on user text input (guided by users). The algorithm will then provide a list of the most likely illnesses [16–19]. Numerous researchers have applied machine learning techniques to the diagnosis of cardiac disease [20]. Another technique is to utilize natural language processing and deep learning to extract text that may be used to identify certain symptoms for the purpose of illness prediction [21–23] and several additional studies. According to a 2014 study conducted by Maree Johson, nearly 20,000 state-of-art papers in the field of voice to text ADPS were published in 2014. These works utilized computational linguistics, natural language processing, human language technology, or text mining [24]. According to the literature study undertaken for this work, the suggested work belongs to the field of ADPS since it does not include linguistics, natural language processing, or speech conversion for the purpose of text mining. This paper developed an ADPS that estimates the degree of pain a patient is experiencing based just on his speech, without requiring any additional language processing. Numerous difficulties were encountered, one of which was the psychological state of the human person and its influence on voice. When a person is psychologically unstable, his vowels are stated differently than when he is stable [25,26]. The other difficulty was dealing with the similarity of children's voices between the ages of three and seven years; hence, children were excluded from the studies used to validate the approach described in this study, and since children rely on adults to speak out for them. Noise and distortion were a challenge in many of the recordings in the used dataset, as they will be in real life. The suggested effort has several limits that must be established, such as its interaction with youngsters and its lack of familiarity with other languages. We were unable to identify the voice signal for each condition in this work, but we were able to differentiate between diseases that have psychological effects and those that do not, as well as diagnose diseases based on the intensity of pain, strong and mild pain.

## 3. Dataset Selection

The dataset used in this study is version 1 of the medical voice, transcription, and intent dataset. The collection is comprised of spoken descriptions in (wav) audio format that detail the patient's medical problems. Each vocal description is accompanied with a transcription in the (csv) text format, and then each is tagged with a specific type of illness. This is a single-language dataset comprised of North American English slang. It enumerates 25 symptoms, all of which are back pain, acne, blurry vision, body feeling weakened,

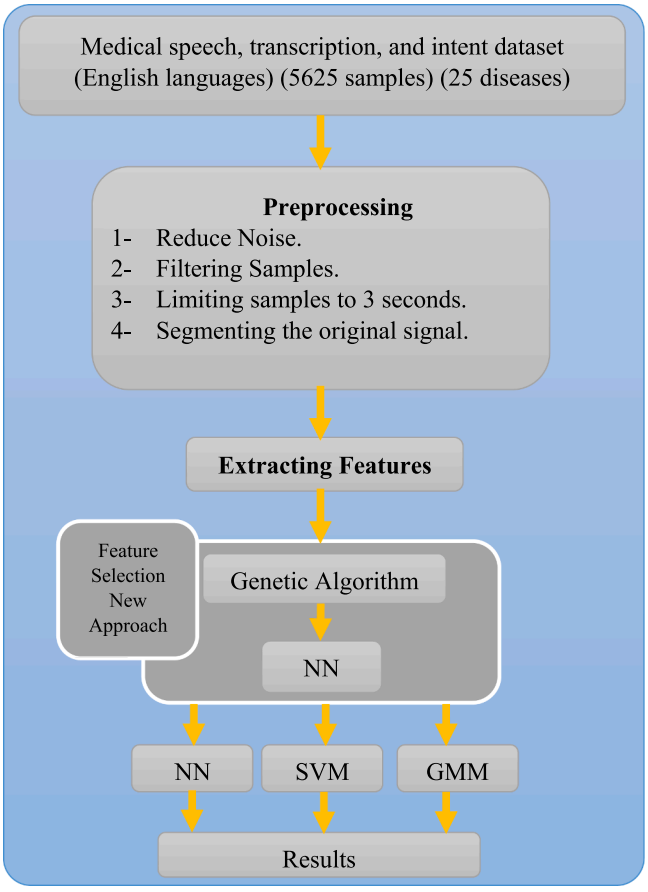



**Fig. 1.** The block diagram of the method proposed in this work.

cough, ear ache, emotional pain, feeling cold, feeling dizzy, foot ache, hair falling out, difficulty breathing, head ache, heart hurts, infected wound, injury from spots, internal pain, joint pain, knee pain, muscle pain, neck pain, open wound, shoulder pain, skin issue, and stomach ache. Each symptom has a unique number of recordings, which required us to disregard some of the recordings for certain symptoms in order to get an equal number of recordings for each symptom, which is 225 records for each symptom. This work has a total of 5625 recordings [27]. We found a lot of distortion and noise in some of the wave recordings throughout our study, which we attribute to the low quality of the microphones used to record those samples and the loud setting in which they were recorded. This necessitated the use of a forward-backward digital filter, dubbed the zero-phase filter, in all recordings. The justification for adopting zero-phase filtering is because it retains filtered features exactly where they appeared in the unfiltered signal, which has no effect on classification performance. These filters conduct zero-phase digital filtering in both directions, forward and backward. After filtering the data in the forward direction, the filter sequence will be reversed, re-filtering the signal [28].

As we have addressed most of the main concepts of health detection systems, we will now explain our framework proposed in this research in detail.

## 4. Framework

The suggested technique is depicted in Fig. 1 using a block diagram. The dataset and the suggested method's major phases will be illustrated in depth in the following four sections.

### 4.1. Preprocessing Phase

Numerous processes were applied during this stage. To begin, there is the noise reduction step. The noise reduction forward-backward filter was used to eliminate noise from wave files recorded online in order to account for the significant level of noise observed in several samples from the medical speech, transcription, and intent dataset utilized in this study. Second, the process of sampling and filtering. Each of the 25 illnesses included in the collection had at least 225 samples, and several had significantly more. Thus, we picked 225 samples for each illness and eliminated samples with excessive noise based on our listening to all samples, which was a lengthy procedure. Trimming was the third procedure. The majority of the recordings were made under various time constraints. As a result, all samples were shortened to three seconds. Segmentation was the fourth procedure. All samples were segmented at a ratio of (0.05%) to the original signal, and the overlap ratio used in this study was (0,025%), which results in a (50%) overlap, and the total number of segments created will be (2n-1), where n is the number of original segments.

### 4.2. Feature Extraction

A feature is a quantifiable quality derived from the substance under observation [29]. The most critical component of feature extraction is identifying the aspects that are most relevant to the issue statement. In this work, the smoothness, mel frequency cepstral coefficient (MFCC) with 12 degrees, and spectral centroid variability (SCV) features were extracted. These features are strongly related to human beings experiencing pain, and were chosen based on experiments conducted on 15 different types of speech features. The mean and standard deviation in multiple degrees were calculated for each feature, therefore, each feature generated 8 features, representing the (mean, and seven different degrees for the standard deviation. As a results, 112 features were generated from the feature extraction phase, all were utilized through experiment (8 smoothness features, 96 MFCC features, and 8 SCV features).

If smoothness is defined as the transaction of speech via air, then smooth speech results in a delayed transaction, whereas rough speech results in a faster transaction [30]. Smoothness was determined in two domains: the temporal domain and the spectral domain. Smoothness is determined using Eq. (1) and (2) [31].

$$GV_t = \frac{1}{P}\sqrt{\sum_{j=1}^{P}(var_t(j))^2} \tag{1}$$

$$GV_s = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(var_s(i))^2} \tag{2}$$

Where $var_t$ and $var_s$ denote the temporal and spectral domain variances of the spectral feature, P denotes the feature's dimension, and N denotes the feature's time domain length

### 4.3. Feature Selection

The evolutionary feature selection method (genetic method) was used in this study was to filter feature groups before giving them to the neural Network (NN), in order to choose the best feature groups according to a given criterion, and then to pass the best selected feature group to the NN.

The idea of the genetic algorithm was to weight each feature group (each two features) with a fitness function, that shows the classification power of that certain group, and the weight will be used to combine the groups to form new ones. Weighted groups will

be checked, and merged, until the criterion is fulfilled, at which point the best group will be handed to the NN. The neural network is used for classification or as an integrated feature selection technique, and it is also utilized in data mining and machine learning [32, 33]. Three considerations motivated the employment of a filter ranking algorithm in this work. To begin, filter out irrelevant variables. Second, to take advantage of the variable selection criteria based on the order of the variable ranking approaches. Finally, and maybe most importantly, their simplicity and success as reported by online apps. Each variable is scored using a ranking criterion, and then a threshold is determined through experimentation and utilized to exclude variables that fall below that threshold [29].

Filter feature selection techniques are those that are applied before to classification; hence, ranking methods are considered filter methods. The fundamental premise behind feature selection methods is to pick unique features that convey important information about various classes in the dataset by leveraging a feature's basic attribute. This quality is referred to as feature relevance, and it quantifies a feature's ability to categorize distinct classes [29,34,35].

### 4.4. Neural Network

In this study, the support vector machine (SVM), backpropagation NN, and Gaussian mixture model (GMM) classifiers were used to predict the patient's discomfort. The structure of the neural network was two hidden neural network layers, the first layer with 80 hidden nodes, and the second layer with 40 hidden nodes. The input were 25 diseases and the output were 3 groups of diseases. The fitness function of the NN, was a NN that evaluated the classification accuracy of each sub group. Three kernel functions used with the SVM, were the polynomial kernel of degree, calculated through Eq. (3):

$$\mathrm{p},\ \left(x_i,\ x_j\right)^{\mathrm{p}} \tag{3}$$

The radial basis function kernel, calculated through Eq. (4):

$$\mathrm{e}^{-\mathrm{a}\left|\left|x_i\ -\ x_j\right|\right|^2},\ \text{with a} > 0 \tag{4}$$

and the sigmoid function, calculated through Eq. (5):

$$\tanh\ \left(a.\ x_i\ .\ x_j + c\right) \tag{5}$$

The radial kernel function achieved the highest accuracies.

## 5. Hardware and Software Platform

All tests were conducted in MATLAB R2019a on a computer equipped with an Intel Core i7-8750H Central Processing Unit (CPU) running at 2.2 GHz and 32 GB of Random Access Memory (RAM) (RAM). The MATLAB instructions are executed on the Windows 10 platform.

## 6. System Setting and Tuning

All of the findings described will be derived using the same system parameters and will not be changed for each sickness category contained in the dataset. The objective was to design a system capable of addressing a wide variety of languages and accents using the same parameters in order to achieve generalizability. The variables that were fixed during the system's setup were tweaked until the highest classification accuracy was achieved, but then they were fixed. Those most important variables are the following:

- **The length of the window dividing the signal** specifies the length of the segment that will be extracted from the original signal in order to extract the desired characteristics. The window length was calculated through Eq. (6):

$$\text{Window}_{\ Length} = 0.05\ \text{X Frequency} \tag{6}$$

- **The length of the step** sets the overlap between the current and previous windows in order to capture the majority of the possible divisions of the original signal. The step length was calculated through Eq. (7):

$$\text{Step}_{\ Size} = 0.025\ \text{X Frequency} \tag{7}$$

- **The standard deviation degree** sets the number of degrees needs for the standard deviation to achieve the highest accuracy.

There are other variables that were fixed, but we have discussed the most important ones.

## 7. Experimental Results

Many experiments were conducted through our work, two of the most important experiments will be discussed that achieved the highest results.

### 7.1. Experiment number 1 (Exp1)

The purpose of this experiment was to anticipate the disease represented in each wave recording and to classify each recording according to one of the 25 diseases represented in the dataset. The confusion matrix of the best accuracy classification results obtained using the three classifiers SVM, NN, and GMM is shown in Figs. 2, 3 and 4. As shown in the three confusion matrices, the classification accuracies produced by the SVM, NN, and GMM were 50.8 %, 48.7 %, and 31.5 %, respectively.

The three classifiers used produced poor results that are unacceptable in any ADPS. Through analyzing the results some conclusions were gained, such as:

1 The classification accuracy of each individual disease lye near the mean of the overall classification accuracy, which shows the difficulty of gaining high classification accuracy from this approach.
2 It has been noticed that some diseases have been misclassified with a certain group of diseases and not with all diseases which led to the idea of clustering the diseases into three different groups. For example, in Fig. 2, column (1), ear ache, foot ache, head ache, Infected wound, injury from spots, internal pain, joint pain, knee pain, muscle pain, neck pain, open wound, shoulder pain, and stomach ache wave samples have been misclassified with back pain by (9, 14, 11, 8, 7, 9, 11, 8, 9, 12, 23, 10, and 7) wave samples. Whereas, the all-other diseases were not misclassified classified with back pain in that high number of wave samples, the misclassifications with other diseases were (1, 1, 0, 0, 1, 1, 0, 1, 4, 2, and 1), which shows that a certain number of diseases share the same effect on speech, different than other diseases.
3 The highest classification accuracy gained from the SVM classifier was 70.2% with the heart hurts disease, which make sense, because any pain in the heart affects the breath of the human being, and as results, will affect his speech.
4 There are some characteristics that are shared among similar recordings of the same condition, as seen by the findings of Exp1. Regardless of the modest findings obtained, the 50.8 % classification accuracy indicates that there are some common characteristics among the recordings, but these characteristics need to be increased.
5 We discovered that illnesses are classified into three groups based on the misclassification distribution of samples in the confusion matrices. Each illness within the same group is categorized incorrectly as another disease within the same category. After analyzing the disorders, we discovered that they may be classified into three categories based on the level of discomfort they produce.
6 There are certain diseases that affect the human voice, primarily due to the amount of pain they cause. These diseases include (painful diseases, back pain, internal pain, joint pain, knee pain, muscle pain, neck pain, open wound, shoulder pain, stomach ache, injury from spots, infected wound, head ache, and ear ache. Foot discomfort) (which will be referred to as Group 1 in Exp2), and these disorders have an effect on the phonetics of articulation. There are several disorders that impact the acoustic phonetics of

| | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| back pain | 75 | 1 | 0 | 2 | 0 | 11 | 1 | 1 | 0 | 14 | 0 | 2 | 7 | 1 | 8 | 8 | 12 | 13 | 14 | 12 | 12 | 13 | 9 | 1 | 13 | 32.6 |
| acne | 1 | 119 | 12 | 23 | 0 | 0 | 2 | 3 | 1 | 1 | 16 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 11 | 1 | 60.7 |
| blurry vision | 1 | 21 | 127 | 21 | 1 | 1 | 1 | 1 | 1 | 1 | 17 | 2 | 0 | 1 | 0 | 3 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 14 | 1 | 58.5 |
| body feels weak | 0 | 24 | 19 | 126 | 1 | 0 | 0 | 1 | 0 | 0 | 13 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 14 | 1 | 61.2 |
| cough | 0 | 1 | 1 | 0 | 129 | 4 | 16 | 12 | 9 | 1 | 0 | 12 | 0 | 8 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 64.5 |
| * ear ache | 9 | 0 | 1 | 1 | 1 | 99 | 0 | 0 | 1 | 8 | 0 | 0 | 8 | 1 | 7 | 12 | 11 | 9 | 11 | 14 | 9 | 10 | 10 | 1 | 12 | 42.1 |
| emotional pain | 1 | 0 | 0 | 0 | 22 | 0 | 131 | 11 | 13 | 1 | 1 | 11 | 1 | 10 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 2 | 63.0 |
| feeling cold | 1 | 1 | 0 | 1 | 15 | 0 | 17 | 144 | 11 | 0 | 0 | 13 | 0 | 9 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 3 | 65.8 |
| * feeling dizzy | 0 | 1 | 0 | 2 | 16 | 1 | 23 | 13 | 135 | 0 | 1 | 11 | 1 | 9 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 1 | 1 | 61.6 |
| foot ache | 14 | 1 | 1 | 0 | 0 | 12 | 0 | 1 | 0 | 108 | 0 | 0 | 10 | 0 | 12 | 9 | 11 | 13 | 13 | 10 | 11 | 14 | 9 | 1 | 17 | 40.4 |
| hair falling out | 1 | 18 | 21 | 19 | 1 | 0 | 0 | 1 | 1 | 1 | 57 | 0 | 3 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 27 | 1 | 61.3 |
| heard to breath | 4 | 0 | 2 | 2 | 15 | 0 | 16 | 15 | 14 | 0 | 1 | 150 | 1 | 17 | 1 | 1 | 0 | 1 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 61.7 |
| *head ache | 11 | 2 | 3 | 0 | 1 | 14 | 0 | 1 | 2 | 12 | 0 | 1 | 105 | 1 | 8 | 6 | 9 | 8 | 7 | 11 | 5 | 34 | 13 | 2 | 14 | 38.9 |
| heart hurts | 2 | 0 | 0 | 0 | 17 | 0 | 13 | 12 | 15 | 0 | 2 | 13 | 1 | 58 | 1 | 0 | 1 | 0 | 2 | 0 | 0 | 1 | 1 | 1 | 3 | 65.0 |
| *Infected wound | 8 | 1 | 1 | 0 | 1 | 8 | 0 | 0 | 1 | 7 | 0 | 0 | 13 | 1 | 116 | 10 | 7 | 7 | 7 | 8 | 6 | 4 | 7 | 2 | 7 | 52.3 |
| injury from spots | 7 | 1 | 0 | 1 | 0 | 9 | 0 | 0 | 2 | 7 | 0 | 1 | 11 | 0 | 7 | 105 | 18 | 7 | 12 | 7 | 5 | 5 | 8 | 1 | 6 | 47.7 |
| internal pain | 9 | 2 | 0 | 0 | 1 | 8 | 0 | 1 | 2 | 7 | 1 | 0 | 7 | 1 | 8 | 8 | 76 | 10 | 8 | 8 | 8 | 6 | 9 | 1 | 8 | 40.2 |
| joint pain | 11 | 1 | 0 | 1 | 0 | 9 | 1 | 1 | 3 | 7 | 0 | 1 | 8 | 1 | 7 | 8 | 12 | 78 | 18 | 7 | 9 | 4 | 7 | 1 | 5 | 39.0 |
| knee pain | 8 | 1 | 1 | 0 | 0 | 8 | 0 | 0 | 3 | 9 | 1 | 1 | 8 | 1 | 8 | 9 | 11 | 13 | 72 | 8 | 9 | 6 | 32 | 0 | 6 | 33.5 |
| muscle pain | 9 | 2 | 1 | 0 | 1 | 8 | 1 | 0 | 3 | 7 | 0 | 1 | 7 | 1 | 7 | 8 | 6 | 7 | 11 | 97 | 9 | 7 | 22 | 0 | 7 | 43.7 |
| neck pain | 12 | 0 | 1 | 0 | 1 | 7 | 0 | 1 | 2 | 7 | 0 | 1 | 9 | 1 | 8 | 8 | 11 | 11 | 11 | 12 | 111 | 7 | 9 | 0 | 7 | 46.8 |
| open wound | 23 | 1 | 1 | 0 | 0 | 7 | 1 | 1 | 2 | 8 | 0 | 1 | 9 | 2 | 9 | 7 | 14 | 16 | 12 | 10 | 9 | 96 | 9 | 0 | 7 | 39.2 |
| shoulder pain | 10 | 1 | 0 | 1 | 1 | 8 | 1 | 1 | 3 | 9 | 0 | 1 | 8 | 0 | 6 | 12 | 14 | 16 | 13 | 12 | 8 | 9 | 62 | 1 | 6 | 30.5 |
| skin issue | 1 | 25 | 32 | 24 | 1 | 0 | 0 | 1 | 1 | 1 | 12 | 1 | 0 | 2 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 144 | 0 | 57.8 |
| *stomach ache | 7 | 1 | 1 | 1 | 0 | 11 | 1 | 3 | 0 | 9 | 3 | 0 | 7 | 0 | 10 | 8 | 8 | 9 | 8 | 6 | 7 | 4 | 13 | 1 | 96 | 44.9 |
| | 33.3 | 52.9 | 56.4 | 56.0 | 57.3 | 44.0 | 58.2 | 64.0 | 60.0 | 48.0 | 69.8 | 66.7 | 46.7 | 70.2 | 51.6 | 46.7 | 33.8 | 34.7 | 32.0 | 43.1 | 49.3 | 42.7 | 27.6 | 64.0 | 42.7 | 50.1 |

**Fig. 2.** The Confusion Matrix of the best classification accuracy gained from predicting 25 diseases in Exp 1 using SVM.

| | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| back pain | 71 | 0 | 2 | 0 | 0 | 9 | 2 | 0 | 2 | 9 | 1 | 2 | 7 | 0 | 6 | 10 | 19 | 8 | 11 | 8 | 10 | 11 | 10 | 2 | 8 | 34.1 |
| acne | 1 | 86 | 31 | 11 | 0 | 0 | 2 | 3 | 2 | 2 | 23 | 0 | 2 | 0 | 2 | 2 | 2 | 1 | 0 | 1 | 2 | 2 | 1 | 11 | 1 | 45.7 |
| blurry vision | 0 | 23 | 56 | 34 | 0 | 1 | 3 | 3 | 3 | 3 | 25 | 0 | 3 | 0 | 2 | 3 | 1 | 0 | 0 | 3 | 2 | 0 | 1 | 13 | 0 | 31.3 |
| body feels weak | 1 | 25 | 33 | 86 | 0 | 1 | 0 | 3 | 1 | 1 | 12 | 0 | 2 | 0 | 2 | 4 | 2 | 0 | 0 | 1 | 2 | 2 | 1 | 18 | 1 | 43.4 |
| cough | 4 | 2 | 0 | 0 | 112 | 3 | 11 | 15 | 17 | 0 | 2 | 20 | 2 | 12 | 4 | 2 | 1 | 0 | 0 | 0 | 2 | 1 | 1 | 2 | 0 | 52.6 |
| * ear ache | 9 | 2 | 1 | 0 | 3 | 71 | 0 | 2 | 0 | 12 | 1 | 3 | 8 | 3 | 7 | 11 | 14 | 11 | 15 | 11 | 13 | 12 | 11 | 4 | 8 | 30.6 |
| emotional pain | 0 | 2 | 1 | 0 | 25 | 2 | 167 | 13 | 19 | 0 | 2 | 21 | 2 | 9 | 2 | 1 | 1 | 0 | 0 | 0 | 3 | 0 | 1 | 2 | 1 | 60.9 |
| feeling cold | 0 | 3 | 1 | 0 | 24 | 3 | 8 | 139 | 16 | 0 | 1 | 19 | 2 | 9 | 3 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 2 | 1 | 59.7 |
| * feeling dizzy | 0 | 4 | 2 | 1 | 18 | 2 | 9 | 13 | 112 | 2 | 0 | 23 | 0 | 12 | 1 | 4 | 2 | 1 | 1 | 1 | 3 | 1 | 0 | 0 | 0 | 52.8 |
| foot ache | 8 | 4 | 3 | 2 | 1 | 21 | 0 | 0 | 1 | 58 | 0 | 5 | 7 | 0 | 8 | 11 | 16 | 12 | 11 | 9 | 9 | 10 | 14 | 0 | 10 | 26.4 |
| hair falling out | 0 | 17 | 33 | 32 | 2 | 1 | 0 | 2 | 0 | 2 | 82 | 0 | 4 | 1 | 2 | 2 | 0 | 0 | 0 | 2 | 2 | 2 | 0 | 17 | 1 | 40.2 |
| heard to breath | 1 | 2 | 4 | 3 | 17 | 2 | 11 | 12 | 18 | 3 | 2 | 95 | 0 | 9 | 1 | 1 | 0 | 1 | 1 | 0 | 2 | 0 | 0 | 0 | 1 | 51.1 |
| *head ache | 9 | 3 | 2 | 4 | 2 | 23 | 0 | 3 | 1 | 11 | 3 | 1 | 121 | 1 | 7 | 13 | 13 | 9 | 13 | 16 | 8 | 14 | 9 | 2 | 2 | 41.7 |
| heart hurts | 1 | 3 | 2 | 4 | 14 | 1 | 11 | 11 | 17 | 3 | 4 | 26 | 2 | 159 | 2 | 1 | 0 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 60.0 |
| *Infected wound | 8 | 2 | 3 | 9 | 2 | 8 | 0 | 0 | 0 | 8 | 3 | 2 | 7 | 1 | 111 | 15 | 8 | 8 | 7 | 8 | 8 | 9 | 8 | 2 | 8 | 45.3 |
| injury from spots | 11 | 3 | 1 | 1 | 1 | 7 | 0 | 0 | 2 | 9 | 4 | 0 | 8 | 0 | 7 | 66 | 7 | 12 | 7 | 9 | 9 | 9 | 12 | 0 | 9 | 34.0 |
| internal pain | 13 | 2 | 2 | 2 | 1 | 8 | 0 | 0 | 1 | 7 | 3 | 0 | 6 | 0 | 7 | 10 | 61 | 11 | 7 | 7 | 12 | 6 | 14 | 0 | 10 | 32.1 |
| joint pain | 13 | 4 | 2 | 1 | 1 | 9 | 0 | 0 | 2 | 18 | 4 | 0 | 0 | 0 | 7 | 10 | 31 | 107 | 8 | 7 | 9 | 9 | 12 | 0 | 7 | 41.0 |
| knee pain | 10 | 3 | 1 | 0 | 0 | 7 | 0 | 0 | 1 | 12 | 3 | 0 | 8 | 1 | 8 | 9 | 8 | 10 | 105 | 7 | 8 | 6 | 11 | 1 | 9 | 46.1 |
| muscle pain | 12 | 2 | 0 | 0 | 1 | 8 | 0 | 0 | 1 | 11 | 2 | 0 | 7 | 2 | 7 | 11 | 6 | 5 | 8 | 106 | 9 | 17 | 14 | 0 | 8 | 44.7 |
| neck pain | 10 | 3 | 0 | 0 | 0 | 9 | 0 | 1 | 1 | 22 | 0 | 1 | 6 | 1 | 8 | 9 | 9 | 8 | 8 | 6 | 77 | 9 | 12 | 1 | 9 | 36.7 |
| open wound | 18 | 2 | 0 | 0 | 0 | 8 | 0 | 1 | 2 | 12 | 0 | 0 | 7 | 0 | 7 | 9 | 7 | 4 | 8 | 5 | 14 | 83 | 12 | 0 | 8 | 40.1 |
| shoulder pain | 14 | 2 | 0 | 0 | 0 | 8 | 0 | 0 | 3 | 17 | 0 | 1 | 6 | 0 | 6 | 8 | 7 | 8 | 8 | 5 | 11 | 14 | 70 | 1 | 8 | 35.5 |
| skin issue | 0 | 24 | 45 | 34 | 1 | 0 | 0 | 2 | 2 | 2 | 45 | 3 | 2 | 2 | 1 | 2 | 1 | 3 | 0 | 1 | 2 | 0 | 0 | 145 | 1 | 45.6 |
| *stomach ache | 11 | 2 | 0 | 1 | 0 | 13 | 1 | 2 | 1 | 3 | 3 | 3 | 6 | 3 | 7 | 11 | 9 | 5 | 6 | 9 | 8 | 8 | 10 | 2 | 114 | 47.9 |
| | 31.6 | 38.2 | 24.9 | 38.2 | 49.8 | 31.6 | 74.2 | 61.8 | 49.8 | 25.8 | 36.4 | 42.2 | 53.8 | 70.7 | 49.3 | 29.3 | 27.1 | 47.6 | 46.7 | 47.1 | 34.2 | 36.9 | 31.1 | 64.4 | 50.7 | 43.7 |

**Fig. 3.** The Confusion Matrix of the best classification accuracy gained from predicting 25 diseases in Exp 1 using NN.

human speech, which may be classified into two categories. To begin, psychological ailments such as acne, blurred eyesight, physical weakness, hair loss, and skin problems (which will be called Group 2 in Exp2). Second, disorders associated with frequency such as (cough, emotional discomfort, feeling chilly, difficulty breathing, heart ache, and dizziness) (which will be called Group 3 in Exp2).

Following the conclusion of our trials, we observed that several disorders had been clustered together. Each category of illnesses is misclassified within itself and not in relation to other groupings. When we classified those diseases into categories, as indicated in Table 1, we discovered that some behaviors were shared by diseases within the same category. Certain disorders are accompanied by pain, while others produce no discomfort but impair the human being's emotional state, and yet others have a direct effect on the vocal folds, affecting the frequency of speech. As a result, Experiment 2 was conducted.

### *7.2. Experiment number 2 (Exp2)*

The purpose of this experiment is to classify illnesses into three categories. Where each set of disorders has a similar pattern of behavior or effect on the human body or mind and speech. Following the implementation of this experiment, some conclusions were reached, such as:

1- High classification accuracy was achieved using all three classifiers used in this study, although the SVM classifier achieved the highest classification accuracy, as seen by the confusion matrix in Fig. 5.
2- Figs. 6 and 7 illustrate the classification results of the NN and GMM classifiers, respectively.
3- Figs. 5, 6, and 7, still show that there is a partial misclassification between the three prevented groups of disease, and this can be justified by the poor representation of the illness in the wave recordings, or there might be some intersection in the features extracted from the three different groups of diseases, that caused this misclassification.
4- The least classification accuracy was gained through the GMM, through classifying group 2, that was misclassified with group 1 via 164 wave samples.
5- No uniform misclassification was found in Exp 2, all three groups are misclassified with each other. the only uniform found was, group 1, achieved the highest classification accuracy with respect to all three classifiers, 96.83%, 94.35%, and 94.92%, with SVM, NN, and GMM, respectively.

| | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| back pain | 34 | 2 | 3 | 2 | 1 | 11 | 1 | 1 | 2 | 12 | 0 | 2 | 19 | 1 | 8 | 9 | 15 | 16 | 14 | 12 | 14 | 11 | 9 | 0 | 9 | 16.3 |
| acne | 2 | 56 | 34 | 24 | 2 | 2 | 1 | 2 | 1 | 2 | 18 | 1 | 2 | 2 | 1 | 1 | 2 | 0 | 1 | 0 | 1 | 1 | 1 | 22 | 1 | 31.1 |
| blurry vision | 0 | 30 | 54 | 33 | 1 | 2 | 0 | 1 | 2 | 3 | 21 | 1 | 2 | 1 | 1 | 2 | 3 | 0 | 1 | 0 | 1 | 1 | 0 | 21 | 1 | 29.7 |
| body feels weak | 3 | 31 | 27 | 70 | 1 | 1 | 1 | 2 | 1 | 2 | 21 | 0 | 3 | 1 | 0 | 1 | 2 | 0 | 1 | 0 | 1 | 1 | 0 | 19 | 1 | 36.8 |
| cough | 4 | 2 | 3 | 3 | 68 | 4 | 20 | 18 | 20 | 2 | 1 | 28 | 0 | 11 | 0 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 35.4 |
| * ear ache | 13 | 3 | 4 | 5 | 2 | 56 | 3 | 2 | 1 | 12 | 2 | 2 | 19 | 1 | 9 | 8 | 11 | 9 | 10 | 10 | 12 | 8 | 12 | 0 | 7 | 25.3 |
| emotional pain | 0 | 3 | 1 | 3 | 23 | 2 | 88 | 15 | 28 | 2 | 0 | 29 | 0 | 17 | 0 | 1 | 1 | 1 | 0 | 1 | 3 | 0 | 2 | 0 | 0 | 40.0 |
| feeling cold | 4 | 2 | 2 | 2 | 20 | 3 | 23 | 78 | 29 | 2 | 1 | 21 | 0 | 18 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 2 | 1 | 0 | 37.0 |
| * feeling dizzy | 2 | 0 | 2 | 2 | 33 | 2 | 23 | 27 | 59 | 0 | 3 | 27 | 1 | 14 | 0 | 0 | 0 | 0 | 0 | 13 | 0 | 0 | 0 | 0 | 1 | 28.2 |
| foot ache | 9 | 0 | 2 | 2 | 1 | 22 | 1 | 3 | 2 | 58 | 0 | 3 | 16 | 2 | 11 | 12 | 12 | 16 | 15 | 9 | 15 | 16 | 15 | 1 | 10 | 22.9 |
| hair falling out | 1 | 32 | 24 | 24 | 2 | 3 | 1 | 1 | 2 | 0 | 73 | 1 | 1 | 3 | 0 | 2 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 32 | 2 | 34.9 |
| heard to breath | 1 | 2 | 4 | 4 | 22 | 3 | 27 | 24 | 33 | 0 | 1 | 67 | 1 | 20 | 1 | 0 | 1 | 0 | 0 | 0 | 2 | 1 | 0 | 1 | 0 | 31.2 |
| *head ache | 12 | 4 | 2 | 2 | 4 | 21 | 2 | 3 | 1 | 11 | 31 | 0 | 54 | 2 | 15 | 10 | 11 | 6 | 12 | 8 | 8 | 11 | 7 | 1 | 12 | 21.6 |
| heart hurts | 0 | 4 | 4 | 4 | 21 | 4 | 24 | 21 | 31 | 1 | 1 | 29 | 0 | 111 | 0 | 0 | 1 | 0 | 0 | 2 | 1 | 1 | 0 | 1 | 0 | 42.5 |
| *infected wound | 10 | 2 | 2 | 2 | 3 | 8 | 1 | 2 | 3 | 14 | 19 | 1 | 18 | 0 | 101 | 10 | 13 | 8 | 8 | 9 | 9 | 10 | 9 | 0 | 7 | 37.5 |
| injury from spots | 14 | 3 | 3 | 1 | 2 | 12 | 1 | 2 | 2 | 15 | 0 | 1 | 12 | 2 | 8 | 66 | 13 | 18 | 9 | 11 | 8 | 17 | 9 | 1 | 8 | 27.7 |
| internal pain | 17 | 4 | 4 | 3 | 4 | 11 | 1 | 3 | 1 | 11 | 0 | 0 | 13 | 3 | 8 | 20 | 56 | 15 | 11 | 12 | 7 | 11 | 8 | 1 | 7 | 24.2 |
| joint pain | 15 | 2 | 3 | 2 | 5 | 11 | 0 | 2 | 1 | 10 | 0 | 1 | 11 | 2 | 9 | 10 | 14 | 68 | 12 | 11 | 9 | 15 | 9 | 0 | 7 | 29.7 |
| knee pain | 13 | 1 | 4 | 2 | 3 | 7 | 1 | 4 | 1 | 9 | 0 | 0 | 10 | 4 | 8 | 15 | 11 | 11 | 65 | 10 | 11 | 16 | 9 | 2 | 7 | 29.0 |
| muscle pain | 14 | 3 | 2 | 2 | 2 | 6 | 0 | 2 | 1 | 8 | 1 | 1 | 9 | 2 | 9 | 14 | 15 | 11 | 14 | 71 | 9 | 14 | 13 | 4 | 7 | 30.3 |
| neck pain | 17 | 2 | 2 | 0 | 1 | 7 | 1 | 4 | 1 | 13 | 1 | 2 | 9 | 5 | 8 | 13 | 11 | 9 | 14 | 9 | 70 | 13 | 10 | 1 | 8 | 30.3 |
| open wound | 14 | 0 | 2 | 0 | 1 | 8 | 1 | 2 | 0 | 11 | 1 | 3 | 9 | 1 | 8 | 8 | 10 | 9 | 13 | 16 | 11 | 50 | 10 | 2 | 9 | 25.1 |
| shoulder pain | 16 | 0 | 3 | 0 | 0 | 8 | 0 | 3 | 3 | 11 | 0 | 0 | 8 | 0 | 9 | 11 | 10 | 14 | 12 | 10 | 13 | 11 | 88 | 1 | 9 | 36.7 |
| skin issue | 1 | 35 | 30 | 32 | 1 | 4 | 4 | 1 | 0 | 1 | 29 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 3 | 112 | 0 | 43.4 |
| *stomach ache | 9 | 2 | 4 | 1 | 2 | 7 | 2 | 2 | 0 | 15 | 1 | 4 | 8 | 1 | 11 | 11 | 9 | 13 | 9 | 8 | 17 | 14 | 9 | 2 | 112 | 41.0 |
| | 15.1 | 24.9 | 24.0 | 31.1 | 30.2 | 24.9 | 39.1 | 34.7 | 26.2 | 25.8 | 32.4 | 29.8 | 24.0 | 49.3 | 44.9 | 29.3 | 24.9 | 30.2 | 28.9 | 31.6 | 31.1 | 22.2 | 39.1 | 49.8 | 49.8 | 31.7 |

**Fig. 4.** The Confusion Matrix of the best classification accuracy gained from predicting 25 diseases in Exp 1 using GMM.

**Table 1**
The groups of diseases generated after analyzing the results of Exp1.

| No. | Acoustic phonetic feature related disease | | Articulator phonetics feature related disease |
|---|---|---|---|
| | Frequency Related | Psychological Related | Painful diseases |
| **1** | cough | acne | back pain |
| **2** | emotional pain | blurry vision | internal pain |
| **3** | feeling cold | body feels weak | joint pain |
| **4** | heard to breath | hair falling out | knee pain |
| **5** | heart hurts | skin issue | muscle pain |
| **6** | feeling dizzy | | neck pain |
| **7** | | | open wound |
| **8** | | | shoulder pain |
| **9** | | | stomach ache |
| **10** | | | injury from spots |
| **11** | | | infected wound |
| **12** | | | head ache |
| **13** | | | ear ache |
| **14** | | | Foot ache |

| | | | | |
|---|---|---|---|---|
| Group1 | 3050 | 50 | 69 | 96.24% |
| Group 2 | 42 | 1056 | 26 | 93.95% |
| Group 3 | 58 | 19 | 1255 | 94.22% |
| | 96.83% | 93.87% | 92.96% | 94.55% |

**Fig. 5.** The Confusion Matrix of the best classification accuracy gained from predicting 3 groups of diseases using SVM.

| | | | | |
|---|---|---|---|---|
| Group1 | 2972 | 113 | 72 | 94.14% |
| Group 2 | 91 | 961 | 35 | 88.41% |
| Group 3 | 89 | 51 | 1243 | 89.88% |
| | 94.35% | 85.42% | 92.07% | 90.62% |

**Fig. 6.** The Confusion Matrix of the best classification accuracy gained from predicting 3 groups of diseases using NN.

| | | | | |
|---|---|---|---|---|
| Group1 | 2990 | 164 | 146 | 90.61% |
| Group 2 | 76 | 904 | 39 | 88.71% |
| Group 3 | 84 | 57 | 1167 | 89.22% |
| | 94.92% | 80.36% | 86.44% | 87.24% |

**Fig. 7.** The Confusion Matrix of the best classification accuracy gained from predicting 3 groups of diseases using GMM.

6- Group 2 have achieved the lowest classification accuracies with NN and GMM, it might be justified by its poor wave samples, but as long as the same groups have achieved the second place in classification with SVM, so that justification is not totally true, but the right justification is, the three group of diseases, still have common features, that are leading to this misclassification.

The limitation of the proposed work, was the time needed by the genetic algorithm to select the best group of features. Adding to this limitation, through experiment, it wasn't applicable to diagnose the 25 diseases represented in the dataset, therefore, it was a must to group identical diseases in three different groups. Those groups were fixed through try and error during experiment. Diseases that are related to acoustic phonetic feature were determined, and diseases related to articulator phonetics feature were also determined. Then the first was divided to frequency related diseases and psychological related diseases.

## 8. Conclusion

Diagnosing illnesses via speech is a difficult process that was previously unattainable, but with the advent of machine learning, everything is conceivable. Our objective was to diagnose illnesses directly from human speech using a machine learning approach, however we were unable to get the desired results, despite the 50.1 percent classification accuracy achieved using the SVM classifier, but on the other hand, high accuracy result of (94.55%) was gained when divided the diseases into three related groups.

This investigation revealed that illnesses may be classified into several categories, including severe and light discomfort, psychological or physical, emotional or non-emotional, and others. Each of the above groups exhibits similar behavior when it comes to communication. This reflection leaves distinct characteristics associated with each form of sickness.

The advantage of the proposed work is the capability of illness diagnoses through speech, which is considered a new approach in this field, and is extremely required in certain applications were speech is the only source of information available to diagnose the illness of a patient.

Through experiment, the proposed work faced a couple of limitation. First, the time, the needed by the genetic algorithm was high. Second, the failure of the proposed work to diagnose each of the 25 diseases represented in the dataset utilized, which means the failure is extracting features that can diagnose each of the 25 diseases separately.

## Future Work

The findings of Exp1 can be increased by using more feature types or machine learning methods. Although the genetic technique utilized in Exp2 produced positive findings, it was a time-consuming procedure. Thus, developing a more efficient feature selection approach can accelerate the ADPS.

As the challenges and the extreme number of random factors of this field have been discussed in the introduction section, the response service methodology (RSM) program can be used to predict the perfect setting to the system proposed, by suggesting the experiments with the highest probability of achieving the best classification results.

Searching for features that are more disease related, that can diagnose each disease separately, will be more applicable, because when patients use ADPS's, they usually are looking for exclusive answers, not broad answers.

The input of this work is only speech, but adding other data such as human hand or arm movement as an input, and using both types of data as combination, and predicting the disease accordingly, might increase the classification accuracy, since human bodies translate pain through moving parts of it, such as moving their arms, hand, shoulders, toughing their foreheads or legs.

## Declaration of Competing Interest

None.

## Data availability

The authors are unable or have chosen not to specify which data has been used.

## References


[1] Zhang Z. Mechanics of human voice production and control. J Acoust Soc Am 2016;140(4):2614–35.
[2] Carr P. English phonetics and phonology: An introduction. John Wiley & Sons; 2019. p. 382.
[3] Anderson C. Essentials of Linguistics. Pressbooks by McMaster University; 2018.
[4] Rabiner LR, Schafer RW. Digital processing of speech signal. Pearson; 1978. p. 528.
[5] Deller JR, Proakis JG, Hansen JH. Discrete time processing of speech signals. Prentice Hall PTR; 1993.
[6] Harrington J, Cassidy S. The acoustic theory of speech production, in Techniques in Speech Acoustics. Springer; 1999. p. 29–56.
[7] Rose P. Forensic Speaker Identification, 1. London: Taylor Francis; 2002. p. 380.
[8] Kreiman J, Papçun G. Voice discrimination by two listener populations. J Acoust Soc Am 1985;77(S1). S9–S9.
[9] EUiott JR. Auditory and F-pattern variations in Australian okay: a forensic investigation. Acoust Aust 2001;(29):41.
[10] Vickers ER, Cousins MJ, Woodhouse AJ. Pain description and severity of chronic orofacial pain conditions. Aust Dent J 1998;43(6):403–9.
[11] Labus JS, Keefe FJ, Jensen MP. Self-reports of pain intensity and direct observations of pain behavior: when are they correlated? Elsevier Sci B.V. 2003;102 (1–2):109–24.
[12] Sullivan MJ, Pascal T, André S, Richard C, John K, William S. The influence of communication goals and physical demands on different dimensions of pain behavior. Elsevier Sci B.V. 2006;125(3):270–7.
[13] Rowbotham S, Holler J, Lloyd D, Wearden A. How do we communicate about pain? A systematic analysis of the semantic contribution of co-speech gestures in pain-focused conversations. J Nonverbal Behav 2012;36(1):1–21.
[14] Rowbotham S, Holler J, Lloyd D, Wearden A. Handling pain: The semantic interplay of speech and co-speech hand gestures in the description of pain sensations. Speech Commun 2014;57:244–56.
[15] Wang X, Chused A, Elhadad N, Friedman C, Markatou M. Automated knowledge acquisition from clinical narrative reports. In: AMIA Annual Symposium Proceedings. American Medical Informatics Association; 2008.
[16] Hossain M, Laskar M, Rahman T. Automated disease prediction system (ADPS): a user input-based reliable architecture for disease prediction. Int J Comput Appl 2015.
[17] Dragu N, Elkhoury F, Miyazaki T, Morelli RA, di Tada N. Ontology-based text mining for predicting disease outbreaks. In: Twenty-Third International FLAIRS Conference; 2010.
[18] Maude J. Patients could provide initial differential. Br J Gen Pract 2021.
[19] Shepperd S, Charnock D, Gann B. Helping patients access high quality health information. Br Med J (Clin Res Ed) 1999;319(7212):764–6.
[20] Das R, Turkoglu I, Sengur A. Effective diagnosis of heart disease through neural networks ensembles. Expert Syst Appl 2009;36(4):7675–80.
[21] Kumar R, Pradhan S, Rebaka T, Prakash J. Disease Prediction from Speech Using Natural Language Processing and Deep Learning Method. In: Congress on Intelligent Systems. Springer; 2020.
[22] Dreisbach C, Koleck TA, Bourne PE, Bakken S. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. Int J Med Informatics 2019;125:37–46.
[23] Gangavarapu T, Krishnan GS, Kamath S, Jeganathan J. FarSight: long-term disease prediction using unstructured clinical nursing notes. IEEE Trans Emerg Topics Comput 2020;9(3):1151–69.
[24] Johnson M, Lapkin S, Long V, Sanchez P, Suominen H, Basilakis J, et al. A systematic review of speech recognition technology in health care. BMC Med Inform 2014;14(1):1–14.
[25] Shaqra FA, Duwairi R, Al-Ayyoub M. Recognizing emotion from speech based on age and gender using hierarchical models. Procedia Comput Sci 2019;151: 37–44.
[26] Abdulmohsin HA, Wahab HA, Hossen AMJA. Speech Emotion Recognition Survey. J Mech Conti. Math Sci 2020;15(9):24.
[27] Mooney, P., medical speech, transcription, intent, 2018, 1.
[28] Gustafsson F. Determining the initial states in forward-backward filtering. IEEE Trans Signal Process 1996;44(4):988–92.
[29] Chandrashekar G, Sahin F. A survey on feature selection methods. Comput Electr Eng 2014;40(1):16–28.
[30] Fant G. Acoustic theory of speech production. Walter de Gruyter; 1970.
[31] Nghia PT, Van Tao N, Huong PTM, Diep NTB, Hien PTT. A Measure of Smoothness in Synthesized Speech. REV Journal on Electronics Communications in Mathematical Physics 2016;6(1-2):35–9.
[32] Yoon H, Park C-S, Kim JS, Baek J-G. Algorithm learning based neural network integrating feature selection and classification. Expert Syst Appl 2013;40(1): 231–41.
[33] Ledesma S, Cerda G, Avina G, Hernandez D, Torres M. Feature selection using artificial neural networks. In: Mexican International Conference on Artificial Intelligence. Springer; 2008. p. 351–9.
[34] Abdulmohsin HA, Wahab HBA, Hossen AMJA. A New Hybrid Feature Selection Method Using T-test and Fitness Function. CMC-COMPUTERS MATERIALS CONTINUA 2021;68(3):3997–4016.
[35] Kohavi R, John GH. Wrappers for feature subset selection. Artif Intell 1997;97(1–2):273–324.



Asst. Prof. Dr. Husam Ali Abdulmohsin (M'41) was born in Baghdad, Iraq in 1979. This author became a Member (M) of IEEE in 2020. He received the B.S. degree in 2000 and M.S. degrees in 2003, both in computer science from the University of Al-Nahrain, Baghdad, Iraq. Fulfilled his Ph.D degree in the computer science department, Technology University, Baghdad, Iraq.

Prof. Dr. Belal Al-Khateeb received the B.Sc. (honors) (first class) degree in computer science from Al-Nahrain University, Baghdad, IRAQ, in 2000, and the M.Sc. degree in computer science from Al-Nahrain University, Baghdad, IRAQ, in 2003, and the Ph.D. degree from the School of Computer Science, University of Nottingham, Nottingham, U.K., in 2011.

Asst. Prof. Dr. Samer Sami hasan was born in Baghdad, Iraq in 1978. This author became a Member (M) of IEEE in 2014. He received the B.S. degree in 2000 and M.S. degrees in 2003, both in computer science from the University of Al-Nahrain, Baghdad, Iraq. Fulfilled his Ph.D degree in the computer science department, UKM University, SELANGOR, MALAYSIA.

Dr Rinky Dwivedi has completed her B.Tech in Computer Science and Engineering from Guru Gobind Singh Indraprastha University, Delhi in 2004 and M.E. in Computer Technology and Application from Delhi College of Engineering, Delhi in 2008. She has received her Doctorate in 2016 from Delhi Technological University, New Delhi. Dr. Rinky has over 19 years of experience in Academics.

**How to Cite**:
Mangla, R., & Chetna, C. (2022). Brain tumor detection and classification by MRI images using deep learning techniques. *International Journal of Health Sciences*, *6*(S3), 5783–5790. https://doi.org/10.53730/ijhs.v6nS3.7233

# Brain tumor detection and classification by MRI images using deep learning techniques

**Rashika Mangla**
Department of Applied Mathematics, Delhi Technological University, New Delhi, India
*Corresponding author email: rashika123mangla@gmail.com

**Chetna**
Department of Applied Mathematics, Delhi Technological University, New Delhi, India
Email: chetnamor4family@gmail.com

***Abstract*---**MRI images play a significant influence in brain tumor classification and detection but instead of having detection and classification using the medical equipment which is a radiologists or clinical professionals do a time-consuming and laborious task where accuracy depends only on the experience only, it can be beneficial to detect and classify the brain tumor by deep learning techniques and algorithms. As a result, the employment of computer-assisted technology becomes increasingly important in order to overcome these constraints. In this paper, the early detected and diagnosed brain tumor images along with their csv data has been used to find out the accuracy of the CNN algorithm for tumor detection and SVM algorithm for tumor classification into benign and malignant. HOG has been used for the feature extraction. After performing the experiment, it was observed that CNN achieved the detection accuracy of 87.02% and further tumor classification by employing SVM, the highest accuracy achieved was 96.35%. The experiment proved a very good accuracy of detection and classification even after using three different methods in the whole procedure.

***Keywords*---**CNN (Convolutional neural network), HOG (Histogram of Oriented Gradients), SVM (Support Vector Machine), Magnetic Resonance Imaging (MRI).

## Introduction

A brain tumor is an abnormal growth in the brain that can be harmful or even cancer-free. Uncontrolled cell proliferation and excessive brain damage cause

tumors in the brain. Primary and secondary tumors are two types of brain tumors that result in malignant or benign brain tumors. Gliomas are the most widely used sort of tumor. In order to collect relevant clinical data, such as the presence of a tumor, location, and type, computer-assisted devices can be used to perform the procedure automatically. However, determining their composition, volume, parameters, plant detection, size and classification remains a difficult task. Because it provides high visual acuity to soft tissues and is non-invasive, magnetic resonance imaging (MRI) [8] is preferred over other therapies and diagnostic methods. Signs and symptoms of a tumor in the brain vary depending on the type and location of the tumor. Because different parts of the brain control different aspects of body processes, some plants do not show symptoms until they are extremely large and cause severe and rapid loss of health. Hearing problems, problems with balance, speech impairment, changes in vision, memory problems, mobility problems, personality changes, lack of concentration, and a weak point in one part of the body are some of the most common symptoms.

Gliomas are divided into three categories, including Oligodendroglioma, Astrocytoma, and Glioblastoma, regardless of the WHO classification or classification of tumor. Due to the remarkable diversity and structure of MRI data, Brain Tumor is a difficult task. Sound and image clarity are limits on MRI imaging. Unwanted information in pictures is called sound. Noise can sometimes affect the edges and details, reducing the brightness adjustment. Because it is difficult to find specific boundaries and to distinguish a tumor from noise, it is an emerging topic of image processing research. In this study, we compare and contrast an effective and efficient method of brain differentiation, detection, and differentiation. Many brain tumor classification techniques have recently been introduced, which may be categorized by machine learning and deep learning (DL) techniques such as CNN, GLCM, OTSU'S THRESHOLDING, SVM, and others, which we will use based on feature selection and learning methods.

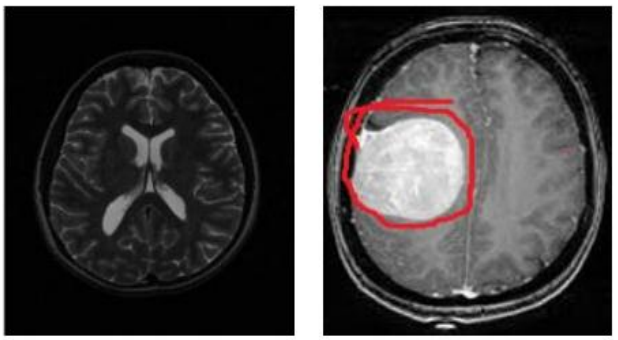



Fig 1. A healthy brain (a) and a tumor-affected brain (b) are depicted in these photos of a human brain.

**Literature review**

Machine learning techniques have been widely used in a variety of fields, including medical diagnosis and prevention. However, few studies have focused on the diagnosis of brain tissue using magnetic resonance imaging (MRI). Most DL

methods use MRI data to train and test old DL algorithms. DL has recently been used to diagnose brain tumors in a number of techniques. G. Hemanth et al. [1] suggested Automated Brain Tumors Automatic Detection System. CNN is used for classifying and identifying tumors. Then a pixel-based detection method is used to obtain a picture of the brain and other affected areas. Then the separation using the extracted features is done using CNN. They found 91% accuracy in sections. Praveena Pedapati and Rama Vaishnavi Tanneedi [2] propose to use SVM and in-depth study to detect and differentiate brain cancer automatically. However, in this case, SVM failed to provide the highest level of accuracy. The accuracy of the features extracted from high glioma (Malignant) images without the use of histogram of the targeted gradients was 97 percent in high glioma images. Low glioma (Benign) images have an accuracy of 68.75 percent. In [3] the authors performed Brain tumor CT imaging classification into malignant and malignant images using SVM with kernel function. The best performance of the categories is obtained through WSVM, based on test data. Moreover, these findings suggest that the proposed method is effective and effective in predicting malignant and malignant tumors from CT scans of the brain, with 74% accuracy of SVM and 76% of WSVM.

The proposed method may be consistent with other types of imaging, such as MRI, in the future, and it may be used to classify and differentiate tumors from other parts of the body. The authors of another study [4] used the effectiveness of pressure strategies to improve accuracy as well as timing of CNN executions in differentiation by brain MRI. Before they were divided into sections, they presented the first step in processing. The method first uses the Probabilistic Neural Network (PNN) to differentiate the Genital Region (ROI) (especially the area of the brain tissue), and then uses Back Propagation Neural Networks to narrow the ROI (BPNN). Finally, compressed images are uploaded to CNN segregation. For comparison, results were obtained for three different grades of formats and the accuracy levels displayed were more than 90%. R-CNN's rapid method of successful detection and classification of brain MRI images was proposed by the authors of another paper [5]. Images are first embedded in a simple CNN, which produces a dynamic feature map, which is then changed to regional suggestions, which are, after that, resized into a feature vector by the integration stage of ROI.

Finally, by categories, this ROI vector feature is integrated into the fast R-CNN. SVM was also used to create the largest margin possible between classes, allowing the algorithm to classify images into categories with greater accuracy. Their algorithm had an accuracy rate of 95%. [6] proposes a flexible approach based on in-depth study of Generative Adversarial Networks (GANs). Their plan was to previously train CNN as a GAN racist using two databases. The productive part of GAN was aided by the development of data on the production of more realistic MRI scans of the brain. To separate, the final bed of CNN discrimination on GAN is changed by SoftMax. In the Figshare database, this precise CNN separator is fine-tuned over time. The outcomes reveal 88 percent precision. [7] Due to the complexity of imaging and there aren't any anatomical models that accurately represent the possible deformities in each and every component, the separation of medical imaging is a difficult task. When it comes to the original size of the collection, the proposed method is quite effective and collection centers. This function recommends a system that requires non-invasive human intervention to

isolate tissue of the brain. The prime purpose of this suggested program is to assist human specialists or surgeons in locating patients in less time.

**Methodology**

The proposed work has been implemented in four phases. The experimental setup involved a system with 4GB Ram, Windows 10 and implemented via Python.

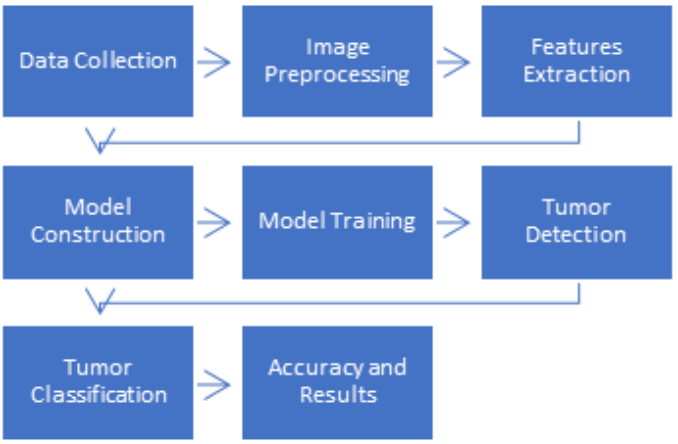



Fig 2. Steps followed for tumor detection and classification

**Phase-I**
**Data Collection and Image Pre-processing**

- Data collection - The standard available dataset [13] was collected for the implementation. The dataset contains 2212 brain MRI images in two folders: yes and no. Yes, there are 1112 tumorous brain MRI images in the folder, and no, there are 1100 non-tumorous brain MRI images in the folder.
- Data Pre-processing - The information we gathered came in a variety of shapes, sizes, and orientations. As a result, the photos had to be resized to a shape of (240, 240, 3) = (image width, image height, number of channels). As a result, all photographs should have the same shape to transmit as an input into the neural network. Normalization was the next step in the process. Normalization is a technique for altering the range of pixel intensity values in image processing. Normalization, also known as histogram stretching or contrast stretching, can be used on photos with poor contrast due to glare. The photos were scaled to the range 0-1 so that they could be viewed regularly. To do so, multiply all of the pixel values by the largest pixel. This equals 255.

**Phase-II**
**Features Extraction using HOG**

HOG [11] is a feature descriptor that is used in image processing to recognize objects. The feature descriptor's goal is to generalize an object in an image so that it offers the same feature descriptors in all images that contain that object, regardless of angle, illumination, distance, or other circumstances. In a detection window, or region of interest, the HOG descriptor approach counts occurrences of 22 gradient orientations. in a limited section of a picture (ROI). One of the most significant topics in computer vision is the gradient vector. A gradient vector is sometimes known as a picture gradient. Feature extraction and edge detection are two applications of gradient vectors. The difference of neighborhood values in both the horizontal and vertical axes is used to calculate the gradient vector for a certain pixel. In HOG, orientation is crucial. The term "orientation" refers to a shift in direction when the pixel intensity value changes. The direction of change can be along the X-axis or the Y-axis. Orientation = arctan (Y/X) is the formula for determining the orientation of a gradient vector. The foundation for histogram grouping and normalization is the grouping of cells into blocks. The normalized group of cells is called a block histogram. The feature descriptor is represented as a set of block histograms. A feature descriptor is a technical term that refers to a picture or a portion of an image that simplifies the original form of the image by removing the key data. In HOG, the distribution of gradient orientation is used as a feature. The picture, orientations, pixels per cell, cells per block, visualize, and multichannel parameters are used as input to the Hog() function.

**Phase-III**
**Tumor Detection using CNN**

A CNN is a Deep Learning system that can take an image as input and give significance to distinct aspects/objects in the image, as well as distinguish between them. In comparison to other partitioning algorithms, ConvNet just requires a little amount of computation. Despite the fact that the fundamental approaches need manual filter manipulation, ConvNets can learn these filters / symbols with enough training. ConvNet can successfully capture Local and Temporary dependencies on a picture by using the right filters. Architecture creates a high degree of image databases because of the reduced number of parameters involved and the reusability of weights. In other words, the network may be trained to recognize higher-quality image technologies. In order to detect brain cancers in multiple MRI scans of the brain we performed detection using various algorithms Like CNN, SVM, OTSU'S THRESHOLDING, KNN and Random Forest. But among all algorithms the best one found out with highest accuracy [12] was CNN. The formula used to determine the model's accuracy is –

$$\text{ACCURACY} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where, TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative. TP and TN mean the samples which have been detected accurately and FP and FN mean the samples which have not been detected accurately.

**Phase-IV**
**Tumor Classification using SVM**

A SVM algorithm's purpose is to find a hyperplane in an N-dimensional space that clearly separates data points. To separate the two sorts of data points, there are several hyperplanes to pick from. Our objective is to locate a plane with a large limit, or a large distance between data points from both groups. Increasing the gene range adds some support, making it simpler to succeed in distinguishing the following data points. SVM [10] has the following advantages: It works really well with a large separation margin. It works well in three-dimensional areas. If the maximum size exceeds the sample size, this method works best. Memory works well because it uses a set of training points (called supporting vectors) in the decision-making process. The HOG characteristics that we retrieved from the MRI images were saved in a.csv file. We employed a radial basis function to accomplish the processing, which involves mapping data into a higher-dimensional space. Kernelling is the term for this process, and the kernel function is the mathematical function that is utilized to alter the data. Then we have fitted our model and predicted the new values.

**Results**

For the CNN algorithm the data collected and the detection done by our algorithm is as follows –

Table 1
Positive and Negative brain tumor raw data

| | No. of MRI Images | Positive MRI Images | Negative MRI Images |
|---|---|---|---|
| RAW DATA | 2212 | 1112 | 1100 |

From all the images of dataset, the accuracy of detection of images is as follows –

Table 2
Positive and Negative brain tumor detected data

| | TP | TN | FP | FN |
|---|---|---|---|---|
| No. of MRI Images detected | 983 | 942 | 58 | 129 |

So, using the result the accuracy of our model CNN is 87.02%. For the SVM algorithm the data taken and the classification into benign and malignant done by our algorithm is as follows - In the dataset that we took for our CNN model - 983 were truly detected positive for brain tumor, this data was then split into training and testing data. 137 were kept into testing data and 846 were kept into the training dataset. Out of these 137-testing dataset, the data of 90 brain tumors

were benign and 47 were malignant. But after evaluating the data in our SVM algorithm, out of these 137 tumors, only 85 were detected to be benign positively and 47 were detected to be malignant positively whereas 5 were detected to be false malignant and there were no false benign tumors.

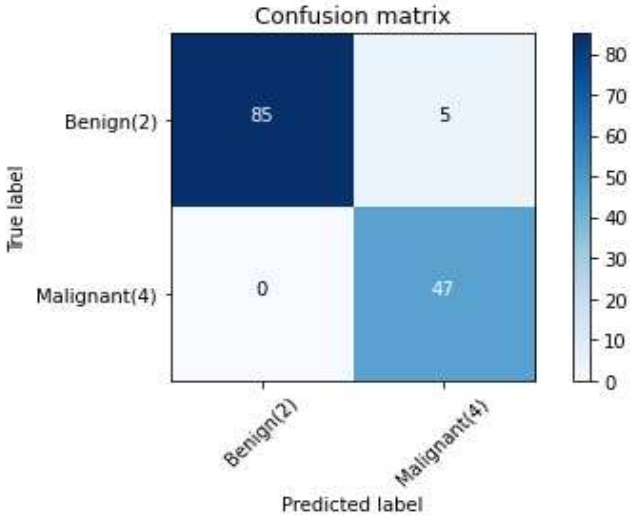



Fig 3. Confusion Matrix

From this matrix it is clear that the benign tumors truly detected are 85 and the malignant tumors truly detected are 47. While the false malignant tumors detected are 5 in number however there are no benign tumors detected falsely. So, the accuracy of our SVM model is 96.35%.

**Conclusion**

The most important aspect of our proposed work was how to use a CNN to detect the presence of a tumor and a HOG to extract features from brain tumor MRI images, followed by a classification method, SVM, to classify brain tumors into benign and malignant. An approach based on a combination of the feature extraction algorithm (HOG), CNN, and SVM for tumor identification and classification from brain pictures is provided. CNN has the ability to identify tumors. When it comes to picking an auto-feature in medical photographs, the CNN comes in handy. Clinicians classified the images obtained at the centers, and then tumor screenings were divided into two categories: normal and patient. The proportion of sick to healthy subjects was proportionate to the proportion of image categorization in two classes. After pre-processing, the images were fed into the CNN. CNN's accuracy is 87.02 percent for appropriately categorizing images into two normal and patient classes. The accuracy of the proposed algorithm increased to 96.35 percent on the test data while using the proposed approach of feature extraction and using SVM for categorization into benign and malignant tumors, which is an improvement over the CNN. Due to the importance of the physician's diagnosis, the model's accuracy can assist doctors in diagnosing the tumor and treating the patient, resulting in a high level of medical accuracy when utilizing the proposed technique. This model can be expanded in the future to

further classify benign and malignant tumors into grades of different types (Grade 1, Grade 2, Grade 3, Grade 4). This could aid the medical system in providing more effective and timely treatments.

**References**


1. G. Hemanth, M. Janardhan and L. Sujihelen, "Design and Implementing Brain Tumor Detection Using Machine Learning Approach," *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2019, pp. 1289-1294
2. Pedapati, Praveena, and Rama Vaishnavi Tannedi. "Brain Tumor Detection using HOG by SVM." (2018).
3. Rezaei, Kimia, Hamed Agahi, and David C. Wyld. "Segmentation and Classification of Brain Tumor CT Images Using SVM with Weighted Kernel Width." *Computer Science & Information Technology* 39 (2017).
4. El Boustani, Abdelhakim, and Essaid El Bachari. "MRI Brain Images Compression and Classification Using Different Classes of Neural Networks." In *International Conference on Model and Data Engineering*, pp. 122-134. Springer, Cham, 2019.
5. Poonguzhali, N., Kagne Raveena Rajendra, T. Mageswari, and T. Pavithra. "Heterogeneous Deep Neural Network for Healthcare Using Metric Learning." In *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, pp. 1-4. IEEE, 2019.
6. Ghassemi, Navid, Afshin Shoeibi, and Modjtaba Rouhani. "Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images." *Biomedical Signal Processing and Control* 57 (2020): 101678.
7. Arif, Muhammad, F. Ajesh, Shermin Shamsudheen, Oana Geman, Diana Izdrui, and Dragos Vicoveanu. "Brain Tumor Detection and Classification by MRI Using Biologically Inspired Orthogonal Wavelet Transform and Deep Learning Techniques." *Journal of Healthcare Engineering* 2022 (2022).
8. https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri
9. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53
10. https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47
11. https://towardsdatascience.com/hog-histogram-of-oriented-gradients-67ecd887675f
12. https://deepchecks.com/how-to-check-the-accuracy-of-your-machine-learning-model/
13. https://www.kaggle.com/jakeshbohaju/brain-tumor

TechRxiv
Powered by IEEE

# Complaint and Severity Identification from Online Financial Content

This paper was downloaded from TechRxiv (https://www.techrxiv.org).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

17-06-2022 / 21-06-2022

CITATION

Bhatia, Rohan; Singh, Apoorva; Saha, Sriparna (2022): Complaint and Severity Identification from Online Financial Content. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.20087675.v1

DOI

10.36227/techrxiv.20087675.v1

# Complaint and Severity Identification from Online Financial Content

Apoorva Singh, Rohan Bhatia, and Sriparna Saha

***Abstract*—The automatic detection of financial complaints can benefit businesses and online merchants. Compared to manually tagged complaints, they can use this information to monitor and address issues and effectively route them to appropriate teams. This can also promote greater transparency and accountability when dealing with consumer financial products and services, strengthening the firm's brand value. In linguistic studies, complaints have been classified into severity categories based on the level of risk the complainant is prepared to accept. Furthermore, since emotions influence every speech act, an individual's emotional state considerably impacts the complaint expression. In this paper, we introduce a Financial Complaints resource, a collection of annotated complaints arising between financial institutions and consumers expressed in English on Twitter. The dataset has been enriched with the associated emotion, sentiment, and complaint severity classes. The dataset comprises 3149 complaint and 3133 non-complaint instances spanning over ten domains (e.g., credit cards, mortgages, etc.). For a comprehensive evaluation of our dataset, we develop a multi-task framework for complaint detection and severity classification with emotion recognition and sentiment classification as the additional tasks and compare it with several existing baselines[1].**

***Index Terms*—Complaint Corpus, Complaint Detection, Complaint Severity Detection, Deep learning, Finance Complaints, Multi-task learning**

## I. Introduction

ARTIFICIAL INTELLIGENCE (AI) is transforming how people and organizations access and manage their finances. While the shift from traditional financial services to digital finance was already underway pre-pandemic, the pandemic accelerated the process as stay-at-home requests became the new normal for financial institutions, and customers decided to seek more self-service options. AI and machine learning in finance cover everything from fraud detection to chatbot assistants and task automation. Other than these mainstream applications, addressing digital service complaints could be a significant commercial application for AI in finance. Compared to manually tagged complaints, automated detection of financial complaints pertaining to fraudulent transactions, delayed fund transfers, lousy customer service, etc., could help direct them to appropriate teams. This can further help promote fairness and transparency while dealing with loans, credit cards, and other consumer financial products and services, thus improving customer experience and increasing the organization's brand value.

In linguistics, complaining is defined as an individual's statement of dissatisfaction with an enterprise, commodity, or event [1]. Complaints have been grouped into various degrees of severity based on the emotional intensity of the complaint, the amount of face-threat that the complainer is willing to undertake, and the complaint's motive [2], [3]. The objective of complaining could be to voice dissatisfaction, seek explanations, or both. Identifying complaints and associated severity levels in natural language is crucial for downstream application developers such as customer service chatbots [4] and commercial organizations to improve their customer support capabilities by identifying and resolving complaints [5].

Various social media and microblogging services, such as Twitter[2], Facebook[3], and Tumblr[4], have skyrocketed in popularity in recent years. The reason being all these forums enable effective communication amongst Web users worldwide. These social media sites, which offer millions of messages and information to be exchanged daily, are multifaceted and invaluable resources for investigating and understanding people's perceptions and evaluating the data generated by participants. Among various social media platforms, Twitter is among the most widely used social media services, with engaging individuals, famous events, and intriguing issues. The social aspect of Twitter, combined with its huge volume, has transformed it into a cost-effective data repository for exploring and evaluating activities, conduct, and user opinions. As a result, the crux of this study is analyzing Twitter-based complaints essentially focused on the Financial domain.

The intended content of every speech act is strongly influenced by a user's emotional condition [6]. Emotion detection is far more sophisticated and fine-grained in terms of research than sentiment classification [7]. Emotion provides a deeper understanding of the customer's mind when combined with sentiment. A statement, for example, could entail only a negative sentiment, whereas emotion could be anger or even disgust. The relationship between

Apoorva Singh and Sriparna Saha are with the Department of Computer Science and Engineering, Indian Institute of Technology, Patna, India (e-mail: apoorva_1921cs19@iitp.ac.in).

Rohan Bhatia, is with Delhi Technological University, Delhi, India (e-mail: rohanbhatia_2k19ep079@dtu.ac.in).

[1]The corpus and code will be made publicly available in the camera-ready version.

[2]http://twitter.com/

[3]https://facebook.com/

[4]https://tumblr.com/

emotion and sentiment urges us to examine the sentiment and emotion of customers while studying complaints.

The following are the major contributions of our proposed work:

- We curate a new corpus *Financial Complaints* (*FINCORP*) for aiding complaint identification and complaint severity classification research.
- We perform good quality annotations for the *FINCORP* dataset for solving four tasks: complaint detection, severity level detection, emotion recognition, and sentiment analysis.
- We present a framework for joint learning of (a) binary complaint classification, (b) complaint severity level classification, (c) emotion recognition, and (d) sentiment analysis. Complaint Identification (CI) and Complaint Severity level classification (CS) are the primary tasks in our multi-task framework. In contrast, Emotion Recognition (ER) and Sentiment Analysis (SA) are considered additional tasks. According to experimental results, the proposed approach surpasses the baselines and state-of-the-art approaches.

The rest of the paper is structured as follows. We present a brief overview of the existing literature in Section 2. In Section 3, we go over the specifics of the proposed dataset. In Section 4, we describe our proposed technique for multi-task experiments. The experimental setting and results analysis are covered in Section 5. Finally, we conclude in Section 6 with a discussion on future work prospects.

## II. Related Work

This section discusses some of the prior works on the complaint identification task and the publicly available annotated complaint corpora.

### A. Complaints in Linguistic Studies

Previous work in linguistics has categorized complaints based on their severity and directness. Complaints can be divided into four fine-grained severity classifications, according to Trosberg [2]: (a) no explicit reproach; (b) disapproval; (c) accusation; and (d) blame. Quite recently, [3] divided complaints into three categories: very direct, moderately direct, and indirect. Clear violations of expectations are referred to as direct complaints (i.e., very direct and moderately direct). On the other hand, indirect complaints do not directly specify or imply a breach of expectations. Furthermore, the contrast between very direct and moderately direct is that the former underlines the obligation of the complaint recipient, while the latter does not.

### B. Multi-task Learning

Previous findings [8], [9], [10] have confirmed the efficiency of multi-task systems by simultaneously learning multiple associated tasks. An individual's emotional state and sentiment have a decisive impact on the intended content [6]. Along with sentiment, emotion offers a deeper understanding of the consumer's mindset [11], [8]. The correlation between emotion and sentiment motivates us to consider complainants' sentiment and emotion while analyzing complaints. Moreover, sentiment and emotion play significant roles in human interactions and thereby contribute toward building efficient and versatile artificial intelligence-based systems [7].

Table I
Statistics of the existing and proposed dataset for complaint identification

| Dataset | Size | Description |
|---|---|---|
| Complaints [12] | 3449 instances out of which 1235 are complaints | Sample instances collected from Twitter grouped in nine generic domains. |
| Product Review [13] | 3449 instances out of which 566 are complaints | Product Reviews collected from Amazon and Youtube websites. |
| FINCORP | 6282 instances out of which 3149 are complaints | Sample instances collected from Twitter based on financial product type & concerns. |

### C. Complaint Identification

Complaint identification on social media is a time-consuming and challenging process that needs identifying complaints from disparate and noisy text samples with character constraints, the use of unpredictable abbreviations, and sarcastic expressions. Text-based complaints have earlier been investigated via semi-supervised strategies, complex feature engineering-based machine learning methods, and deep learning models [14], [12], [15]. In the work [12], authors proposed a logistic regression model with hand-crafted features for detecting complaints. The authors of the work [16] used an ensemble of neural language models enhanced by the usage of transformer networks on a Twitter-based Complaint dataset. In linguistics, complaints were earlier classified based on responsible branch, the level of urgency and the possibility of receiving a response [17], [18], [19].

### D. Complaint Datasets

The *Complaints* dataset[5] was published in the work [12]. The authors selected 93 customer service handles from Twitter and sampled original tweets addressed to these accounts. Only tweets that received a response from the customer support accounts were considered. A total of 1971 tweets were collected using this method. Additionally, 1478 tweets were sampled to guarantee a diverse and representative range of tweets. Overall, the

[5]https://github.com/danielpreotiuc/complaints-social-media

Table II
ANNOTATION EXAMPLES FOR EACH COMPLAINT SEVERITY LEVELS

| Severity level | Annotation Example |
|---|---|
| Accusation | I ordered a mattress from your site. Item is missing now. You are doing fraud. |
| Blame | One payment was made with my consent the another payment seems a fraud transaction gone to an unkown account. |
| Disapproval | Can't even begin to tell how disappointed I am with crypto market. Lost so much money. |
| No explicit reproach | I have a problem and need help, support site kept putting me in a loop. Help! |

Table III
ANNOTATION EXAMPLES FOR EACH EMOTION CLASS

| Emotion | Annotation Example |
|---|---|
| Anger | Most of the time your ATM machines does not work. Over that you will charge ATM usage charges also! What the hell is your problem? |
| Disgust | What kind of service is this? Absolutely pathetic and disgusting packaging. I want my money back. |
| Fear | I am scared due to this pandemic and lockdown will worsen our world economy. |
| Happiness | I love your mobile app, makes it so easy to do all my payments from one place. Awesome! |
| Sadness | I am not able to use netbanking and mobile since last evening. When will downtime end? |
| Surprise | Best investment I've ever made. The knowledge and understanding there is amazing!! |
| Other | Can we apply debit card in our nearest branch or only in home branch and what documents are required for this purpose? |

dataset consists of 2214 non-complaints and 1235 complaints in English. The dataset was grouped into nine domains (i.e., Apparel, Cars, Electronics, Food, Retail, Service, Software, Transport, and Other). Recently in the work [15], authors extended the *Complaints* dataset with the complaint severity classes and analyzed the effect of severity levels on CI.

The *Product Review* corpus[6] [13] is a Hindi language-based product review corpus. This corpus is a collection of product reviews posted on the retail website Amazon[7] and the video-sharing platform, YouTube[8] comment section. The corpus comprises 3711 instances, with 3145 being non-complaints and 566 being complaints. The dataset was further grouped into five domains (i.e., Book, Headphones, Phone, Watch, and Miscellaneous).

[6]https://github.com/MrRaghav/Complaints-mining-from-Hindi-product-reviews/blob/master/amazon-youtube-hindi-complaints-data.xlsx
[7]https://www.amazon.in/
[8]https://www.YouTube.com/

Figure 1. One hundred most frequent words used for expressing complaints by Twitter users.

In this work, we introduce the *Financial Complaints* dataset (*FINCORP*), which unlike the *Complaints* and *Product Review* datasets, is a balanced dataset with nearly equal percentage of complaints (50.13%) and non-complaints (49.87%) samples. Unlike the other two datasets, which contain generic complaints from various domains, our dataset focuses on grievances arising between financial institutions and consumers on the social media network Twitter. Another unique feature of our dataset is that, in addition to the complaint identification axis, we have expanded the dataset on three other closely related axes (severity level classification, sentiment classification, and emotion recognition) to better understand the nuances of the complaint identification task. To the best of our knowledge, this is the first gold standard complaint dataset specifically dealing with Financial complaints.
We create a word cloud to graphically depict the frequency of words that appear more frequently in complaints expressed on Twitter. The more frequently a term occurred in the complaint tweets, the larger the term is in the image. Figure 1 depicts the top 100 terms used by 6282 Twitter users to convey their dissatisfaction. If we observe Figure 1 closely, we will see that Twitter users complain more frequently using key phrases like response, refund, and please, alongside other words like RBI, fraud, and help. Table I shows the details of various existing and proposed CI-related datasets.

## III. Corpus Description

The existing complaint datasets [12], [13] constitute only binary-class complaints. To encourage complaint identification and severity prediction research in the financial domain, we curate a *Financial Complaints* dataset (*FINCORP*) labeled with complaint labels, severity classes,

emotion, and sentiment classes. We chose Twitter as the data source as it is a forum that allows for a lot of self-expression, and users can communicate directly with each other or with corporate support handles. Tweets are freely available and are a popular choice for data analysis in various other similar tasks such as sentiment and emotion analysis ([9], [8]), etc. Here, we discuss the details of the FINCORP dataset.

### A. Data Collection

Complaints account for a small percentage of all Twitter posts. Many companies employ Twitter to offer customer support and deal with customer complaints. To collect finance-related complaints and opinions from tweets, we use Tweepy[9] library to fetch tweets containing specific keywords and phrases using its *cursor class*. Based on an analysis of previous complaint datasets, we used a combination of keywords such as credit card/ debit card/ net banking/ UPI in combination with transaction failure/ fraud/ no response/ customer support/ complaint, etc., to fetch finance-related complaint tweets. Twitter Developers'[10] account is required to use this library, in which API tokens are provided by Twitter necessary to make the tweepy API object. In order to fetch tweets, queries in the form of strings have to be passed to the cursor object, and tweets containing those strings can be fetched and saved into data structures. We collected 6945 tweets to be considered for annotation using this technique.

Furthermore, we manually sorted the tweets into ten domains (Credit card, Customer service, Debit card, Economy, Financial policies, Investments, Loans, Retail banking, Salary, and Transaction failure) based on the financial product type and concerns expressed. We have done this to enable complaint analysis on a more granular level. In related studies on complaint analysis, tweets and product reviews were similarly sorted based on domains ([12], [13]).

We anonymize all usernames and URLs in the tweet and substitute them using placeholder tokens for preprocessing. Using langid.py ([20]), tweets were filtered for English, and retweets were discarded. The Unicode emojis in the tweets were converted to emoji short text with the help of a Python module called Emoji. The Python library Textblob was used to correct the spellings of many improperly spelled words. We use the NLTK library to remove stopwords and stem the corpus using Snowball Stemmer.

### B. Data Annotation

We selected three graduate students[11] who are articulate in English to annotate the tweets with suitable complaint/non-complaint classes, severity levels, as well as emotion and sentiment tags. Before the annotation process began, the guidelines for annotation were provided.

[9]https://www.tweepy.org/
[10]https://developer.twitter.com/
[11]Annotators were recruited from the author's institution.

Table IV
DOMAIN-WISE DATA DISTRIBUTION IN THE *FINCORP* DATASET.

| Domain | Non-Complaints | Complaints |
|---|---|---|
| Credit card | 402 | 238 |
| Customer service | 12 | 697 |
| Debit card | 668 | 191 |
| Economy | 600 | 4 |
| Financial policies | 238 | 29 |
| Investments | 906 | 70 |
| Loans | 102 | 285 |
| Retail banking | 45 | 303 |
| Salary | 28 | 140 |
| Transaction failure | 132 | 1256 |
| Total | 3133 | 3149 |

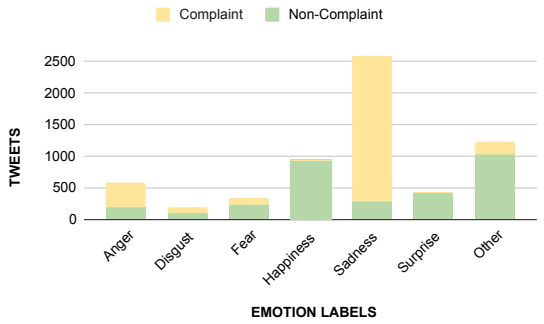



Figure 2. Distribution of emotion labels across *FINCORP* dataset.

Some sample cases outside the corpus were provided to train the annotators for the annotation work.
Annotators were selected using the following criteria:

- They are at least 22 years old and fluent in English (reading, writing, and speaking).
- Prepared to read, interpret, and annotate twitter instances according to the provided guidelines.

A few examples were provided before annotation for complaint severity level, and emotion task annotation as shown in Table II, III, respectively. Initially, we construct a binary annotation task to determine whether or not a tweet involves a complaint. Tweets are most often brief and convey a single idea. As a result, if a tweet includes a single complaint speech act, we consider the overall tweet to be a complaint. We utilize the complaint definition from earlier linguistic research [21] for complaint annotation: *A complaint presents a state of affairs that breaches the writer's clear expectation.*

We utilize the complaint severity levels ( *no explicit reproach, disapproval, accusation, and blame*) defined in [2] as it is considered a benchmark in linguistic studies[12]. According to [2], classes are disconnected. In the 'No explicit reproach' class, the reason for the complaint is not specified, and the complaint is not harsh. The 'Disapproval' class conveys clear negative emotions, including discontent, anger, contempt, or disapproval. The primary distinction between 'Accusation' and 'Blame' is that the

[12]We assign a 'No Complaint Severity' class for non-complaint tweets.

Table V
SAMPLE INSTANCES FROM THE *FINCORP* DATASET. COMP: COMPLAINT, NON-COMP: NON-COMPLAINT

| Domain | Tweet | Label | Severity | Emotion | Sentiment |
|---|---|---|---|---|---|
| Loan | Why is your Panvel branch team not working on disbursing my loan which is approved since Friday? | Comp | Accusation | Sadness | Negative |
| Transaction Failure | I am not able to do online transaction from my credit card and customer care number also not responding. | Comp | Blame | Sadness | Negative |
| Credit Card | Rejecting Lost Credit Card Replacement without any valid reason! Using your card service for 10years. Highly disappointed with the way you treated your existing loyal customer. | Comp | Disapproval | Anger | Negative |
| Debit Card | I've found a Debit Card on my walk. I'd like to report it found. What do I do next? | Comp | No explicit reproach | Other | Neutral |
| Credit Card | Kpop concert tickets are great for my credit card points I got like $100 off a flight thanks. | Non-Comp | Non-Comp | Happiness | Positive |

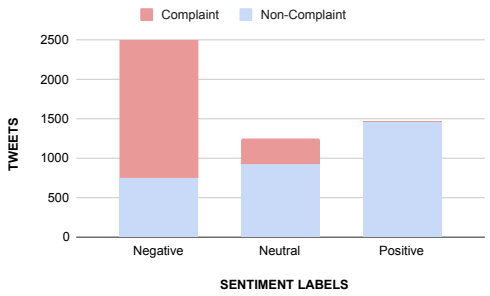



Figure 3. Distribution of sentiment polarities across *FINCORP* dataset

Table VI
DISTRIBUTION OF COMPLAINT INSTANCES ACROSS THE INDIVIDUAL SEVERITY LEVELS.

| Severity class | Amount |
|---|---|
| Accusation | 121 |
| Blame | 876 |
| Disapproval | 604 |
| No explicit reproach | 1548 |
| Total | 3149 |

complainant assumes the complainee is responsible for the violation in the latter. For the emotion annotation of the *FINCORP* dataset, we consider Ekman's [22] six basic emotions (*anger, disgust, fear, happiness, sadness, and surprise*). Other than these six basic emotions, we add a category 'Other' to represent such tweets that do not fall under the six emotion classes' scope. For the sentiment annotation, we consider three sentiment classes (*negative, neutral, positive*).
The final complaint, severity level, emotion, and sentiment labels are selected by utilizing the majority voting technique. The final annotated dataset excluded tweets with no common severity level, emotion, or sentiment label.

Finding agreement among the different annotators is vital in any annotation task requiring two or more annotators to obtain a credible annotated dataset. One such metric is Fleiss' kappa [23] score. We computed the Fleiss' kappa score on the annotated dataset to assess inter-rater agreement among the three annotators. We attained agreement scores of 0.83, 0.69, 0.68, and 0.82 on the complaint, severity level, emotion, and sentiment tasks, respectively, indicating that the annotations are of good quality.

### C. Statistics

The *FINCORP* dataset comprises 6282 tweets with the corresponding complaint, severity, emotion, and sentiment labels. It consists of 3133 tweets in the non-complaint category (49.87%) and 3149 tweets in the complaint category (50.13%). Distributions of sentiment and emotion labels across the dataset are shown in Figure 2 and 3, respectively. Table IV presents the domain-wise data distribution statistics. Table V shows sample instances along with the corresponding complaint, severity, emotion, and sentiment labels. As can be observed from Figure 3, most of the complaint-type tweets have a negative polarity, with a few being neutral. In the case of emotion labels, most complaint tweets show sadness, and the next significant emotion is anger. The distribution of tweets across complaint severity classes is shown in Table VI.

## IV. METHODOLOGY

In this section, we outline our problem and discuss the details of the baselines used for the evaluation of the dataset.

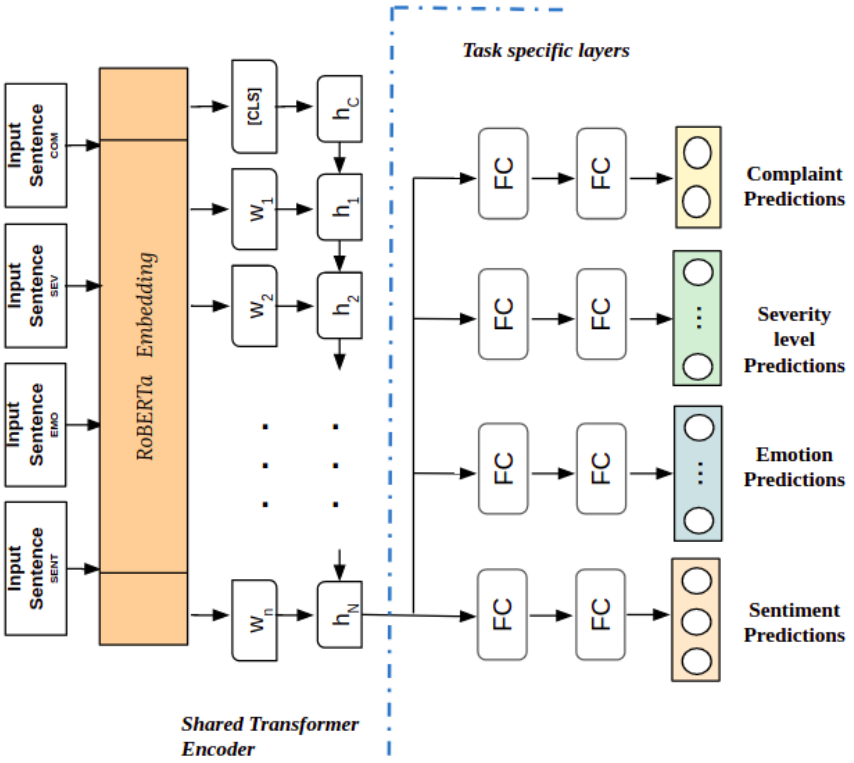



Figure 4. The architectural diagram of Multi-task Learning with RoBERTa (MTL-RoBERTa) framework. FC: Fully connected layers

### A. Problem Definition:

We intend to learn four closely related tasks at the same time, including binary complaint identification (main-task), complaint severity prediction (main-task), emotion recognition, and sentiment analysis (auxiliary tasks). Let $(t_{\mathrm{i}}, e_{\mathrm{i}}, s_{\mathrm{i}}, m_{\mathrm{i}}, c_{\mathrm{i}})_{i=1}^{T}$ be a set of T instances where $e_{\mathrm{i}}$, $s_{\mathrm{i}}$, $m_{\mathrm{i}}$ and $c_{\mathrm{i}}$ represent the matching emotion, sentiment, severity and complaint labels for ${t_{\mathrm{i}}}^{\mathrm{th}}$ tweet, respectively. Here, $t_{\mathrm{i}} \epsilon T$, $e_{\mathrm{i}} \epsilon E$ (emotion classes), $s_{\mathrm{i}} \epsilon S$ (sentiment classes), $m_{\mathrm{i}} \epsilon M$ (severity classes) and $c_{\mathrm{i}} \epsilon C$ (binary complaint classes).

Our multi-task learning framework's objective is to maximize the function $f$ (Eqn 1) that draws a new instance $t_{\mathrm{i}}$ to its fitting emotion label $e_{\mathrm{i}}$, sentiment label $s_{\mathrm{i}}$, severity label $m_{\mathrm{i}}$ and complaint label $c_{\mathrm{i}}$, simultaneously.

$$argmax(\Pi_{k=0}^{R} P(m_i, s_i, e_i, c_i | t_i; \theta)), \tag{1}$$

where $t_i$ is the input tweet whose complaint label $(c_i)$, severity label $(m_i)$, emotion label $(e_i)$, and sentiment label $(s_i)$ and are to be predicted. $\theta$ denotes the model's parameters we aim to optimize.

### B. Multi-task learning with RoBERTa (MTL-RoBERTa):

Based on the work in [24], we employ multi-task learning with hard parameter sharing (MTL) to boost the binary complaint classification utilizing severity levels, emotion, and sentiment tasks. MTL with hard parameter sharing comprises a single encoder that is shared and modified by all the tasks and succeeded by task-specific branches. MTL allows two or more tasks to be learned simultaneously by sharing model knowledge and parameters. This boosts the binary complaint classification utilizing severity levels, emotion, and sentiment task features. The overall MTL-RoBERTa framework is shown in Figure 4.

Bidirectional Encoder Representations from Transformers (BERT) is a multi-layer bidirectional transformer encoder [25] based on the original work shown in [26]. It is trained on masked language modeling, which involves arbitrarily masking certain tokens from the input to predict them only based on context. RoBERTa [27] is an extension of BERT that has exhibited enhanced performance on social media analysis tasks [28], [29], by training on more data with varying hyperparameters.

The proposed MTL-RoBERTa model utilizes a single encoder layer that is shared and modified by all four tasks, with independent feed-forward networks for each task. We use the pre-trained RoBERTa model to embed the words of each tweet instance. The embedding representations are then fed to the shared transformer encoder (the 'roberta-base' model supports an inbuilt encoder as an attribute) to generate the instance's latent representation. This representation is then passed in parallel to four fully connected layers (for each of the four tasks), each of which has two hidden layers, to predict the output for each task separately. Each task-specific loss function is assigned a weight by the final loss function to be optimized. Finally, we add four separate output layers with softmax activation functions for binary and multi-class prediction.

*Calculation of Loss:* The categorical-cross entropy (CE) losses are calculated for the complaint (C), severity (M), emotion (E), and sentiment (S) tasks. Our proposed model's (MTL-RoBERTa) integrated loss function (L) is realized as follows:

$$L = a * L^{\mathrm{C}}{}_{\mathrm{CE}} + b * L^{\mathrm{M}}{}_{\mathrm{CE}} + c * L^{\mathrm{E}}{}_{\mathrm{CE}} + d * L^{\mathrm{S}}{}_{\mathrm{CE}}. \quad (2)$$

To calculate the overall loss, we calculated the weighted sum of the losses from the four tasks. The loss weights indicating the per-task loss-share to the overall loss are determined by the constants a, b, c, and d, which range from 0 to 1.

## V. Experiments and Results

This section discusses the result and analysis of various baseline models and our proposed paradigm, as tested on the FINCORP corpus. The dataset and codes will be released publicly via the corresponding repository.

### A. Baselines

Since *FINCORP* is a new dataset, for a comprehensive evaluation, we evaluate the dataset on multiple baselines, discussed as follows:

- **Logistic Regression (LR-BOW):** To measure the performance of the deep learning models, we also employ Logistic Regression, a common machine-learning-based classifier that has repeatedly served as a good baseline in many classification tasks. We employ Logistic Regression with standard Bag-of-Words (BOW) for CI and CS tasks.
- **Single-task systems:** For binary complaint identification and complaint severity prediction, we develop a single-task deep learning model based on RoBERTa ($\mathrm{STL_{CI}}$, $\mathrm{STL_{CS}}$).
- **Multi-task system:** Based on existing works in CI, we develop $\mathrm{Baseline_1}$[10] model as the multi-task baseline. $\mathrm{Baseline_1}$ is a shared private multi-task model for simultaneous classification of complaint, emotion, and sentiment tasks. For training the model, the authors employed three different sources of data. We also build an $\mathrm{MTL_{GloVe}}$ model wherein the textual embeddings are generated using the pre-trained GloVe[13] [30]. The word sequence encoder receives the embedding layer's output and analyses it to attain contextual knowledge from the text. The rest of the framework is the same.
- **Ablation models:** To evaluate the influence of complaint severity, emotion, and sentiment prediction on the complaint task independently and in varying combinations, we developed dual-task ($\mathrm{MTL_{CI+ER}}$, $\mathrm{MTL_{CI+SA}}$, $\mathrm{MTL_{CI+CS}}$) and tri-task variants ($\mathrm{MTL_{CI+ER+SA}}$, $\mathrm{MTL_{CI+ER+CS}}$, $\mathrm{MTL_{CI+SA+CS}}$) of our proposed model.

[13]GloVe:urlhttp://nlp.stanford.edu/data/wordvecs/glove.840B.300d.zip

### B. Experimental Setup

We employ Python-based libraries to implement our proposed framework as well as the baselines, namely TensorFlow[14], Scikit-learn[15] [31]. The pre-trained model (RoBERTa) is from the open-source library[16] huggingface transformer. We ran all of the experiments on a single NVIDIA GeForce RTX 2080 Ti GPU. We evaluated the predictive performance of the proposed model and the baselines using accuracy and macro-F1 score. We have utilized 60% of the *FINCORP* dataset as training data, 15% for validation, and 25% as testing data for all the experiments. A seed value of 40 was used for a fair assessment of the models, allowing them to encounter the same training and testing data. To effectively employ the GPU, we kept the batch size at 8. We use *Adam* optimizer [32] with a learning rate of $lr$ = 1e-3, $lr$ $\epsilon$ {1e-1, 1e-2, 1e-3, 1e-4}. We use the base uncased model for RoBERTa, and fine-tune it with learning rate of $lr$ = 5e-5, $lr$ $\epsilon$ {1e-4, 5e-5, 1e-5, 1e-6}. The maximum sentence length is set at 50 characters, which spans 93% of the tweets in the training set. We utilize *ReLU* [33] activation and a *dropout* [34] of 30%, *dropout* $\epsilon$ {20%, 30%, 50%}. The output layers for complaint, severity, emotion, and sentiment classification tasks use $Softmax$ activation with 2, 4, 6, and 3 neurons, respectively. To train across all channels, categorical cross-entropy is employed as the loss function. *All of these parameter values were chosen following a thorough sensitivity study.*

### C. Results

*1) **Results Discussion:**: Kindly note that the current paper's objective is to boost the efficacy of the CI and CS tasks by utilizing the two secondary tasks (ER and SA). As a result, we present the results and analysis with CI and CS as the primary tasks in all task pairings.*

The classification results from the various experiments are shown in Table VII. It can be observed that combining all associated tasks, such as severity, emotion, and sentiment, improves performance compared to single-task, dual-task, and tri-task variants. This improvement indicates that the proposed model makes effectual usage of interaction between the four tasks. As seen in Table VII, the MTL-RoBERTa model, which includes all four tasks (CI, CS, ER, and SA), beats the single-task complaint and severity variants ($\mathrm{STL_{CI}}$, $\mathrm{STL_{CS}}$). $\mathrm{MTL_{CI+CS}}$ outperforms $\mathrm{MTL_{CI+ER}}$ and $\mathrm{MTL_{CI+SA}}$ in the dual-task variants. This supports our assumption that the complaint severity level information can boost the complaint identification task while simultaneously learning these two tasks. This could also be influenced by the notion that sentiment by itself is usually unable to convey essential details about the complainant's predicament. For instance, negative feelings toward a company or a product

[14]https://www.tensorflow.org/

[15]https://scikit-learn.org/stable/

[16]https://huggingface.co/model and https://tfhub.dev/google/collections/bert/1

Table VII
RESULTS OF THE PROPOSED MODEL AND OTHER BASELINES WITH RESPECT TO MACRO-F1 SCORE AND ACCURACY VALUE FOR CI TASK. BOLD-FACED VALUES INDICATE THE MAXIMUM SCORES ACHIEVED. THE * INDICATES THAT THE RESULTS ARE STATISTICALLY SIGNIFICANT.

| Model | Accuracy (%) | F1-score(%) |
|---|---|---|
| **$SOTA_{CI}$[16]** | 84.26 ± .15 | 83.53 ± .03 |
| **LR-BOW** | 67.52 ± .03 | 64.28 ± .03 |
| **$Baseline_1$[10]** | 82.38 ± 1.02 | 81.84 ± 1.05 |
| **$MTL_{GloVe}$** | 80.45 ± 1.03 | 79.32 ± 1.02 |
| **Single-task baseline** | | |
| **$STL_{CI}$** | 81.26 ± 0.03 | 80.16 ± 0.05 |
| **Dual-task baselines** | | |
| **$MTL_{CI+ER}$** | 81.79 ± 1.07 | 81.54 ± 1.13 |
| **$MTL_{CI+SA}$** | 80.58 ± 1.12 | 79.51 ± 1.06 |
| **$MTL_{CI+CS}$** | 83.87 ± 1.04 | 83.28 ± 1.07 |
| **Tri-task baselines** | | |
| **$MTL_{CI+ER+SA}$** | 82.48 ± 1.17 | 82.14 ± 1.03 |
| **$MTL_{CI+ER+CS}$** | 84.28 ± 1.05 | 83.54 ± 1.02 |
| **$MTL_{CI+SA+CS}$** | 83.39 ± 1.02 | 81.89 ± 1.06 |
| **Quad-task baseline** | | |
| **MTL-RoBERTa** | **86.33*** ± .03 | **85.82*** ± .05 |

Table VIII
RESULTS OF ALL THE PROPOSED MODEL AND OTHER BASELINES WITH RESPECT TO THE MACRO-F1 SCORE AND ACCURACY VALUE FOR CS TASK. BOLD-FACED VALUES INDICATE THE MAXIMUM SCORES ACHIEVED. THE * INDICATES THAT THE RESULTS ARE STATISTICALLY SIGNIFICANT.

| Model | Accuracy (%) | F1-score(%) |
|---|---|---|
| **$SOTA_{CS}$[15]** | 65.67 ± .11 | 64.17 ± .12 |
| **LR-BOW** | 48.52 ± .14 | 46.28 ± .12 |
| **$MTL_{GloVe}$** | 60.49 ± 1.03 | 60.12 ± 0.14 |
| **Single-task baseline** | | |
| **$STL_{CS}$** | 57.89 ± 1.04 | 57.19 ± 1.12 |
| **Dual-task baselines** | | |
| **$MTL_{CS+CI}$** | 61.22 ± 1.14 | 60.08 ± 1.03 |
| **$MTL_{CS+ER}$** | 58.67 ± 1.02 | 56.88 ± 1.04 |
| **$MTL_{CS+SA}$** | 57.88 ± 1.11 | 57.37 ± 1.03 |
| **Tri-task baselines** | | |
| **$MTL_{CS+ER+SA}$** | 62.58 ± 1.07 | 61.74 ± 1.03 |
| **$MTL_{CS+ER+CI}$** | 64.68 ± 1.07 | 63.14 ± 1.13 |
| **$MTL_{CS+SA+CI}$** | 63.69 ± 1.12 | 60.19 ± 1.04 |
| **Quad-task baseline** | | |
| **MTL-RoBERTa** | **67.89*** ± .15 | **66.52*** ± .12 |

might be influenced by numerous emotions such as anger, sorrow, etc. Consequently, distinct or minor variations in emotion cannot be accurately determined and conveyed solely through sentiment.

Specifically, in the case of tri-task variants, $MTL_{CI+ER+CS}$ outperforms other tri-task baselines with the macro-F1 score of 83.54%. One of the reasons for the inferior results of $MTL_{CI+ER+SA}$ and $MTL_{CI+SA+CS}$ variants could be that the sentiment analysis task does not contribute much to the multi-task architecture as compared to severity and emotion tasks.

In Table VIII, we report the results of the MTL-RoBERTa model for the complaint severity task. Overall, the MTL-RoBERTa model yields the best result with the macro-F1 score of 66.52%. Moreover, of all the baselines, $MTL_{CS+ER+CI}$ outperforms other baselines with the macro-F1 score of 63.14%. This is in line with [2], where the author claims that the way people convey their grievances is linked to their emotional states.
*The findings reported here are all statistically significant.*[17] *[35]..*

*2)* ***Comparison with State-of-the-art (SOTA):****:* For single-task CI, we compare the MTL-RoBERTa model to the existing state-of-the-art technique [16]. $SOTA_{CI}$ makes use of a variety of neural language models that are aided by transformer networks. We also compare the results of the MTL-RoBERTa model with those of the existing state-of-the-art approach [15] for single-task CS task. $SOTA_{CS}$ uses a wide range of neural language models with several linguistic features augmented by transformer networks. We re-implement it on the FINCORP dataset and report the results in Table VII and VIII. The MTL-RoBERTa model outperforms the SOTA techniques for complaint identification and severity prediction tasks.

*3)* ***Qualitative Analysis:****:* Table IX presents a few examples of MTL-RoBERTa and single-task (complaint task) model predictions. Even though the 1st and 2nd cases in Table IX communicate dissatisfaction in a less explicit tone, the MTL-RoBERTa model accurately predicts the examples as complaints. Even though the 3rd and 4th examples are indirect and vague in nature, the $MTL_{All}$ model identifies that there is no violation of trust. These examples highlight the importance of detecting complaints based on severity, emotion, and sentiment knowledge.

*4)* ***Error Analysis:****:* We examine some of the plausible reasons for the MTL-RoBERTa model's complaint and severity prediction inaccuracies:

- **Elusive Complaints:** Non-complaints are incorrectly predicted when complaints are communicated in an implicit or cryptic manner. One justification for this could be the complainant's lack of explicit accusation or blame. For example, *'Why am I waiting 60mins just so I can ok a debit card transaction!?'*. The predicted class is non-complaint, whereas the instance's actual class is complaint. Unless the tweet does not mention the particular company or use a critical phrase, the model misinterprets it as non-complaint.
- **Neighbouring Severity Level Errors:** Instances that belong to neighboring severity levels tend to be semantically and compositionally similar, sometimes leading to misclassification. For example, *'You are looting consumer's money. It's very disappointing behavior that they didn't refund back failed transaction money even after weeks and their costumer support show us the way to our bank, who doesn't recognize their transaction number.'*; predicted severity level: 'disapproval'. The correct severity level for the preceding example is 'accusation', but due to the presence of negative terms such as *disappointing* which

[17]For the significance test, we utilized the Student's t-test. When testing the null hypothesis, the results are statistically significant (p-value $< 0.05$).

Table IX
Few example instances along-with associated predicted labels for the best performing multi-task (MTL-RoBERTa) model and its single-task CI equivalent. ($STL_{CI}$).

| Sentence | MTL-RoBERTa | $STL_{CI}$ | Actual Label |
|---|---|---|---|
| This has been a great way to spend an afternoon, trying to "chase" down that $832 deposit-student loan funds-that disappeared, then you mysteriously closed the account | Complaint | Non-Complaint | Complaint |
| I had to report my visa as STOLEN after purchasing from Amazon! Thanks alot. And no proper response from the team. | Complaint | Non-Complaint | Complaint |
| I actually don't prefer AMEX. A great card but I only use cards that have a zero annual fee. I personally think annual fees on a credit card are ridiculous. | Non-Complaint | Complaint | Non-Complaint |
| Great job, dude why there's no crypto payment for VIP membership? | Non-Complaint | Complaint | Non-Complaint |

usually appear in the 'disapproval' severity class, the proposed model misclassifies it.

- **Satirical Instances:** Instances where the underlying tone is positive or neutral, but the instance is of complaint type, the MTL-RoBERTa model inaccurately predicts such instances as non-complaint. For example, *'Yes, I need to pay a transaction that cannot be done online/over the phone because it exceeds the daily limit.'*. For this instance, the correct class is complaint, but the model predicts it as non-complaint. One reason could be the neutral undertone and the usage of less explicit words to signify complaint.
- **Probing Complaints:** We also identified that the model incorrectly labels cases of weak discontent with a inquiring tone as non-complaints. For example, *'What is this restaurant packaging charges and where is this money going when McDonald's on their website is not charging it?'*. The correct class for the given example is complaint. The model classifies it as non-complaint because there is no clear objection, and the sentence is interrogative. To effectively classify such occurrences, the system requires background information and knowledge about the key aspects of the sentence.

## VI. Conclusion

In this work, we have introduced FINCORP, an annotated dataset of complaints originating between financial institutions and consumers on Twitter. We have addressed the task of complaint identification and severity level prediction (primary tasks) jointly using a multi-task framework assisted by emotion and sentiment detection as auxiliary tasks. The paper goes into a detailed description of the dataset and the complete process of creating it. Every instance in the FINCORP dataset has been annotated across four axes: complaint, severity level, emotion, and sentiment classes, resulting in a dataset with a lot of diversity. We evaluated the dataset and reported the findings using several single-task, multi-task and other existing baselines. Given the scarcity of task-specific (complaint identification) data in English and even some low-resource languages, we believe this dataset would add value to social media analytics research and practice.

We plan to expand our work in complaint identification at the sentence level in english as well as code-mixed languages in the future. Furthermore, the annotated corpus can aid research in other emotion and sentiment classifications tasks.

## Acknowledgments

This publication is an outcome of the R&D work undertaken in the project under the Visvesvaraya Ph.D. Scheme of Ministry of Electronics & Information Technology, Government of India, being implemented by Digital India Corporation (Formerly Media Lab Asia).

## References

[1] E. Olshtain and L. Weinbach, "Complaints: A study of speech act behavior among native and nonnative speakers of hebrew. the prag-matic perspective," 1985.

[2] A. Trosborg, *Interlanguage pragmatics: Requests, complaints, and apologies*. Walter de Gruyter, 2011, vol. 7.

[3] L. N. Kakolaki and M. Shahrokhi, "Gender differences in complaint strategies among iranian upper intermediate efl students," *Studies in English Language Teaching4 (1)*, pp. 1–15, 2016.

[4] A. Xu, Z. Liu, Y. Guo, V. Sinha, and R. Akkiraju, "A new chatbot for customer service on social media," in *Proceedings of the 2017 CHI conference on human factors in computing systems*, 2017, pp. 3506–3510.

[5] G. Van Noort and L. M. Willemsen, "Online damage control: The effects of proactive versus reactive webcare interventions in consumer-generated and brand-generated platforms," *Journal of interactive marketing*, vol. 26, no. 3, pp. 131–140, 2012.

[6] M. Lewis, J. M. Haviland-Jones, and L. F. Barrett, *Handbook of emotions*. Guilford Press, 2010.

[7] A. Kumar, A. Ekbal, D. Kawahra, and S. Kurohashi, "Emotion helps sentiment: A multi-task model for sentiment and emotion analysis," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.

[8] T. Saha, A. Upadhyaya, S. Saha, and P. Bhattacharyya, "Towards sentiment and emotion aided multi-modal speech act classification in twitter," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2021, pp. 5727–5737.

[9] M. S. Akhtar, A. Ekbal, and E. Cambria, "How intense are you? predicting intensities of emotions and sentiments using stacked ensemble [application notes]," *IEEE Computational Intelligence Magazine*, vol. 15, no. 1, pp. 64–75, 2020.

[10] A. Singh and S. Saha, "Are you really complaining? a multi-task framework for complaint identification, emotion, and sentiment classification," in *International Conference on Document Analysis and Recognition*. Springer, 2021, pp. 715–731.
[11] N. Majumder, S. Poria, D. Hazarika, R. Mihalcea, A. Gelbukh, and E. Cambria, "Dialoguernn: An attentive rnn for emotion detection in conversations," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 6818–6825.
[12] D. Preotiuc-Pietro, M. Gaman, and N. Aletras, "Automatically identifying complaints in social media," *arXiv preprint arXiv:1906.03890*, 2019.
[13] R. P. Singh, R. Haque, M. Hasanuzzaman, and A. Way, "Identifying complaints from product reviews: A case study on hindi," in *Proceedings of the 17th International Conference on Natural Language Processing (ICON)*, 2020, pp. 108–116.
[14] A. Singh, S. Saha, M. Hasanuzzaman, and A. Jangra, "Identifying complaints based on semi-supervised mincuts," *Expert Systems with Applications*, p. 115668, 2021.
[15] M. Jin and N. Aletras, "Modeling the severity of complaints in social media," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2021, pp. 2264–2274.
[16] ——, "Complaint identification in social media with transformer networks," in *Proceedings of the 28th International Conference on Computational Linguistics*, 2020, pp. 1765–1771.
[17] S. Tjandra, A. A. P. Warsito, and J. P. Sugiono, "Determining citizen complaints to the appropriate government departments using knn algorithm," in *2015 13th International Conference on ICT and Knowledge Engineering (ICT & Knowledge Engineering 2015)*. IEEE, 2015, pp. 1–4.
[18] W. Yang, L. Tan, C. Lu, A. Cui, H. Li, X. Chen, K. Xiong, M. Wang, M. Li, J. Pei *et al.*, "Detecting customer complaint escalation with recurrent neural networks and manually-engineered features," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Industry Papers)*, 2019, pp. 56–63.
[19] D. Law, R. Gruss, and A. S. Abrahams, "Automated defect discovery for dishwasher appliances from online consumer reviews," *Expert Systems with Applications*, vol. 67, pp. 84–94, 2017.
[20] M. Lui and T. Baldwin, "langid. py: An off-the-shelf language identification tool," in *Proceedings of the ACL 2012 system demonstrations*, 2012, pp. 25–30.
[21] A. D. Cohen and E. Olshtain, "The production of speech acts by efl learners," *Tesol Quarterly*, vol. 27, no. 1, pp. 33–56, 1993.
[22] P. Ekman, W. V. Friesen, M. O'sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti *et al.*, "Universals and cultural differences in the judgments of facial expressions of emotion." *Journal of personality and social psychology*, vol. 53, no. 4, p. 712, 1987.
[23] J. L. Fleiss, "Measuring nominal scale agreement among many raters." *Psychological bulletin*, vol. 76, no. 5, p. 378, 1971.
[24] R. Caruana, "Multitask learning," *Machine learning*, vol. 28, no. 1, pp. 41–75, 1997.
[25] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 4171–4186.
[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
[27] Z. Liu, W. Lin, Y. Shi, and J. Zhao, "A robustly optimized bert pre-training approach with post-training," in *China National Conference on Chinese Computational Linguistics*. Springer, 2021, pp. 471–484.
[28] W. Liao, B. Zeng, X. Yin, and P. Wei, "An improved aspect-category sentiment analysis model for text sentiment analysis based on roberta," *Applied Intelligence*, vol. 51, no. 6, pp. 3522–3533, 2021.
[29] Y. Liu, H. Liu, L.-P. Wong, L.-K. Lee, H. Zhang, and T. Hao, "A hybrid neural network rbert-c based on pre-trained roberta and cnn for user intent classification," in *International Conference on Neural Computing for Advanced Applications*. Springer, 2020, pp. 306–319.
[30] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
[31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.
[32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: http://arxiv.org/abs/1412.6980
[33] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 2011, pp. 315–323.
[34] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.
[35] B. L. Welch, "The generalization of 'student's'problem when several different population varlances are involved," *Biometrika*, vol. 34, no. 1-2, pp. 28–35, 1947.

**Apoorva Singh** is a research scholar in the Department of Computer Science and Engineering, IIT Patna. She completed her M.Tech from BIT, Mesra in 2016. She has been pursuing research in Text Mining, Information Retrieval and Applications of Federated Learning.

**Rohan Bhatia** is an undergraduate at Delhi Technical University, Delhi. He is pursuing his B.Tech in Engineering Physics with a minor in Computer Science. His interest lies in Data Science and Quantum Computation.

**Dr. Sriparna Saha** is currently an Associate Professor in the Department of Computer Science and Engineering, Indian Institute of Technology Patna, India. She is the author of a book published by Springer-Verlag. She has authored or coauthored more than 290 papers. Her current research interests include deep learning, natural language processing, machine learning, information extraction, text mining, bioinformatics and multiobjective optimization. She is involved with different sponsored research projects in the broad areas of Artificial Intelligence and Machine Learning technologies, funded by different Govt. and private agencies.

Uncorrected Proof

Water Supply

© 2022 The Authors

Water Supply Vol 00 No 0, 1 doi: 10.2166/ws.2022.255

# Computation of energy dissipation across the type-A piano key weir by using gene expression programming technique

Deepak Singh and Munendra Kumar *

Department of Civil Engineering, Delhi Technological University, Delhi
*Corresponding author. E-mail: munendrakumar@dtu.ac.in

MK, 0000-0003-3922-424X

## ABSTRACT

Energy dissipation across the weir and dam structures is a vital economic and technical solution for designing the downstream morphology of any hydraulic system. Accurately estimating the energy over any hydraulic system using traditional empirical formulas is tedious and challenging. Consequently, employing new and precise techniques still in high demand is crucial. In this study, the authors developed an empirical model for estimating the residual energy downstream of the type-A PKW using gene-expression programming (GEP) by considering six non-dimensional parameters: headwater ratio, magnification ratio, inlet to outlet width ratio, inlet to outlet key bottom slopes, inlet to outlet overhang portions and the number of cycles. The performance of the proposed models has been compared to empirical equations using the statistical factors coefficient of determination ($R^2$), concordance coefficient (CC), and root mean square error (RMSE). The computed relative residual energy values using the proposed models are within $\pm$10% of the observed ones. The proposed GEP model predicted the relative residual energy satisfactorily, with coefficients of determination ($R^2 = 0.978$ for training, 0.980 for testing) and root mean square errors (RMSE) of 0.032, 0.029 for the training and testing datasets, respectively.

**Key words:** energy dissipation, gene expression programming, low head efficient structure, residual energy

## HIGHLIGHTS

- This study presents an efficient method for predicting residual energy downstream the type-A PKW that has yet to be published.
- The experimental investigation was carried out to assess the energy dissipation across the type-A PKW.
- GEP technique is used for estimating relative residual energy downstream of the PKW.
- The performance of the proposed models has been compared to empirical equations using statistical factors.

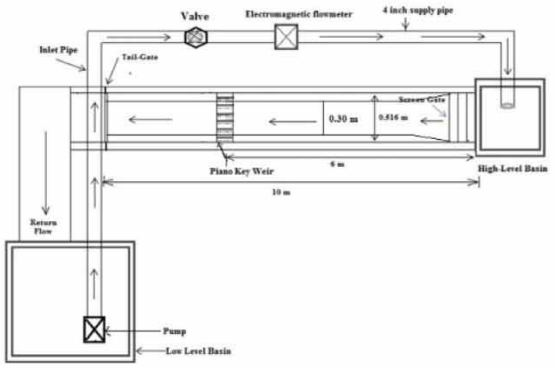
This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (http://creativecommons.org/licenses/by/4.0/).

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2022.255/1073735/ws2022255.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

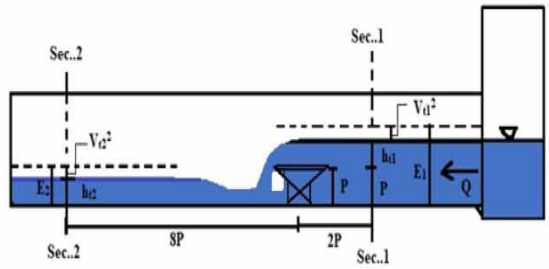
Uncorrected Proof



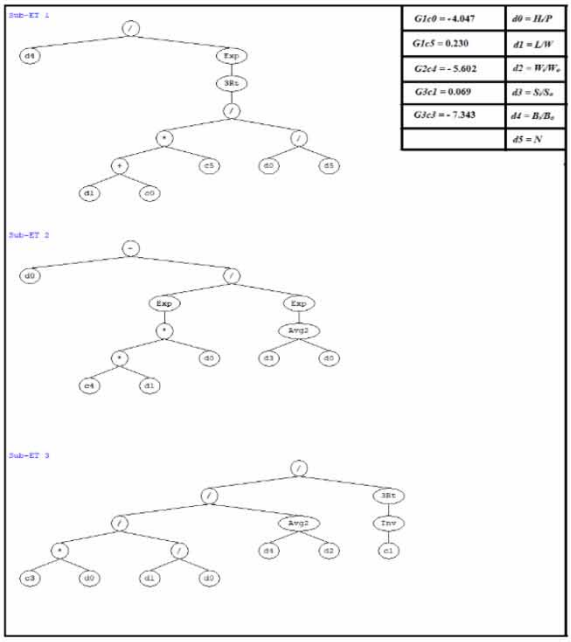
## GRAPHICAL ABSTRACT



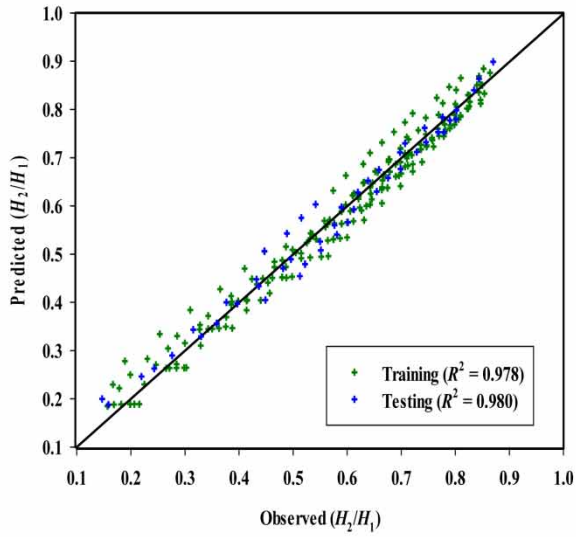

Schematic plan view of the experimental setup.




Schematic plot for specific energy measurement.




Expression Tree (ET) for GEP formulation




Compares the observed and projected relative residual energy ratios for different data phases ( training and testing phases).

## NOTATIONS

| | |
|---|---|
| $B$ | Length of side weir $(B_i + B_b + B_o)$ |
| $B_b$ | Base length |
| $B_i$ | Length of overhang portions at the inlet side |
| $B_o$ | Length of overhang portions at the outlet side |
| $CC$ | Coefficient of correlation |
| $H$ | Specific energy at section' i' |
| $H_r$ | Relative residual energy |
| $H_L$ | Relative energy dissipation |
| $g$ | Acceleration of gravity |
| $H_t$ | Total energy head |
| $y_t$ | Piezometric head |

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2022.255/1073735/ws2022255.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof



| | |
|---|---|
| $i$ | Represents the section |
| $L$ | Total developed crest length |
| $ME$ | Mean Percentage Error |
| $MAE$ | Mean Absolute Error |
| $MAPE$ | Mean absolute percentage error |
| $N$ | Number of cycles |
| $P$ | Weir height |
| $Q$ | Discharge over the PKW |
| $R^2$ | Coefficient of determination |
| $RSME$ | Root mean square error |
| $S_i$ | Inlet key slope |
| $S_o$ | Outlet key slope |
| $V_t$ | Mean flow velocity |
| $V_t^2/2\,g$ | Approach velocity head |
| $W$ | Channel width/Width of PKW |
| $W_i$ | Inlet key width |
| $W_o$ | Outlet key width |

## 1. INTRODUCTION

Efficient computation of the energy dissipation over the hydraulic structures helps in designing the downstream dissipative systems and also helps in reducing downstream hazards during the flood season. Flooding is the most common natural disaster globally, and it has been increasing at an alarming rate over the last two decades. Extreme flooding is expected to become more common due to climate change (Chanson 2021). Flood control generally ensures that floods pass through and are released without causing damage to structural frames or their surroundings (Chanson 1994). The proper estimation of energy dissipation over the hydraulic structures can minimize the downstream expenses. Protection is an essential aspect of dam construction and exploration because an accident can often have significant consequences, to varying degrees, depending on the amount of water contained (Pinto 2017). During a flood or extreme hydrologic conditions, the overtopping conditions may occur due to undersized spillways, culminating in dam overtopping. Therefore, spillways or weir structures must be designed to efficiently spill large amounts of water while maintaining high structural performance (Garg 2010). Consequently, increasing these structures' flow release capability is critical for improving their protection.

Many researchers have shown their interest in the investigation of hydraulic jumps, including Rajaratnam (1990) and Hager *et al.* (1990), who expanded their research to include a jump with a control sill (Hager and Li 1992). Novak (2010) examined the hydraulics of the jump and its implications for the submerged jump stilling basin and proposed that the morphology of the river bed usually determines the shape of the stilling basin. As a result, it is critical to understand the excavation required for its construction and operational purposes. It would be necessary to understand better the energy dissipation phenomenon downstream of the hydraulic structures to support the preceding statement. One of the most common solutions for increasing the spillway's discharging capacity is the installation of a labyrinth-style weir. Indeed, this form of the weir will increase discharge while maintaining the same length as a traditional linear weir (Anderson & Tullis 2012; Leite Ribeiro *et al.* 2012). Energy dissipation at the base of the PKW is an important phenomenon to consider when designing dissipative structures; special care should be taken to avoid undesirable effects such as scouring and cavitation (Silvestri *et al.* 2013a, 2013b).

Ho Ta Khanh *et al.* (2011) presented the first experimental study with a stepped spillway or the survey over the stilling basins has been examined by (Troung Chi *et al.* 2006; Pfister *et al.* 2017). Moreover, the general project-specific research and special studies on the energy dissipation of labyrinth and PKWs were investigated by (Leite Ribeiro *et al.* 2007; Bieri *et al.* 2011; Ho Ta Khanh *et al.* 2011; Leite Ribeiro *et al.* 2011; Erpicum *et al.* 2013). In addition, Silvestri *et al.* (2013c) explored the energy dissipation over-stepped chutes with a PKW crest and found that the low residual energy at the spillway toe increases with discharge and spillway length. The effects of the slope of the PKW on energy dissipation were examined by Al-Shukur & Al-Khafaji (2018). They concluded that the amount of energy dissipated decreases as the slope decreases. Recently, (Eslinger & Crookston 2020) have conducted an experimental study to clarify the energy dissipation analysis at the base of type-A PKWs. In addition, Singh & Kumar (2022a) and (Singh & Kumar 2022b) presented the experimental investigation and the computational technique based on Gene Expression Programming (GEP) to estimate the residual energy at the base of type-B PKW, respectively. They observed that the PKW's energy dissipation is not linear and more at a low head. Zounemat-Kermani & Mahdavi-Meymand (2019) used artificial intelligence data-driven models (ANFIS & MLPNN) embedded with several meta-heuristic algorithms (GA, PSO, FA & MFO) to simulate the passing flow over PKW, and

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2022.255/1073735/ws2022255.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof



compared the results. General results indicated that the ANFISs and MLPNNs could simulate the discharge coefficient of the PKW more accurately than empirical relations.

Many researchers have applied the GEP in the various field of hydraulics to predict the accurate estimates of the different hydraulics characteristics (Azamathulla *et al.* 2013; Karbasi & Azamathulla 2016); Azamathulla *et al.* (2018). Karbasi & Azamathulla (2016) used Gene Expression Programming (GEP) to predict the characteristics of a hydraulic jump over a rough bed. They compared it with the standard artificial intelligence (ANN and SVR) techniques. They found that the artificial intelligence techniques indicated that the performance of these models is slightly better than the GEP model, but the application of the GEP model due to derivation of explicit equations is easier for practical purposes. Further, Azamathulla *et al.* (2018) used the GEP to predict the atmospheric temperature in Tabuk, Saudi Arabia. Abhash & Pandey (2020) and Singh & Kumar (2021) summarised the geometrical and hydraulic evaluation of PKWs over the last decade. Yazdi *et al.* (2021) investigated the effects of weir geometry on scouring development downstream of the PKW. According to them, the geometry of the weir and the discharge rate impact the scour characteristics. Zhao *et al.* (2020) discovered that the saltation height and length increase if the particle shape is not spherical. Moreover, Kumar *et al.* (2021) investigated sediment passage over type-A PKWs and observed that sediment passage at a lesser rate as it goes through the intake key and speeds up at the key entrance. Plunging and impinging jets originating from the inlet and outflow keys were credited with forming the ridge and dip (Kumar & Ahmad 2022).

Studies of local scour at the toe of PKWs placed in canals and rivers are also related to the energy dissipation of PKWs. Jüstrich *et al.* (2016) investigated the formation of scour holes and ridges caused by PKWs without scouring protection, concluding that the overall process is jet-induced scour. Pfister *et al.* (2017) conducted an experimental study to investigate the toe dug at the PKW and noticed that if the foundation on rock is not possible, then toe-scour occurring during flood discharges is relevant to weir stability. Shivashankar *et al.* (2022) describe the different methodologies for estimating the velocity phenomenon. This study proposed a hybrid generalized reduced gradient-genetic algorithm (hybrid GRG-GA) to assess the fall velocity. The hydraulic performance of the trapezoidal labyrinth-shaped stepped spillways was investigated by Ghaderi *et al.* (2020). The scour depth and volume of sediment removed at the toe of PK weirs are determined primarily by sediment properties, discharge, residual energy, and tailwater depth. In addition, several experimental studies and machine learning algorithms/techniques were used in the prediction of the scour depth around submerged weirs, spur dikes, and a circular pier (Rashki Ghaleh Nou *et al.* 2019; Pandey *et al.* 2020a, 2020b; Birbal *et al.* 2021; Ghasempour *et al.* 2021; Pandey & Md Azamathulla 2021; Emadi *et al.* 2022; Singh *et al.* 2022). Pandey *et al.* (2021) suggested some critical points to Mohammad Najafzadeh and Ali Reza Kargar for their article on 'gene-expression programming, evolutionary polynomial regression, and model tree to evaluate local scour depth at culvert outlets.' Despite numerous significant experimental and computational investigations on energy dissipation across the labyrinth and PKWs, designers lack the knowledge to predict using traditional empirical models. As a result, new and precise approaches are still in high demand.

This article aims to create new equations based on gene-expression programming to determine the relative residual energy downstream of the type-A PKW, which will aid hydraulic structure designers and engineers. During training, the measured flow and geometrical parameters are entered as input parameters into GEP, and the target parameter is the residual energy downstream of the PKW. The proposed GEP-based approach is compared to the conventional ones.

## 2. MATERIALS AND METHOD

### 2.1. Experimental setup

The tests were carried out under free-flow circumstances at the 'Fluid Mechanics and Hydraulics Research Laboratory' at Delhi Technological University, Delhi, India, using a tilting steel flume of dimensions 10 m long and 0.516 m wide, 0.6 m deep. The width of the test portion was reduced to 0.30 m during the experiment to increase the water head over the weir to eliminate scaling effects. A 20 Horse Power (HP) pump attached to a series of 4-inch supply pipes provided water to the channel, which was controlled by a flow regulating valve that predicted discharges of up to 50 L/s and calibrated using the flume supply line and an orifice meter (0.25 percent uncertainty). A 4–20 mA electromagnetic flowmeters (uncertainty $\pm 0.2\%$) were used to measure the discharge (see Figure 1). In order to regulate the water elevations in various sections, the flume was equipped with a 4–20 mA ultrasonic level sensor ($\pm$ 0.2 percent uncertainty) instrumentation carriage and a pointer gauge of least count $\pm 0.1$ mm. The PKW's head was measured at a distance of $2P$ upstream and $8P$ downstream from the lateral centreline of the weirs after the water surface had been permitted steady-state for at least 3–5 minutes, as shown in Figure 1(b). The mean approach flow velocity $V_t$ was calculated as the average velocity measured at the same cross-section at

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2022.255/1073735/ws2022255.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof



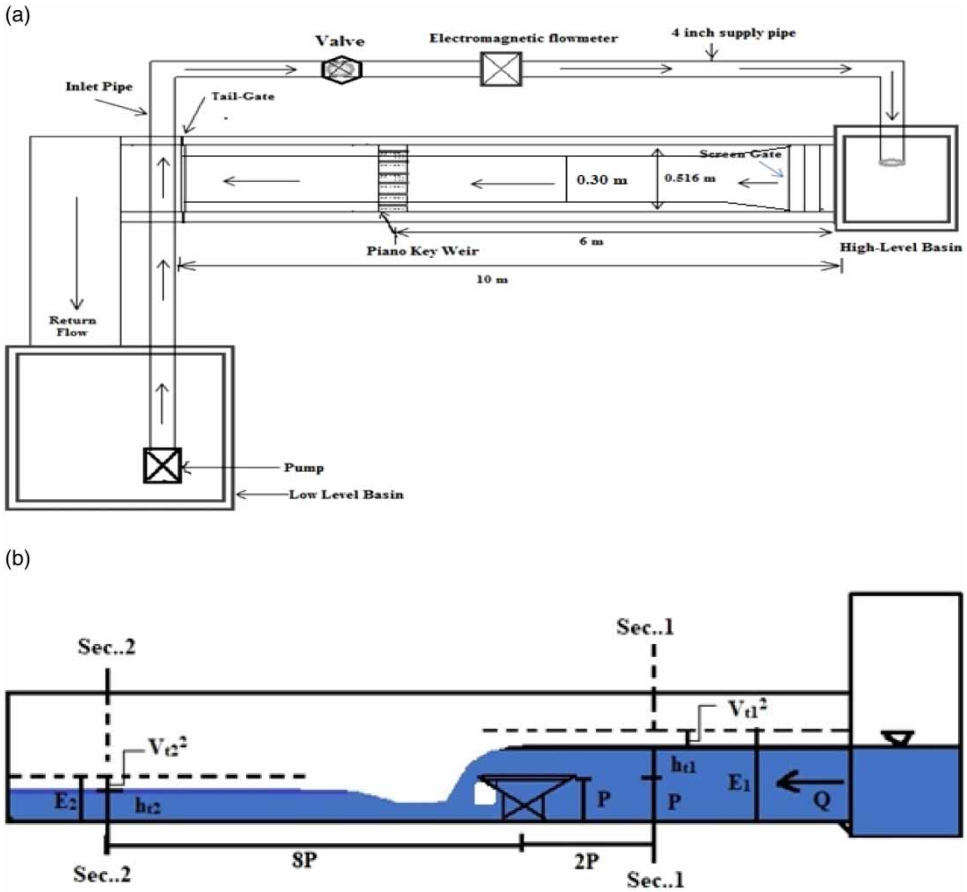



**Figure 1** | (a) Schematic plan view of the experimental setup. (b) Schematic plot for specific energy measurement.

0.25, 0.5, and 0.75$W$ across the width $W$ of the flume each for one-minute records using a Sontek ADV (Acoustic Doppler Velocimeter). The ADV has proven highly dependable in recording turbulence characteristics in an open channel stream. The results of the velocity analysis revealed agreement between these average cross-sectional velocities and mean approach velocities (by ADV), resulting in a difference in $H_t$ of less than 5% for the $Q$ ranges. The details of the constructed models are presented in Table 1.

This research aims to develop a prediction equation for the relative residual energy downstream of a type-A PKW in rectangular horizontal channels under free-flow conditions. To this end, twelve different type-A PKW models were tested and assessed. The models' configurations are as follows: magnification ratio $L/W = 5$–$6$, where $L$ is the total developed crest length, $W$ is the width of the weir or channel, the height of the weir $P$ varies from 0.15 to 0.20 m, relative width ratio $W_i/W_o$ ranged between ($1.0 \leq W_i/W_o \leq 1.5$) where, $W_i$ and $W_o$ are the inlet and outlet key widths, respectively, and overhang portions ($B_i = B_o$) are alike for all models, where $B_i$ and $B_o$ are the inlet-outlet overhang portions, respectively (see Table 1). The discharges range ($0.005\ \text{m}^3/\text{s} \leq Q \leq 0.05\ \text{m}^3/\text{s}$), head to weir ratio ($0.085 \leq H_t/P \leq 0.85$) for 0.20 m tall models, $0.11 \leq H_t/P \leq 0.88$ for 0.15 m tall models, and the total head over crest ($0.0168\ \text{m} \leq H_t \leq 0.167\ \text{m}$), other dimensionless ratios ($0.43 \leq (B_i/P = B_o/P) \leq 0.95$). A total of 240 tests have been conducted over the twelve different types of PKW, of which 20 tests comprise each model.

## 2.2 Methodology

In this study, water head over the weir was measured under free-flow conditions, and the upstream and downstream energy across the PKW was calculated as follows:

$$H_i = P + y_i + \frac{V_i^2}{2g} \tag{1}$$

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2022.255/1073735/ws2022255.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof



**Table 1** | Range of data collected in the present study [type-A PK weirs]

| Model No. | Range of Q ($m^3/s$) | Range of $H_t$ (m) | $W_i/W_o$ | $P$ (m) | $\frac{L}{W}$ | $B$ (m) | $B_i$ (m) | $B_o$ (m) | Range of $(H_L/H_1)$ | Range of $(H_r = H_2/H_1)$ | $N$ (No. of cycles) | No. of runs |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PKW-1 | 0.005–0.050 | 0.0170–0.165 | 1.0 | 0.20 | 5 | 0.28 | 0.093 | 0.093 | 0.8413–0.153 | 0.158–0.8419 | 3 | 20 |
| PKW-2 | 0.005–0.050 | 0.0168–0.167 | 1.1 | 0.20 | 5 | 0.28 | 0.093 | 0.093 | 0.8310–0.1460 | 0.1682–0.8504 | 3 | 20 |
| PKW-3 | 0.005–0.050 | 0.0169–0.164 | 1.2 | 0.20 | 5 | 0.28 | 0.093 | 0.093 | 0.8180–0.1531 | 0.189–0.853 | 3 | 20 |
| PKW-4 | 0.005–0.050 | 0.0168–0.154 | 1.3 | 0.20 | 5 | 0.28 | 0.093 | 0.093 | 0.8012–0.1601 | 0.198–0.843 | 3 | 20 |
| PKW-5 | 0.005–0.050 | 0.0171–0.151 | 1.4 | 0.20 | 5 | 0.28 | 0.093 | 0.093 | 0.7931–0.1476 | 0.206–0.852 | 3 | 20 |
| PKW-6 | 0.005–0.050 | 0.0168–0.160 | 1.5 | 0.20 | 5 | 0.28 | 0.093 | 0.093 | 0.7841–0.1301 | 0.215–0.861 | 3 | 20 |
| PKW-7 | 0.005–0.050 | 0.0171–0.147 | 1.28 | 0.15 | 5 | 0.343 | 0.115 | 0.115 | 0.8110–0.1471 | 0.1889–0.8531 | 3 | 20 |
| PKW-8 | 0.005–0.050 | 0.0169–0.145 | 1.28 | 0.15 | 5 | 0.259 | 0.086 | 0.086 | 0.833–0.136 | 0.167–0.8642 | 4 | 20 |
| PKW-9 | 0.005–0.050 | 0.0172–0.146 | 1.28 | 0.15 | 5 | 0.208 | 0.069 | 0.069 | 0.853–0.1501 | 0.147–0.8498 | 5 | 20 |
| PKW-10 | 0.005–0.050 | 0.0172–0.145 | 1.28 | 0.15 | 6 | 0.427 | 0.142 | 0.142 | 0.8001–0.1302 | 0.199–0.869 | 3 | 20 |
| PKW-11 | 0.005–0.050 | 0.0168–0.141 | 1.28 | 0.15 | 6 | 0.322 | 0.107 | 0.107 | 0.8215–0.1492 | 0.179–0.8511 | 4 | 20 |
| PKW-12 | 0.005–0.050 | 0.0181–0.148 | 1.28 | 0.15 | 6 | 0.259 | 0.086 | 0.086 | 0.8439–0.1538 | 0.1574–0.8461 | 5 | 20 |

where $H$ represents the specific energy, $P$ represents the height of the weir in (m) (for d/s $P = 0$ m), $y$ is the head over the weir, and $V$ is the mean velocity at section $i$, respectively. The 'g' represents the gravitational acceleration, $i$ represents the section (i.e., $i = 1, 2\ldots$,) as shown in Figure 1(b)). $H_1$ and $H_2$ were then utilized to determine relative energy dissipation and relative residual energy ($H_r = H_2/H_1$) in the following way:

$$\frac{\Delta H}{H_1} = \frac{(H_1 - H_2)}{H_1} \times 100 = \left(1 - \frac{H_2}{H_1}\right) \times 100 \tag{2}$$

and,

$$H_r = 1 - \frac{\Delta H}{H_1} = \frac{H_2}{H_1} \tag{3}$$

where $\Delta H/H_1$ denotes total relative energy dissipation or energy dissipation ratio, and $H_r$ ($=H_2/H_1$) denotes the PKW's downstream residual energy. The energy dissipation over the PKW depends on various hydraulic and geometrical parameters: such as magnification ratio ($L/W$), head to weir height ratio ($H_t/P$), relative width ratio ($W_i/W_o$), inlet to outlet key slopes ($S_i/S_o$), number of cyclic variation ($N$), and overhang portions ($B_i$, $B_o$), etc. The relative length of the overhangs ($B_i/B_o$) ratio and relative key slopes ($S_i/S_o$) ratio were assumed to be one throughout the study for all models in the current study.

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2022.255/1073735/ws2022255.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof



The energy dissipation over the PKWs depends on different geometrical and hydraulic parameters. In the case of the present study, the following geometrical and hydraulic parameters were taken into account for calculating the residual energy downstream of the PKW.

$$\frac{H_2}{H_1} = f\ (\rho,\ g,\ V_t,\ H_t,\ L, P, W, W_i, W_o, S_i,\ S_o, B,\ B_i, B_o) \tag{4}$$

In which $\rho$ is the density of flowing fluid, $g$ is the acceleration of gravity, $H_t$ is the total upstream hydraulic head ($h_t + V_t^2/2g$) over the PKWs, and $V_t$ is the average velocity of flowing fluid. $L$ = total developed crest length; $P$ = height of PKW, $W$ = total width of PKW; $W_i$ and $W_o$ = widths of inlet and outlet keys sections, respectively, and $S_i$ and $S_o$ represent the inlet and outlet key slopes variation, respectively. $B_i$ and $B_o$ are the inlet-outlet overhang portions, B is the total length of the weir in the flow direction, $R$ = height of parapet wall, and $N$ = is the number of cycles. Six non-dimensional factors were used to create the GEP model: headwater ratio, magnification ratio, relative width ratio, inlet-outlet key slope ratio, relative overhang ratio, and the number of cyclic variations. The required dimensionless equation can be written as (Using the Buckingham theorem):

$$\frac{H_2}{H_1} = f\left(\frac{H_t}{P},\ \frac{L}{W},\ \frac{W_i}{W_o},\ \frac{S_i}{S_o},\ \frac{B_i}{B_o},\ N\right) \tag{5}$$

The above relationship describes the relative residual energy ratio as a function of geometric and hydraulic factors. The ranges of various parameters and data collected in the present study are summarized in Table 1.

### 2.3. An overview and application of GEP

Gene Expression Programming (GEP) is the learning algorithm behind GeneXproTools. It learns explicitly about relationships between variables in data sets and then builds models to explain these relationships. Gene expression programming uses character linear chromosomes, made up of genes structurally organized in a head and a tail, first encoded by (Ferreira 2001a). In GEP, chromosomes of various sizes and shapes can code in a simple graph (2001a, 2001b), and, like other evolutionary methods, GEP begins by randomizing early population chromosomes. It combines elements of genetic programming and genetic algorithms. The chromosomes function as a genome and are subject to mutation, transposition, root transposition, gene transposition, gene recombination, and one- and two-point recombination. The chromosomes encode the expression trees that are the aims of selection. In the GEP model, various fitness functions such as mean squared error (MSE), root mean squared error (RMSE), relative standard error (RSE), and root relative squared error (RRSE) can be used (Ferreira 2001a). The most advantageous chromosomes are likely to be passed down to future generations. Genetic operators perform the same acts with minor variations following selecting the best chromosomes.

The first step in choosing a GEP model is to select a fitness function. As a result, this study employs the RMSE function. The next step is to select the set of terminals and functions used to construct the chromosomes. The modeling process adopted in this study designates the relative residual energy ($H_2/H_1$) as the target value and the six independent parameters ($H_t/P$, $L/W$, $W_i/W_o$, $S_i/S_o$, $B_i/B_o$, and $N$) as input variables which are discussed in Equation (5). The basic operators (+, −, ×, /, ln, $x^2$, $e^x$, 1/x, $\sqrt[3]{x}$, Avg. of 2) were used to develop the GEP model.

The functions were chosen based on their coherence to the quiddity of the problem in order to achieve an uncomplicated and sensible GEP model. The general sampling strategy included selecting 30 chromosomes, three genes, and eight different head sizes. Table 2 shows that 240 data points were used in modeling and distributed randomly for the training and testing data phases. Approximately 80% of the data is used for training, while the remaining 20% is used to test the current project. The GEP's training and testing data were chosen randomly from the original dataset.

GeneXpro Tools 5.0 software package is used to compute in this study. It performs admirably on a personal computer. The defining criterion was the maximum fitness function, the mean square error function. The proposed GEP model's analytical

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2022.255/1073735/ws2022255.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof



**Table 2** | Functional set and operational parameters used in the GEP model

| S. No. | Description of Parameter (1) | Setting of Parameter (2) |
|---|---|---|
| 1. | Function Set | +, −, ×, /, ln, $x^2$, $e^x$, 1/x, $\sqrt[3]{x}$, Avg. of 2 |
| 2. | No. of Chromosomes | 30 |
| 3. | Head Size | 8 |
| 4. | No. of Genes | 3 |
| 5. | Gene Size | 26 |
| 6. | Linking Function | Addition |
| 7. | Fitness Function | RMSE |
| 8. | Program Size | 41 |
| 9. | Literals | 14 |
| 10. | Number of Generations | 1,15,310 |
| 11. | Constants per Gene | 10 |
| 12. | Data Type | Floating-point |
| 13. | Mutation | 0.00138 |
| 14. | Inversion | 0.00546 |
| 15. | Gene recombination rate | 0.00277 |
| 16. | One-point recombination rate | 0.00277 |
| 17. | Two-point recombination rate | 0.00277 |
| 18. | Gene transposition rate | 0.00277 |
| 19. | Insertion sequence (IS) transposition rate | 0.00546 |
| 20. | Root insertion sequence (RIS) transposition rate | 0.00546 |

form is as follows:

$$\frac{H_2}{H_1} = \left[\frac{\left(\frac{B_i}{B_o}\right)}{\left[e^{\left[\frac{\left(\left(\frac{H_t}{P}\right)+(-4.04684)\times(0.230272)\times N\right)}{\left(\frac{H_t}{P}\right)}\right]}\right]^{\frac{1}{3}}}\right] + \left[\left(\frac{H_t}{P}\right) - \left[\frac{e^{\left\{-5.60185\times\left(\frac{L}{W}\right)\times\left(\frac{H_t}{P}\right)\right\}}}{e^{\left\{\frac{\left(\frac{S_i}{S_o}\right)+\left(\frac{H_t}{P}\right)}{2}\right\}}}\right]\right] + \left[\frac{\left[\frac{\left\{\frac{-7.34305\times\left(\frac{H_t}{P}\right)^2}{\left(\frac{L}{W}\right)}\right\}}{\left\{\frac{\left(\frac{B_i}{B_o}\right)+\left(\frac{W_i}{W_o}\right)}{2}\right\}}\right]}{\left(\frac{1}{0.06952}\right)^{\frac{1}{3}}}\right] \quad (6)$$

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2022.255/1073735/ws2022255.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof



Further, simplified the above equation as

$$\frac{H_2}{H_1} = \left[ \frac{\left(\frac{B_i}{B_o}\right)}{e^{\left\{ \frac{\left(\left(\frac{H_t}{P}\right) - 4.04684\right) \times (0.230272) \times N}{\left(\frac{H_t}{P}\right)} \right\}^{\frac{1}{3}}} \right] + \left[ \left(\frac{H_t}{P}\right) - \left\{ \frac{e^{\left\{ -5.60185 \times \left(\frac{L}{W}\right) \times \left(\frac{H_t}{P}\right) \right\}}}{e^{\left\{ \frac{\left(\frac{S_i}{S_o}\right) + \left(\frac{H_t}{P}\right)}{2} \right\}}} \right\} \right] + \left[ \frac{-6.03869 \times \left(\frac{H_t}{P}\right)^2}{\left(\frac{L}{W}\right) \times \left\{ \left(\frac{B_i}{B_o}\right) + \left(\frac{W_i}{W_o}\right) \right\}} \right] \tag{7}$$

Figure 2 depicts an expression tree (ET) and Equation (6) show the representation of the GEP model for predicting downstream relative residual energy. In Figure 2, $d0$, $d1$, $d2$, $d3$, $d4$, and $d5$ represent the $H_t/P$, $L/W$, $W_i/W_o$, $S_i/S_o$, $B_i/B_o$, and the number of cycles ($N$), respectively. Whereas $G1c0$ and $G1c5$ show the numerical constants used in the first gene of the model, $G2c4$ is the constant used for the second gene model; similarly, $G3c1$ and $G3c3$ denote the constant used in the third gene of the model.

## 3. RESULTS AND DISCUSSION

The flow over the PKW is three-dimensional and follows a complicated pattern. It (Flow over PKW) is the sum of the flow over the downstream crest, the flow over the upstream crest, and the sidelong flow over the side crest (Machiels 2012), as shown in Figure 3. The literature said the geometry of the PKW has significantly influenced the flow capacity of the weir. Increasing the inlet key width decreases flow velocities at the inlet key and substantially increases PKW efficiency (Eslinger & Crookston 2020; Singh & Kumar 2022c). However, for a given weir width and crest length, increasing the inlet width causes a decrease in the outlet width. Too narrow outlet key may not evacuate the outlet flow under supercritical conditions, lowering the overall weir efficiency. A high $W_i/W_o$ value improves the flow approach and distribution within the inlet keys and increases submergence effects in outlet keys of PKWs. Thus, an optimal $W_i/W_o$ ratio must be found by balancing inlet width increase and outlet resilience capacity. Similarly, the $L/W$, $S_i/S_o$, and $B_i/B_o$ proportions influence the discharging capacity of the weir at a significant level.

In order to better predict or accurate estimation of the residual energy downstream of the PKW, the Gene Expression Programming (GEP) technique was used to measure the downstream residual energy. The GEP approach demonstrates a highly nonlinear relationship between relative residual energy and the input parameters ($H_t/P$, $L/W$, $W_i/W_o$, $S_i/S_o$, $B_i/B_o$, and $N$) with high accuracy and relatively low errors. After multiple generations, the program was discontinued due to no progress in the fitness function value or the coefficient of determination. After 1,15,310 generations, there was no discernible difference noticed. All of the parameters mentioned were chosen through trial and error in order to obtain the best model of the GEP in the form of an algebraic equation between output and input variables.

According to the author, the proposed equation of second-order exponential approximation may be adequate for conceptual designs and alternative analyses of the relative residual energy estimation. The $R^2$ value in the current approach for the two datasets (testing = 0.978 and training = 0.980) demonstrates the model's adequacy. GEP reported RMSE values of 0.032 and 0.029 for the training and testing data sets. ME, MAE, and MAPE values for training data were 3.26%, 0.025, and 5.30%, and for the testing, data were 4.21%, 0.016, and 5.40%, respectively, showing the performance and accuracy of the predicted model.

As a result, the developed GEP is a reliable method for predicting relative residual energy ($H_2/H_1$) with good generalization and no overtraining. The proposed GEP has a significant advantage over traditional regression-based models (traditional equations). Depending on the number of generations, the GEP model can predict the relative residual energy with high accuracy and in a short amount of time. It is capable of mapping the data into a high-dimensional feature space, where various

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2022.255/1073735/ws2022255.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof



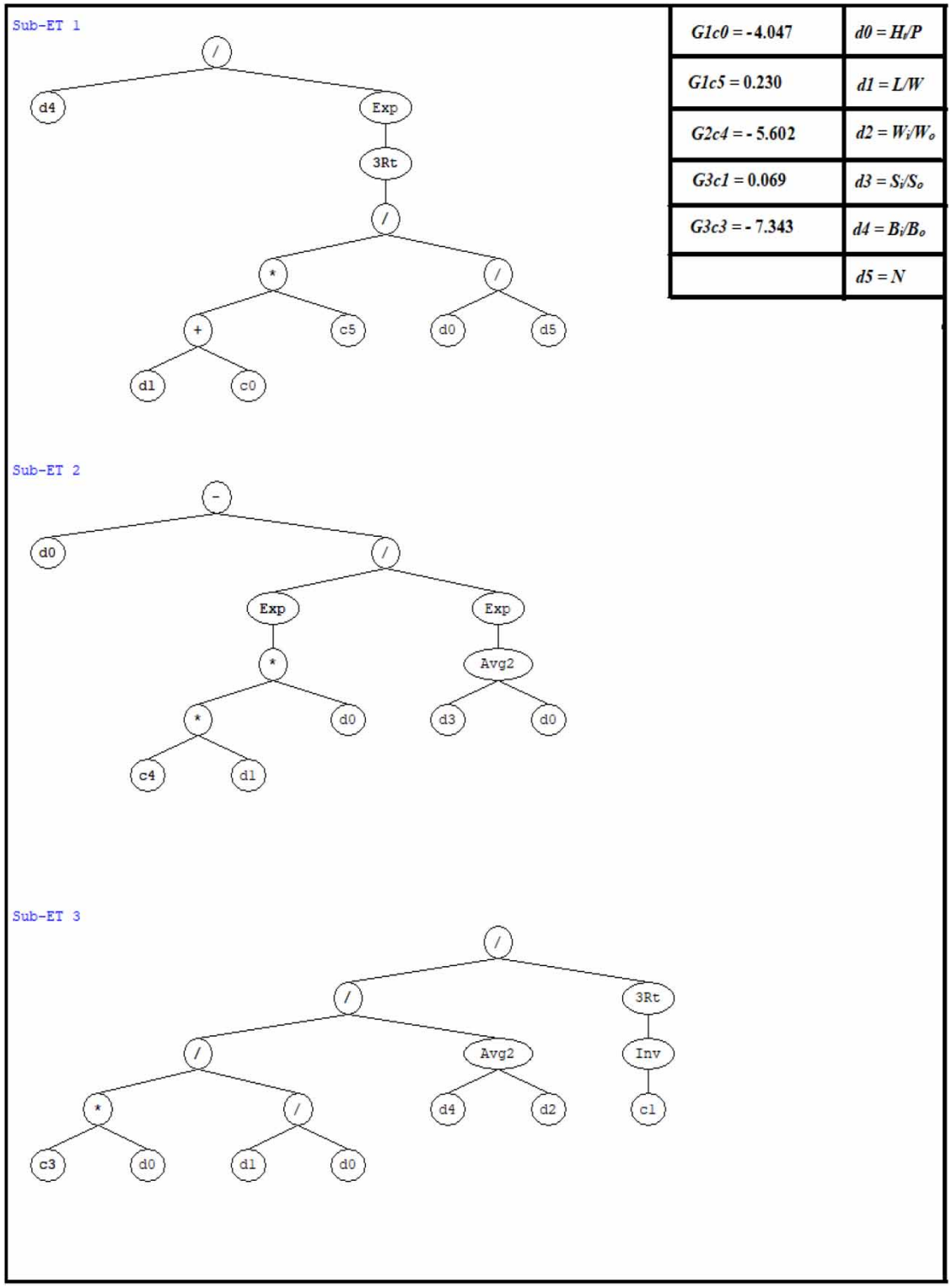



**Figure 2** | Expression Tree (ET) for GEP formulation.

methods are used to discover data relationships. The relationships are as diverse as the mapping. The constructed GEP model demonstrated excellent agreement with the observed values in terms of coefficient of determination ($R^2$) and root mean square error (RMSE) for training and testing datasets.

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2022.255/1073735/ws2022255.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof



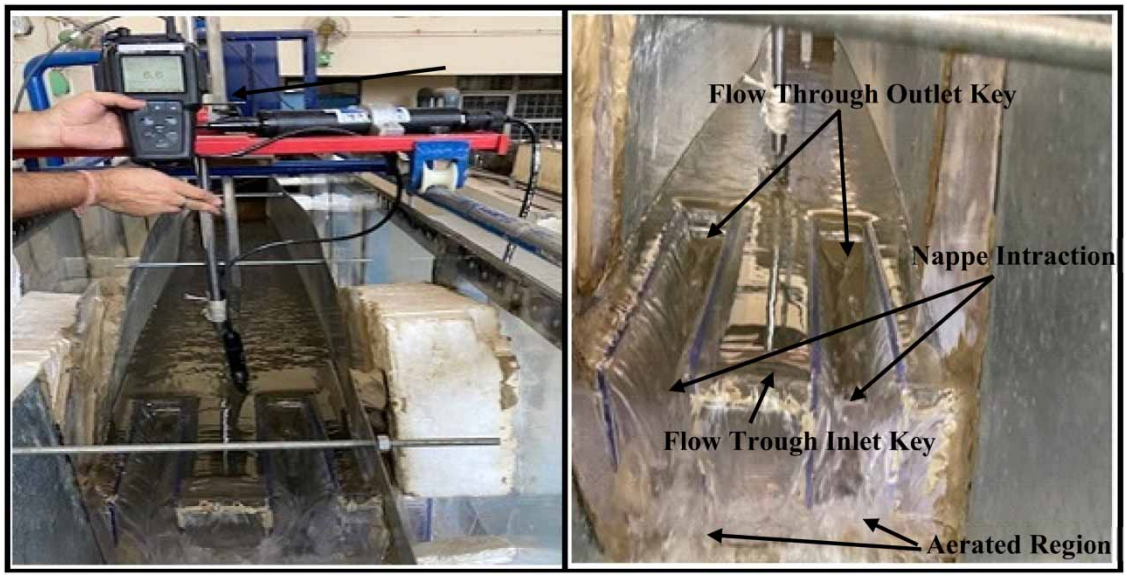



**Figure 3** | Flow pattern over PKW.

Figure 4 represents the performance matrics of the observed and predicted values of relative residual energy ($H_2/H_1$) achieved by the GEP approach Equation (7) for the training and testing data phases separately. Following the development of the GEP model (Equation (6)), an attempt is made to test its robustness. The authors used Equation (6) to demonstrate the accuracy of the predicted residual energy ($H_2/H_1$) value with the observed value, as shown in Figure 5, for different data used in the training and testing data phases individually. The performance of the developed GEP model was assessed using the various statistical parameters shown in Table 3.

## 4. CONCLUSION

Accurately estimating the relative energy dissipation or residual energy downstream of the weir structures is critical for hydraulic engineers and designers. This paper developed an empirical equation based on Gene Expression Programming (GEP) to predict the relative residual energy downstream of the type-A PKW. The GEP approach produced a highly nonlinear

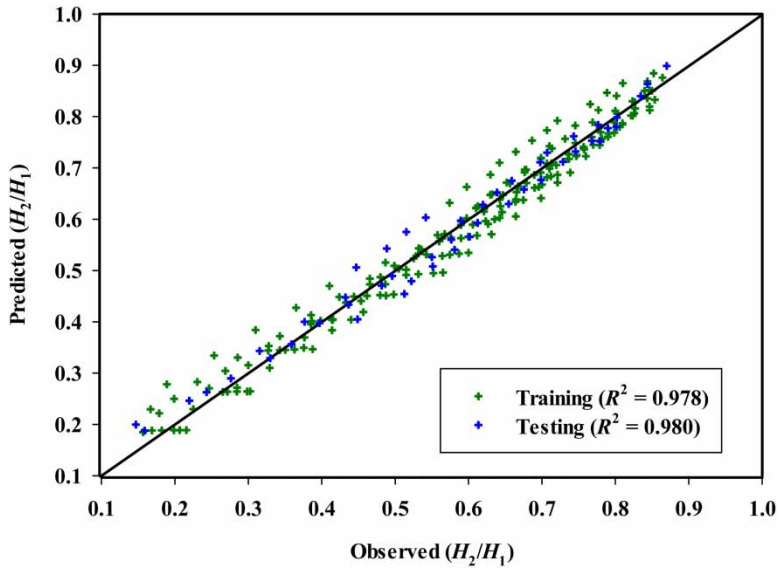



**Figure 4** | Compares the observed and projected relative residual energy ratios for different data phases (training and testing phases).

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2022.255/1073735/ws2022255.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof



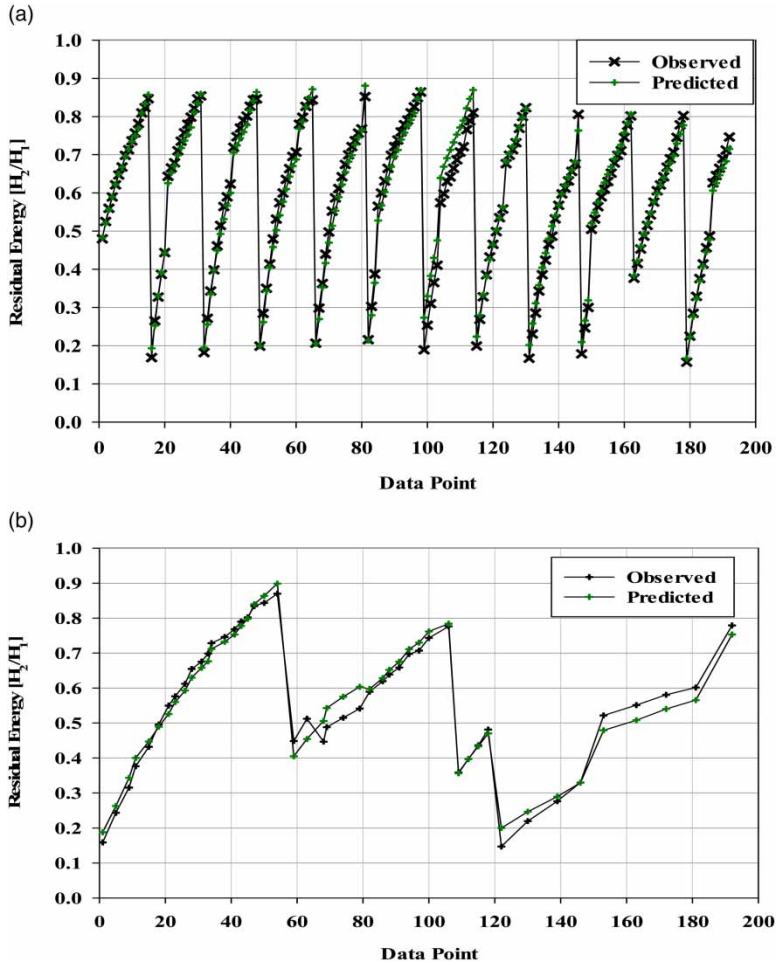



**Figure 5** | Performance matrices for predicting the relative residual energy ($H_2/H_1$) ratios for the different data phases; (a) Training phase; (b) testing phase.

**Table 3** | Performance evaluation of predicted $H_2/H_1$ by GEP model for training and testing dataset

| Data Set | CC | $R^2$ | ME % | MAE | RMSE | MAPE % |
|---|---|---|---|---|---|---|
| Training | 0.986 | 0.978 | 3.26 | 0.025 | 0.032 | 5.30 |
| Testing | 0.988 | 0.980 | 4.21 | 0.016 | 0.029 | 5.40 |

CC, Coefficient of correlation; $R^2$, Coefficient of determination; ME, Mean Percentage Error; MAE, Mean Absolute Error; RMSE, Root Mean Square Error; MAPE, Mean Absolute Percentage Error.

relationship between relative residual energy and input parameters according to the proposed equation for relative residual energy. The GEP model's comparisons over the existing models show that the GEP models provide better predictions, and practitioners can use them as design approximations to assist researchers and design engineers in designing PKW. The GEP model had the lowest $R^2$ (0.978 for training and 0.980 for testing), and root means square error (RMSE) values of 0.032 and 0.029 for the training and testing datasets. The results suggest the efficiency of the GEP model and its potential use for practical applications within a similar range of non-dimensional parameters tested in this work. According to the findings of this study, the GEP model is more beneficial for any condition with no limitations. More research into the energy dissipation estimation of PKW is recommended. This could include other geometries like PKW types B, C, and D, or it could use turbid water to calculate the energy surplus.

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2022.255/1073735/ws2022255.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof



## DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

## CONFLICT OF INTEREST

The authors declare there is no conflict.

## REFERENCES


Abhash, A. & Pandey, K. K. 2020 A review of piano key weir as a superior alternative for dam rehabilitation a review of piano key weir as a superior alternative for dam rehabilitation. *ISH Journal of Hydraulic Engineering* **00** (00), 1–11. Taylor & Francis.

Al-Shukur, A.-H. K. & Al-Khafaji, G. H. 2018 Experimental study of the hydraulic performance of piano key weir. *International Journal of Energy and Environment* **9**, 63–70.

Anderson, R. M. & Tullis, B. P. 2012 Piano key weir: reservoir versus channel application. *Journal of Irrigation and Drainage Engineering* **138** (8), 773–776.

Azamathulla, H. M., Ahmad, Z. & Ghani, A. A. 2013 An expert system for predicting Manning's roughness coefficient in open channels by using gene expression programming. *Neural Computing and Applications* **23** (5), 1343–1349.

Azamathulla, H. M., Rathnayake, U. & Shatnawi, A. 2018 Gene expression programming and artificial neural network to estimate atmospheric temperature in Tabuk, Saudi Arabia. *Applied Water Science* **8** (6), 1–7. Springer International Publishing.

Bieri, M., Federspiel, M., Boillat, J.-L., Houdant, B., Faramond, L. & Delorme, F. 2011 Energy dissipation downstream of piano key weirs – case study of Gloriettes Dam (France). In: *Labyrinth and Piano Key Weirs – PKW 2011*. CRC Press, Leiden, The Netherlands, pp. 120–130.

Birbal, P., Azamathulla, H., Leon, L., Hosein, J. & Kumar, V. 2021 Predictive modelling of the stage-discharge relationship using gene-expression programming. *Water Supply* **21** (7), 3503–3514.

Chanson, H. 1994 Comparaison de la dissipation d'énergie pour des écoulements en nappe et extrémement turbulents sur des évacuateurs de crue en marches d'escaliers. *Journal of Hydraulic Research* **32** (2), 213–218.

Chanson, H. 2021 Hydraulics and energy dissipation on stepped spillways – prototype and laboratory experiences. **15** (1), 5–8.

Emadi, A., Sobhani, R., Ahmadi, H., Boroomandnia, A., Zamanzad-Ghavidel, S. & Mohammad Azamathulla, H. 2022 Multivariate modeling of river water withdrawal using a hybrid evolutionary data-driven method. *Water Supply* **22** (1), 957–980.

Erpicum, S., Laugier, F., Pfister, M., Pirotton, M., Cic, M. & Schleiss, J., *et al.* 2013 Labyrinth and Piano Key Weirs II PKW 2013. In: *Labyrinth and Piano Key Weirs II – PKW 2013* (Erpicum *et al.* ed.). © 2014 Taylor & Francis Group, London. ISBN 978-1-138-00085-8

Eslinger, K. R. & Crookston, B. M. 2020 Energy dissipation of type a piano key weirs. *Water*. MDPI, Basel, Switzerland.

Ferreira, C. 2001a Gene expression programming: a new adaptive algorithm for solving problems. 1–22.

Ferreira, C. 2001b Gene expression programming in problem-solving. In *6th Online World Conference on Soft Computing in Industrial Applications (Invited tutorial)*.

Garg, S. K. 2010 *Irrigation Engineering and Hydraulic Structures*. Khanna publishers, New-Delhi- 110002.

Ghaderi, A., Abbasi, S., Abraham, J. & Azamathulla, H. M. 2020 Efficiency of Trapezoidal Labyrinth Shaped stepped spillways. *Flow Measurement and Instrumentation* **72** (February), 101711. Elsevier Ltd.

Ghasempour, R., Roushangar, K., Azamathulla, H. M. & Roushangar, K. 2021 EEMD- And VMD-based hybrid GPR models for river streamflow point and interval predictions. *Water Supply* **21** (7), 3960–3975.

Hager, W. H., Bremen, R. & Kawagoshi, N. 1990 Classical hydraulic jump: length of roller. *Journal of Hydraulic Research* **28** (5), 592–608.

Ho Ta Khanh, M., Sy Quat, D. & Xuan Thuy, D. 2011a *'PKWs Under Design and Construction in Vietnam (2010).' Labyrinth and Piano key Weirs-PKW 2011: 225-232*. CRC Press, London.

Ho Ta Khanh, M., Hien, T. C. & Hai, N. T. 2011b *'Main Results of the PKW Models Test in Vietnam. (2004–2010)'. Labyrinth and Piano key Weirs-PKW 2011: 191-198*. CRC Press, London.

Jüstrich, S., Pfister, M. & Schleiss, A. J. 2016 Mobile riverbed scour downstream of a piano key weir. *Journal of Hydraulic Engineering* **142** (11), 1–12.

Karbasi, M. & Azamathulla, H. M. 2016 GEP to predict characteristics of a hydraulic jump over a rough bed. *KSCE Journal of Civil Engineering* **20** (7), 3006–3011.

Kumar, B. & Ahmad, Z. 2022 Scour downstream of a piano key weir with and without a solid apron. *Journal of Irrigation and Drainage Engineering* **148** (1), 1–12.

Kumar, B., Kadia, S. & Ahmad, Z. 2021 Sediment movement over type A piano key weirs. *Journal of Irrigation and Drainage Engineering* **147** (6), 04021018.

Leite Ribeiro, M., Boillat, J. L., Schleiss, A., Laugier, F. & Albalat, C. 2007 Rehabilitation of St-Marc Dam – Experimental Optimization of a Piano Key Weir. In *Proceedings of the 32nd Congress of Int. Association for Hydro-Environment Engineering and Research (IAHR) (CD-ROM),(683) Theme C (C2.B-149-O)Venice, Italy 1–6 July 2007*.

Leite Ribeiro, M., Boillat, J.-L. & Schleiss, A. J. 2011 Experimental parametric study for hydraulic design of PKWs. In *' Proc. Int. Conf. Labyrinth and Piano Key Weirs, Liege B, 183–190*. CRC Press, Boca Raton, FL.


Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2022.255/1073735/ws2022255.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof



Leite Ribeiro, M., Bieri, M., Boillat, J. L., Schleiss, A. J., Singhal, G. & Sharma, N. 2012 Discharge capacity of piano key weirs. *Journal of Hydraulic Engineering* **138** (2), 199–203.

Machiels, O. 2012 *Experimental Study of the Hydraulic Behaviour of Piano Key Weirs*. *PhD Thesis ULgetd-09252012- 224610*, University of Liege (Belgium).Acknowledgments.' (June).

Pandey, M. & Md Azamathulla, H. 2021 Discussion of 'Gene-expression programming, evolutionary polynomial regression, and model tree to evaluate local scour depth at culvert outlets' by Mohammad Najafzadeh and Ali Reza Kargar. *Journal of Pipeline Systems Engineering and Practice* **12** (2), 07021001.

Pandey, M., Zakwan, M., Khan, M. A. & Bhave, S. 2020a Development of scour around a circular pier and its modelling using genetic algorithm. *Water Science and Technology: Water Supply* **20** (8), 3358–3367.

Pandey, M., Zakwan, M., Sharma, P. K. & Ahmad, Z. 2020b Multiple linear regression and genetic algorithm approaches to predict temporal scour depth near circular pier in non-cohesive sediment. *ISH Journal of Hydraulic Engineering* **26** (1), 96–103. Taylor & Francis.

Pandey, M., Jamei, M., Karbasi, M., Ahmadianfar, I. & Chu, X. 2021 Prediction of maximum scour depth near spur dikes in uniform bed sediment using stacked generalization ensemble tree-based frameworks. *Journal of Irrigation and Drainage Engineering* **147** (11), 04021050.

Pfister, M., Jüstrich, S. & Schleiss, A. J. 2017 Toe-scour formation at Piano Key Weirs. In: *Labyrinth and Piano Key Weirs III – PKW 2017*. CRC Press, Leiden, The Netherlands, pp. 147–156.

Pinto, M. 2017 Energy dissipation on stepped spillways with a piano key weir : experimental study. 1–11.

Rajaratnam, N. 1990 Skimming flow in stepped spillways. *Journal of Hydraulic Engineering, ASCE* **116** (4), 587–591.

Rashki Ghaleh Nou, M., Azhdary Moghaddam, M., Shafai Bajestan, M. & Azamathulla, H. M. 2019 Estimation of scour depth around submerged weirs using self-adaptive extreme learning machine. *Journal of Hydroinformatics* **21** (6), 1082–1101.

Shivashankar, M., Pandey, M. & Zakwan, M. 2022 Estimation of settling velocity using generalized reduced gradient (GRG) and hybrid generalized reduced gradient–genetic algorithm (hybrid GRG-GA). *Acta Geophysica*. Springer International Publishing, (0123456789).

Silverstri, A., Erpicum, S., Archambeau, P., Dewals, B. & Pirotton, M. 2013a Stepped spillway downstream of a Piano Key weir- Critical length of uniform flow. In *International Workshop on Hydraulic Structures*. Bundesanstalt fur, Wasserbau, Karlsruhe, Germany, pp. 99-107.

Silverstri, A., Archambeau, P., Pirotton, M., Dewals, B. & Erpicum, S. 2013b Comparative analysis of the energy dissipation on a stepped spillway downstream of a Piano Key weir. In: *Labyrinth and Piano key Weirs II-PKW 2013*. CRC Press, London, pp. 111-120.

Silvestri, A., Archambeau, P., Pirotton, M., Dewals, B., Erpicum, S., 2013c Comparative Analysis of the Energy Dissipation on a Stepped Spillway Downstream of a Piano Key Weir. In: *Labyrinth and Piano Key Weirs II* (Erpicum, S., Laugier, F., Pfister, M., Pirotton, M., Cicero, G. M. & Schleiss, A. J., eds). CRC Press/Balkema, Leiden, The Netherlands, pp. 111–120.

Singh, D. & Kumar, M. 2021 Hydraulic design and analysis of piano key weirs : a review. *Arabian Journal for Science and Engineering*. Springer Berlin Heidelberg.

Singh, D. & Kumar, M. 2022a Study on aeration performance of different types of piano Key weir. *Water Supply* **00** (0), 1–12.

Singh, D. & Kumar, M. 2022b Energy dissipation of flow over the type-B Piano Key Weir. *Flow Measurement and Instrumentation* **83** (November 2021), 102109. Elsevier Ltd.

Singh, D. & Kumar, M. 2022c Gene expression programming for computing energy dissipation over type-B Piano Key Weir. *Renewable Energy Focus*. Elsevier Ltd, (March 2022). https://doi.org/10.1016/j.ref.2022.03.005.

Singh, U. K., Jamei, M., Karbasi, M., Malik, A. & Pandey, M. 2022 Application of a modern multi-level ensemble approach for the estimation of critical shear stress in cohesive sediment mixture. *Journal of Hydrology* **607** (February).

Truong Chí, H., Huynh Thanh, S. & Ho Ta Khanh, M. 2006 Results of Some Piano Keys Weir Hydraulic Model Tests in Vietnam. In *Proceedings of the 22nd ICOLD Congress*, 18–23 June 2006, Barcelona, Spain.

Yazdi, A. M., Abbas Hoseini, S., Nazari, S. & Amanian, N. 2021 Effects of weir geometry on scour development in the downstream of Piano Key Weirs. *Water Science and Technology: Water Supply* **21** (1), 289–298.

Zhao, C., Fang, H., Liu, Y., Dey, S. & He, G. 2020 Impact of particle shape on saltating mode of bedload transport sheared by turbulent flow. *Journal of Hydraulic Engineering* **146** (5), 04020034. https://doi.org/10.1061/(ASCE) HY.1943-7900.0001735.

Zounemat-Kermani, M. & Mahdavi-Meymand, A. 2019 Hybrid meta-heuristics artificial intelligence models in simulating discharge passing the piano key weirs. *Journal of Hydrology* **569** (November), 12–21. Elsevier.

First received 31 March 2022; accepted in revised form 30 June 2022. Available online 6 July 2022

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2022.255/1073735/ws2022255.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

ENVIRONMENTAL ENGINEERING SCIENCE
Volume 00, Number 00, 2022
© Mary Ann Liebert, Inc.
DOI: 10.1089/ees.2022.0020

Open camera or QR reader and scan code to access this article and other resources online.

# COVID-19 Lock-down in Delhi: Understanding Trends of Particulate Matter in Context of Land-Use Patterns, GIS Mapping, and Meteorological Traits

Amrit Kumar,[1] Kiranmay Sarma,[2] Abhinav Pandey,[3,4,*]
Rajeev Kumar Mishra,[3] and Panuganti China Sattilingam Devara[1]

[1]*Amity Centre for Ocean-Atmospheric Science and Technology, Amity University Haryana, Gurugram, India.*
[2]*School of Environment Management, Guru Gobind Singh Indraprastha University, Delhi, India.*
[3]*Department of Environmental Engineering, Delhi Technological University, Delhi, India.*
[4]*Engineering Design & Research Centre, Larsen & Toubro Limited, New Delhi, India.*

*Received: January 20, 2022* *Accepted in revised form: June 11, 2022*

Downloaded by 14.139.251.98 from www.liebertpub.com at 07/31/22. For personal use only.

## Abstract

The effectiveness and cost implications are always top factors for policy makers while deciding upon the appropriate air pollution abatement measures. The present study aimed to understand the actual particulate matter ($PM_{2.5}$ and $PM_{10}$) patterns during different phases of COVID-19 lockdown periods and depict their spatial distributions covering the 36 major areas in Delhi, India. Drastic visible reduction in both the pollutants was found during lockdown phase 1 and 2. Average $PM_{2.5}$ reductions of 41.97%, 39.24%, 56.04%, and 56.77% were recorded comparing lockdown and/or study period with the years 2018, 2019, 2021, and 2022, respectively. Similar average reduction of $PM_{10}$ to the magnitude of 51.72%, 48.95%, 48.24%, and 49.00% was found for the referred years. However, the reduction during the before-lockdown period of 2018 and 2019 and the year 2020 did not follow such radical reduction returning the values for $PM_{2.5}$ as 7.66–14.88% and that for $PM_{10}$ as 12.86–20.67%. The geospatial maps generated for Delhi city followed the similar findings at macro level depicting huge reduction in PM distribution classes for the study period. For instance, the percent surface area under ''moderately high'' polluted due to $PM_{2.5}$ came down to 0.61 during lockdown phase 2 from 13.96 during January 2020. Further, about 15 of the 36 locations reported compliance to the National Ambient Air Quality Standards (NAAQS) for either of the pollutants during the study period. Nevertheless, such reductions are short-lived because the levels went up again in the years 2021 and 2022 (except similar lockdowns) as the situation got back to normal daily life activities postlockdown. Although, lockdown may be imposed in case of severe ambient air quality in a densely populated megacity like Delhi, it remains a temporary or quick-fix solution, to be looked as a last line of defense.

***Keywords:*** air quality; COVID-19; curfew day; lockdown; particulate matter; pre-lockdown

## Introduction

THE URBAN CONGLOMERATIONS worldwide are subject to adverse environmental effects attributed by its high population density, intense traffic volume aggravated by the ever-growing motorized transport, vehicle characteristics and also, the driving characteristics. Compared to other similar urban setup(s), the ambient concentration of air pollutants in Delhi has been reported to be in violation of limits prescribed by the World Health Organization (WHO) and National Ambient Air Quality Standard (NAAQS, India) year after year. This perennial scenario has been raising serious concerns pointing toward the remarkably worse environmental quality in the region (Mohan and Kandya, 2007; Dandona *et al.*, 2017; Kumar *et al.*, 2020). The deteriorating air quality of Delhi, mainly during the winters, compelled the state government implementing odd-even numbers driving system on the roads, which brought small but positive changes in air quality (Mishra *et al.*, 2019).

**Corresponding author:* Department of Environmental Engineering, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi 110042, India. *Phone:* (+91)-9818774654; *Fax:* (+91)-11-27871023; *E-mail:* abhigkp@lntecc.com

Downloaded by 14.139.251.98 from www.liebertpub.com at 07/31/22. For personal use only.

The mass concentration of particulate matter (PM) largely comprises $PM_{10}$ (inhalable particles, with diameters $\leq 10\,\mu$m, e.g., dust, pollen, mold etc.) and $PM_{2.5}$ (fine inhalable particles, with diameters, in-general, $\leq 2.5\,\mu$m, e.g., burning particles, carbon-based compounds, metals etc.). This heterogeneous mass of particles is known to be able to travel acutely into the lungs of human bodies and thereby causing respiratory problems like emphysema and bronchitis leading to an increase in hospital visits and early death(s) (Adar *et al.*, 2013).

A local outbreak of pneumonia of primarily unknown cause, was noticed at Wuhan in December 2019 (Hubei, China), and was rapidly determined to be produced by a novel coronavirus, specifically, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Dutheil *et al.*, 2020; Fann *et al.*, 2012; Gautam and Hens, 2020). The outbreak since then extended to each area of mainland China as well as other continents with almost no country untouched (Devara *et al.*, 2020; Filonchyk and Peterson, 2020). Following the epidemic, as on February 21, 2020, the total confirmed cases of COVID-19 in the world were 110,763,898 with a casualty of 2,455,331 numbers (WHO, 2021).

At this time, India was placed third in terms of confirmed case of COVID-19 (Table 1). Delhi had recorded 145 new COVID-19 cases on this date (registering just about 2.15% of numbers registered on November 20, 2020 as 6,746 numbers) whereas its tally of total cases rose to 6,37,900 from 5,29,863 recorded on November 20, 2020, thereby reporting 16.9% higher figures.

There have been several researches focusing on the air quality assessment in view of the lockdown restrictions imposed by the Governments worldwide. A decline of about 14% (from 30 to 26 $\mu g/m^3$) in $PM_{2.5}$, $\sim$30% (10–7 $\mu g/m^3$) in Delhi and Chennai was reported. Further, a similar trend was found for lockdown period March to April 2020 while comparing air quality data of the previous year March to April 2019 except for $PM_{2.5}$ (89 $\mu g/m^3$) in Delhi and $PM_{10}$ levels in Delhi (at 222 $\mu g/m^3$), Kolkata (at 102 $\mu g/m^3$), and Bangalore (at 118 $\mu g/m^3$) (Jain and Sharma, 2020).

The metropolitan cities of Kolkata, Bangalore, and Mumbai reported a similar decline of $PM_{10}$ as 34%, 42%, and 47%, respectively, which were observed ''during'' lockdown phase in comparison to the ''before'' lockdown phase. The study also reported alike reduction in $PM_{10}$ concentration levels in Vienna (−60.7%), Bern (−53.3%), Paris (−52.7%), Warsaw (−45.9%), and Dublin (−44.3%) during February. On the contrary, Ulaanbaatar (−43.4%), Delhi (−32.5%), Nanjing (−31.8%), Wuhan (−26.8%), and Lima (−25.7%) witnessed maximum reduction in March (Shrestha *et al.*, 2020). A study on the impact of the lockdown period (partial, from January to March 2020) on ambient air quality in 6 cities of Hubei province, China, along with six megacities in India reported that Cities in China and India post-1-week-of-lockdown witnessed an average decline in $AQIPM_{2.5}$ and $AQINO_2$ of 11.32% and 48.61% as well as 20.21% and 59.26%, subsequently.

The results indicated that such a drop in $AQINO_2$ was instantaneous when compared with the rather gradual decline in $AQIPM_{2.5}$ (Agarwal *et al.*, 2020). In an alike study, Air Quality Index (AQI) variations were analyzed the before and after nationwise lockdown reporting a major sustaining type of decline in the average AQI values for major Type-1 and Type-2 cities due to the reduction in public movement and industrial activities, whereas the reduction in AQI values for Type-3 and Type-4 cities was observed to be fluctuating due to the continued small-scale industrial activities and low awareness level about ban on public gathering programs (Selvi *et al.*, 2020).

An analysis using the National Air Quality Index (NAQI) to demonstrate the spatial pattern of air quality before and during lockdown phases in Delhi witnessed positive change therein. PM revealed maximum reduction (>50%) in comparison to the prelockdown phase, while ambient levels of other gaseous pollutants, such as $NO_2$ (−52.68%) and CO (−30.35%) were also reduced during-lockdown phase reporting about 40% to 50% betterment. Further, about 54%, 49%, 43%, 37%, and 31% reduction in NAQI were observed in Central, Eastern, Southern, Western, and Northern areas of the Indian capital city (Mahato *et al.*, 2020).

Observation data from the China National Environmental Monitoring Center (CNEMC) revealed that compared to levels in 2019, the average concentration of $PM_{2.5}$ got reduced by 35%, 29%, and 19% in Wuhan, Hubei (Wuhan excluded), and China (Hubei excluded), respectively. The temporal variation and spatial distribution reported decrease in $PM_{2.5}$ and $NO_2$ concentrations, which was found to be relatively consistent (Chu *et al.*, 2020). An investigation on the impact of short-term lockdown during the period from March 15 to April 12, 2020 on the ambient levels of gaseous and PM in about 11 Spanish cities found that lockdown had a considerable impact on observed reduction in $PM_{10}$ in almost all cities, however, did lead to increase in $O_3$ levels in a few areas (Briz-Redón *et al.*, 2020).

Interestingly, lockdown scenarios contributed to significant reduction of $PM_{2.5}$ and $PM_{10}$ in ambient air quality by way of surface and off-surface transport activities and also through fuel combustion in commercial and institutional buildings; however, such reductions were counter-balanced

Table 1. Status of Cases Related to COVID-19 (as on February 21, 2021)

| *WHO region* | *Total cases* | *New cases in last 7 days* | *Total deaths* | *New deaths in last 7 days* |
|---|---|---|---|---|
| Africa | 2,789,884 | 66,453 | 70,322 | 2,038 |
| Americas | 49,296,115 | 1,066,990 | 1,171,294 | 34,386 |
| Eastern Mediterranean | 6,181,023 | 1,81,969 | 141,915 | 2,443 |
| Europe | 37,574,211 | 939,271 | 838,761 | 24,102 |
| South-East Asia (including India) | 13,345,590 | 157,379 | 204,796 | 2,189 |
| Western Pacific | 1,576,330 | 44,964 | 28,220 | 1,201 |
| India alone | 10,991,651 | 86,711 | 156,302 | 660 |
| Overall | 110,763,898 | 2,457,026 | 2,455,331 | 66,359 |

WHO, World Health Organization.

Table 2. Schedule of Lockdown Restrictions in Delhi

| *Restriction phase* | *Period* | *Span (days)* |
|---|---|---|
| Prelockdown (including one Curfew day) | January to March 2020 | 84 |
| Lockdown—Phase 1 (LDP1) | March to April 2020 | 20 |
| Lockdown—Phase 2 (LDP2) | April to May 2020 | 20 |
| Lockdown—Phase 3 (LDP3) | May 2020 | 14 |
| Lockdown—Phase 4 (LDP4) | May 2020 | 14 |

by an increase of PM emissions from the household activities (e.g., domestic heating and biomass burning (Devara *et al.*, 2020; Sicard *et al.*, 2020).

After a thorough review of concurrent literature and understanding the gaps, the present study was undertaken with the following objectives:

- Land-use pattern (LUP)-wise average variation of $PM_{2.5}$ and $PM_{10}$ over the 36 study sites during lockdown phases.
- LUP-wise variation in $PM_{2.5}$:$PM_{10}$ ratios during lockdown phases and through 2019–2022.
- LUP-wise average changes in $PM_{2.5}$ and $PM_{10}$ through 2019–2022.
- Average meteorological conditions vis-à-vis $PM_{2.5}$ and $PM_{10}$ variations through 2019–2022.
- Spatial distribution of $PM_{2.5}$ and $PM_{10}$ over the entire city of Delhi comprising different distribution classes, namely, ''low,'' ''moderately low,'' ''moderate,'' ''moderately high,'' and ''high.'' The geographic information system (GIS) tool-generated maps were presented as a supplemental analysis to get a glimpse of how the PM was distributed over the entire city and was limited only to the study period (January 2020 to May 2020).
- Assessment of compliance of PM levels toward the extant NAAQS.

## Research Methodology

### *Study area*

The National Capital Territory (NCT) of Delhi has a total geographical area of 1,483 $km^2$ with diverse LUPs (CPCB,

Table 3. Selected Locations of Study Area Covered by Existing Air Quality Monitoring Stations

| *Name of location (AQMS)* | *Symbol* | *Location (longitude, latitude)* | *LUP type* |
|---|---|---|---|
| Alipur | L1 | 28°47′50.52″N, 77°7′59.29″E | Residential |
| Anand Vihar | L2 | 28.650278N, 77.302778E | Transport |
| Ashok Vihar | L3 | 28°41′19.24″N, 77°10′32.97″E | Residential |
| Aya Nagar | L4 | 28.468881N, 77.127188E | Residential |
| Bawana | L5 | 28°47′35.62″N, 77°2′54.01″E | Commercial |
| CRRI Mathura Road | L6 | 28.551003N, 77.273836E | Institutional |
| DTU | L7 | 28.749981N, 77.117236E | Institutional |
| Dr. Karni Singh Shooting Range | L8 | 28°29′59.07″N, 77°16′1.52″E | Residential |
| Dwarka-Sector | L9 | 28°34′19.34″N, 77°4′15.01″E | Residential |
| East Arjun Nagar | L10 | 28°39′22.22″N, 77°17′41.09″E | Residential |
| IGI Airport (T3) | L11 | 28.554927N, 77.0844E | Transport |
| IHBAS, Dilshad Garden | L12 | 28.68115N, 77.304583E | Institutional |
| ITO | L13 | 28.627521N, 77.243774E | Commercial |
| Jahangirpuri | L14 | 28°43′46.62″N, 77°9′59.87″E | Residential |
| Jawaharlal Nehru Stadium | L15 | 28°34′58.25″N, 77°14′3.72″E | Commercial |
| Lodhi Road | L16 | 28.59246N, 77.236729E | Commercial |
| Major Dhyan Chand National Stadium | L17 | 28°36′45.15″N, 77°14′14.40″E | Commercial |
| Mandir Marg | L18 | 28.634167N, 77.200556E | Commercial |
| Mundka | L19 | 28°40′56.33″N, 77°2′5.77″E | Residential |
| NSIT-Dwarka | L20 | 28.609088N, 77.03465E | Institutional |
| Najafgarh | L21 | 28°36′32.45″N, 76°59′7.63″E | Residential |
| Narela | L22 | 28°51′17.51″N, 77°5′21.15″E | Residential |
| Nehru Nagar | L23 | 28°34′14.79″N, 77°15′1.72″E | Residential |
| North Campus | L24 | 28.687613N, 77.21023E | Institutional |
| Okhla Phase-2 | L25 | 28°32′8.79″N, 77°16′35.14″E | Commercial |
| Patparganj | L26 | 28°38′5.03″N, 77°18′16.46″E | Commercial |
| Punjabi Bagh | L27 | 28.661944N, 77.124167E | Residential |
| Pusa | L28 | 28.63432N, 77.170913E | Institutional |
| R.K. Puram | L29 | 28.56611N1, 77.176667E | Residential |
| Rohini | L30 | 28°44′17.76″N, 77°4′55.97″E | Residential |
| Shadipur | L31 | 28.652533N, 77.155121E | Commercial |
| Sirifort | L32 | 28.54991N, 77.219907E | Residential |
| Sonia Vihar | L33 | 28°43′59.69″N, 77°14′58.52″E | Residential |
| Sri Aurobindo Marg | L34 | 28°33′22.72″N, 77°12′22.82″E | Commercial |
| Vivek Vihar | L35 | 28°40′16.48″N, 77°19′3.55″E | Residential |
| Wazirpur | L36 | 28°41′51.16″N, 77°9′37.58″E | Commercial |

AQMS, Air Quality Monitoring Stations; DTU, Delhi Technological University; LUP, land-use pattern.

Downloaded by 14.139.251.98 from www.liebertpub.com at 07/31/22. For personal use only.

2011, 2012). The study area is bordered by the peripheral districts of Gurgaon, Sonipat, Bahadurgarh, and Faridabad of Haryana and Ghaziabad, Noida, and Greater Noida districts of Uttar Pradesh state. A total number of 36 locations (Table 2) denoted from L1 to L36 have been selected within the NCT of Delhi for the present study (Supplementary Fig. SM1 in the Supplementary Data).

### *Materials and methods*

Base data. The base data that is, $PM_{2.5}$ and $PM_{10}$ concentrations, have been retrieved from the 36 Air Quality Monitoring Stations (AQMS) in NCT of Delhi for the entire period of study and data analysis (i.e., from January to May; 2019–2022). The AQMS located across Delhi are owned/operated by three agencies, viz., Central Pollution Control Board (CPCB), Indian Institute of Tropical Meteorology (IITM), and Delhi Pollution Control Committee (DPCC). The data recording, quality control and other aspects may be referred from Data Recording & Quality Control section in the Supplementary Data.

The location of the monitoring stations is representative of all four types of LUPs (i.e., commercial, residential, transport, and institutional) falling within urban areas of the city (Table 2). Twenty-four hours' average secondary data for all the parameters used in the study (namely, PM, and meteorological) were used in the present study corresponding to the lockdown, pre-, and post-lockdown (or extended) phases. The schedule of lockdown follows as in Table 2. The other particulars of the lockdown imposed in Delhi, India, can be had from the from Particulars & Schedule of Lock-Down in the Wake of COVID-19 section in the Supplementary Data.

The PM that is, $PM_{2.5}$ and $PM_{10}$ concentration levels were analyzed from January 2020 to May 2020 during the pre- and post-COVID-19 pandemic phases and additionally for the corresponding periods of year 2018, 2019, 2021, and 2022 for the trend analysis purposes (Table 3).

The spatial interpolation tools and technique. The GIS was used for the spatial dissemination of the point data recorded over the environment span (Cao *et al.*, 2007; Adams and Kanaroglou, 2016; Li *et al.*, 2018; Kumar *et al.*, 2020). The point locations were interpolated to generate thematic (spatial variability) maps that exhibit the trends of Spatio-temporal dynamics in relationship to the corrosion-scaling potential possible within the surface boundaries. Such dynamics was created using the spatial analyst function

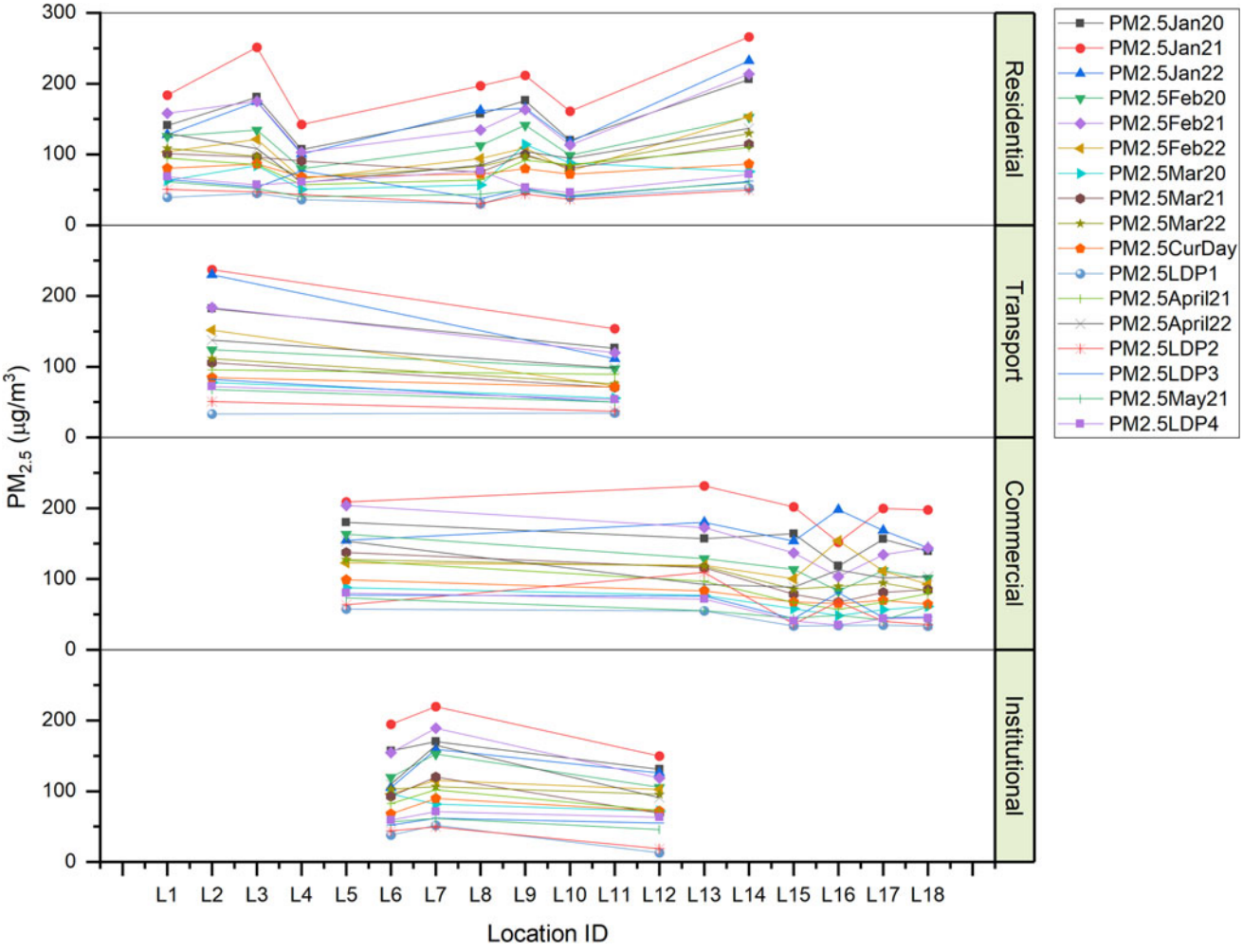



**FIG. 1.** Average $PM_{2.5}$ variations during pre- and lockdown phases and corresponding periods of 2021 and 2022 (L1–L18). PM, particulate matter.

Downloaded by 14.139.251.98 from www.liebertpub.com at 07/31/22. For personal use only.

Downloaded by 14.139.251.98 from www.liebertpub.com at 07/31/22. For personal use only.

available in the ArcGIS tool. The inverse distance weighting interpolation technique has been applied for estimating the values between the computed data points.

This technique encompasses the calculation of the weighted values for the unsampled point considering that it is inversely proportional to the squared distance between the observed point data and that of the unsampled location (Qu *et al.*, 2010; Burrough *et al.*, 2015; Jiang and Bai, 2018). This is to be noted that, however, the exactness of the surface interpolation generated using this technique is highly dependent upon the distribution, density, accuracy, and resolution of the input point dataset (Tyagi and Sarma, 2020).

## Results

### *Effect of different lockdown phases and corresponding periods on $PM_{2.5}$ concentration*

In January 2020, the maximum mass concentration range of $PM_{2.5}$ (over 200 $\mu g/m^3$) was found at three locations, namely, Nehru Nagar (L23: 219.65 $\mu g/m^3$), Jahangirpuri (L14: 206.15 $\mu g/m^3$), and Wazirpur (L36: 200.28 $\mu g/m^3$). Former two areas belonged to residential LUP, while the later to the commercial.

Further, in February 2020, highest monthly average concentrations (>150 $\mu g/m^3$) were recorded at six study locations, namely North campus (L24: 165.18 $\mu g/m^3$); Mundka (L19: 163.23 $\mu g/m^3$); Narela (L22: 159.69 $\mu g/m^3$); ITO (L13: 156.97 $\mu g/m^3$); Sirifort (L32: 152.77 $\mu g/m^3$), and Anand Vihar (L2: 152.58 $\mu g/m^3$). The lowest value range, that is, <100 $\mu g/m^3$ corresponded to five locations, viz., Rohini (L30), Punjabi Bagh (L27), Okhla Phase 2 (L25), IHBAS-DG (L12), and Patparganj (L26) as 79.80, 82.88, 97.34, 98.38, and 98.58 $\mu g/m^3$ correspondingly. All the locations, irrespective of LUPs, exceeded NAAQS for $PM_{2.5}$ that is, 60 $\mu g/m^3$ (Figs. 1 and 2).

In March 2020, 11 of 36 locations complied with NAAQS for $PM_{2.5}$ by recording average values below 60 $\mu g/m^3$. These were Shadipur (L31: 46.0 $\mu g/m^3$), Lodhi Road (L16: 48.1 $\mu g/m^3$), Sri Aurobindo Marg (L34: 49.9 $\mu g/m^3$), Aya Nagar (L4: 50.5 $\mu g/m^3$), R.K. Puram (L29: 51 $\mu g/m^3$), IGI Airport-Terminal 3 (L11: 55.5 $\mu g/m^3$), Patparganj (L26: 56.4 $\mu g/m^3$), Pusa (L28: 56.5 $\mu g/m^3$), Major Major Dhyan Chand National Stadium (MDCNS, L17: 56.5 $\mu g/m^3$), Dr. Karni Singh Shooting Range (DKSSR, L8: 56.9 $\mu g/m^3$), and Jawaharlal Nehru Stadium (JNS, L15: 58.1 56.5 $\mu g/m^3$). All remaining 25 locations (majority of residential and institutional LUP areas) exceeded the NAAQS with

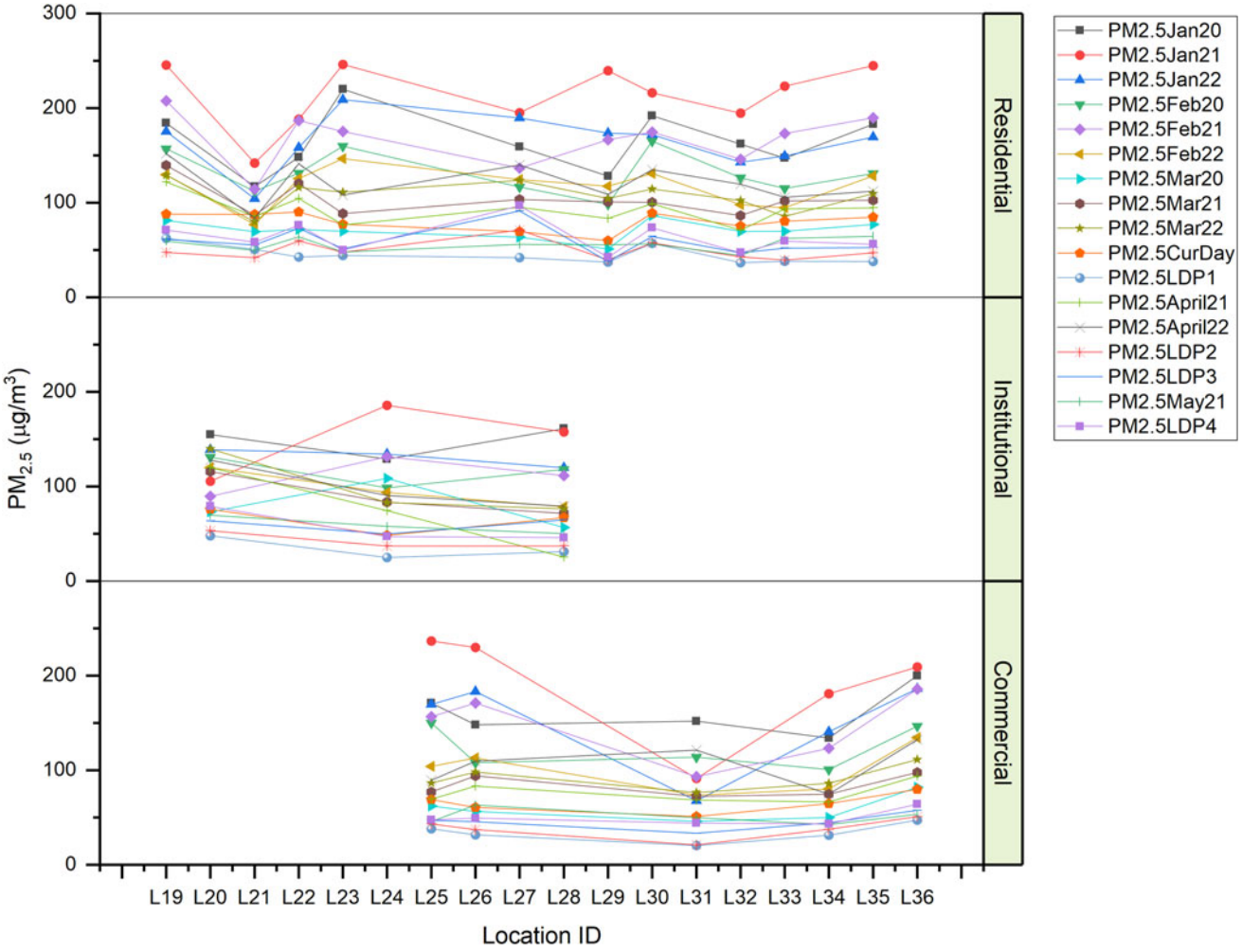



**FIG. 2.** Average $PM_{2.5}$ variations during pre- and lockdown phases and corresponding periods of 2021 and 2022 (L19–L36).

Dwarka-sector and North Campus recording the highest values over 100 $\mu g/m^3$, that is, 114.07 and 108.60 $\mu g/m^3$, respectively.

During 14 h public curfew on March 22, 2020, only three of the locations complied with the prescribed annual average $PM_{2.5}$ concentration of 60 $\mu g/m^3$, that is, R.K. Puram (L29: 59.71 $\mu g/m^3$), Shadipur (L31: 50.99 $\mu g/m^3$), and North Campus (L24: 48.32 $\mu g/m^3$). All remaining 33 locations violated such norms, majority being a mix of residential and institutional LUP areas (Fig. 3). During LDP1, interestingly, 35 locations conformed to $PM_{2.5}$ concentration of 60 $\mu g/m^3$ (from L1 to L18 and L20 to L36) with only Mundka (L19, residential LUP) recording the highest average concentrations as 62.31 $\mu g/m^3$.

During LDP2, except 4 locations, namely ITO (L13: 109.21 $\mu g/m^3$), Punjabi Bagh (L27: 71.07 $\mu g/m^3$), Lodhi Road (L16: 68.20 $\mu g/m^3$), and Bawana (L5: 63.31 $\mu g/m^3$), most of them representing commercial LUP, seeing a violation of the prescribed standard, all other 32 locations complied to it by having values lower than 60 $\mu g/m^3$. Further, DKSSR (L8: 30.45 $\mu g/m^3$), Shadipur (L31: 21.13 $\mu g/m^3$), and IHBAS-DG (L12: 18.78 $\mu g/m^3$) witnessed the lowest concentrations.

During LDP3, as many as 14 locations recorded values above 60 $\mu g/m^3$ violating the NAAQS norm, such as Punjabi Bagh (L27: 91.19 $\mu g/m^3$). Anand Vihar (L2: 81.99 $\mu g/m^3$), Lodhi Road (L16: 80.76 $\mu g/m^3$), Bawana (L5: 77.13 $\mu g/m^3$), Aya Nagar (L4: 76.77 $\mu g/m^3$), ITO (L13: 75.82 $\mu g/m^3$), Narela (L22: 72.94 $\mu g/m^3$), Pusa (L28: 64.67 $\mu g/m^3$) Alipur (L1: 64.51 $\mu g/m^3$), Rohini (L30: 64.33 $\mu g/m^3$), NSIT-Dwarka (L20: 63.59 $\mu g/m^3$), Delhi Technological University (DTU; L7: 61.75 $\mu g/m^3$), Jahangirpuri (L14: 61.66 $\mu g/m^3$), and Mundka (L19: 61.26 $\mu g/m^3$). All remaining 22 locations were found in conformance to the norm having recorded values below 60 $\mu g/m^3$. R.K. Puram (L29), DKSSR (L8) and Shadipur (L31) recorded the lowest values as 37.82, 37.44, and 33.44 $\mu g/m^3$, respectively.

LDP4 saw as 15 locations recording average concentrations >60 $\mu g/m^3$, thus flouting NAAQS. These were Punjabi Bagh (L27: 96.71 $\mu g/m^3$), Bawana (L5: 80.26 $\mu g/m^3$), NSIT-Dwarka (L20: 78.96 $\mu g/m^3$), DKSSR (L8: 76.73 $\mu g/m^3$), Narela (L22: 76.10 $\mu g/m^3$), Rohini (L3: 73.36 $\mu g/m^3$), Anand Vihar (L2; 71.94 $\mu g/m^3$), Jahangirpuri (L14: 71.76 $\mu g/m^3$), ITO (L13: 71.38 $\mu g/m^3$), Mundka (L19: 71.04 $\mu g/m^3$), DTU (L7: 70.99 $\mu g/m^3$), Alipur (L1: 68.43 $\mu g/m^3$), Wazirpur (L36: 63.99 $\mu g/m^3$), IHBAS-DG (L12: 62.89 $\mu g/m^3$), and Aya Nagar (L4: 60.99 $\mu g/m^3$).

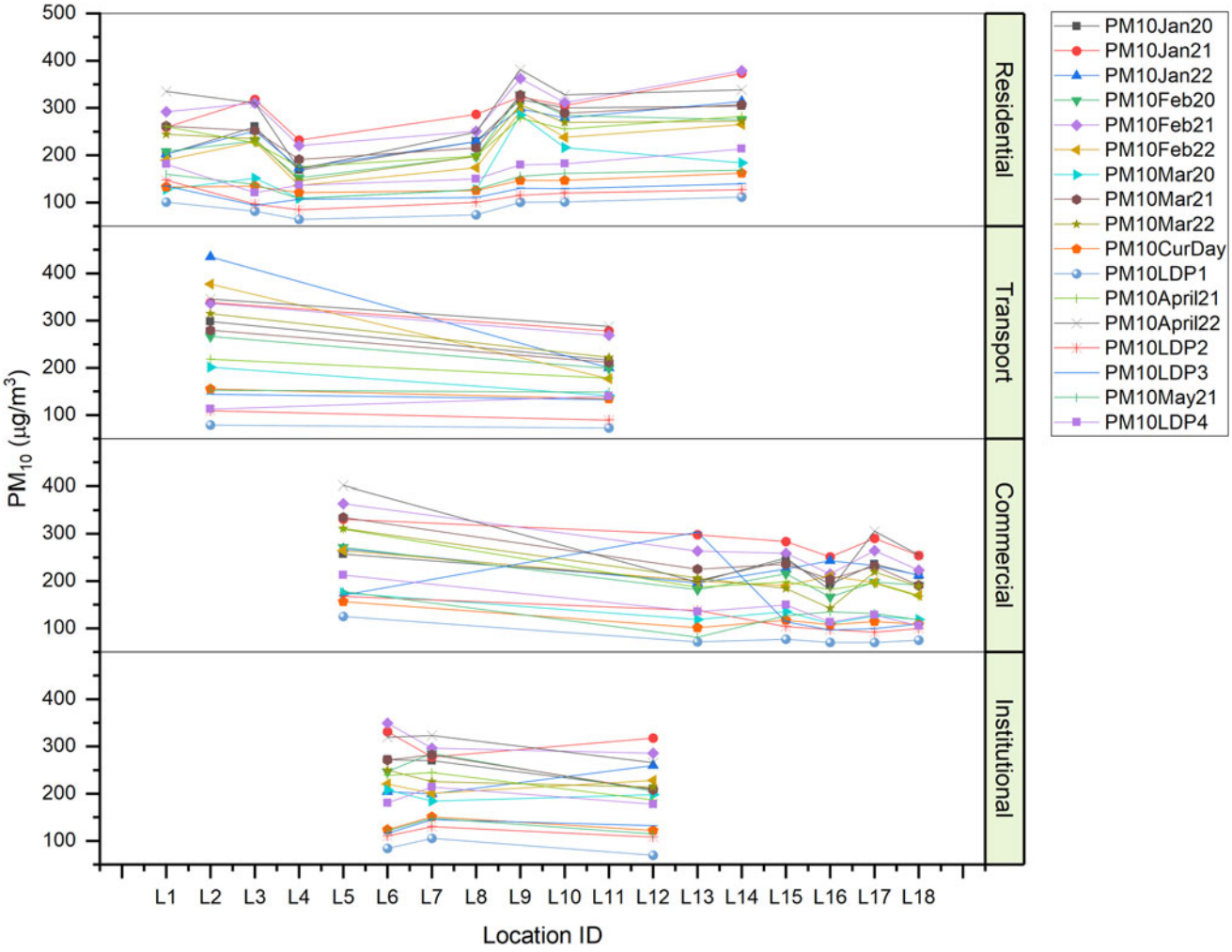



**FIG. 3.** Average $PM_{10}$ variations during pre- and lockdown phases and corresponding periods of 2021 and 2022 (L1–L18).

Downloaded by 14.139.251.98 from www.liebertpub.com at 07/31/22. For personal use only.

Downloaded by 14.139.251.98 from www.liebertpub.com at 07/31/22. For personal use only.

*Effect of different lockdown phases and corresponding periods on $PM_{10}$ concentration*

During January 2020, all locations violated NAAQS norm of 100 $\mu g/m^3$ (24 h average) prescribed for $PM_{10}$. Four locations even recorded more than three times of NAAQS, that is, 317.73, 303.55, and 301.57 $\mu g/m^3$ (L9, L14, and L36, respectively). Values >150 and <200 $\mu g/m^3$ were reported at least four locations, namely L16 (192.89 $\mu g/m^3$), L34 (169.26 $\mu g/m^3$), L4 (167.91 $\mu g/m^3$), and L21 (158.1 $\mu g/m^3$), respectively (Fig. 3). In February 2020, all the selected locations were flouted the corresponding air quality standard (100 $\mu g/m^3$) as set by CPCB. Twenty-three locations were recorded double the prescribed norm and up to 328.43 $\mu g/m^3$ (recorded at L9), whereas 13 locations witnessed values between 150 and 200 $\mu g/m^3$ (L34, L4, L16, L21, L13, L26, L18, L24, L27, L28, L8, L17, and L11 in descending order) (Fig. 3). A variation in effect of LUP on $PM_{10}$ from that on $PM_{2.5}$ was noted.

During March 2020 too, all the locations recorded values greater than the prescribed norm of $PM_{10}$ except location L34, which reported an average monthly concentration of 90.7 $\mu g/m^3$. More than 15 locations depicted concentration >150 and <200 $\mu g/m^3$ (L9, L24, L6, L36, L2, L32, L19, L7, L14, L5, L30, L22, L35, L25, and L3 in reducing order).

Similar to March 2020 data, during 14 h curfew period, the average concentration of $PM_{10}$ at all the locations exceeded the NAAQS norms expect location L34, which reported the concentration of 93.44 $\mu g/m^3$. However, only 5 of 36 locations recorded concentration >150 and <200 $\mu g/m^3$ (L22 > L14 > L5 > L2 > L7), (Fig. 4).

During LDP1, 23 locations conformed to $PM_{10}$ concentration of 100 $\mu g/m^3$ (L34, L4, L28, L26, L17, L16, L13, L11, L8, L24, L18, L32, L15, L2, L25, L33, L29, L3, L23, L6, L35, L27, and L36 in ascending order), whereas remaining locations did not comply to the norm. During LDP2, only 15 locations complied with the prescribed $PM_{10}$ norm, whereas remaining locations did not. The highest recorded value increased to LDP1 (147.15 $\mu g/m^3$ at L21 vis-à-vis 212.1 $\mu g/m^3$ at L24). The compliance scenario worsened during LDP3 with only 13 locations complying with $PM_{10}$ norm, whereas 23 locations remained noncompliant. Location L13 recorded maximum concentration at 303.48 $\mu g/m^3$ about three times the standard.

The LDP4 witnessed worst compliance to the prescribed $PM_{10}$ standard as none of the locations complied with it. As many as six locations even showed just double concentration values compared to the norm (L19, L22, L7, L14, L5, and L27).

The positive impact on ambient average $PM_S$ concentrations during lockdown phases are attributable to its strict

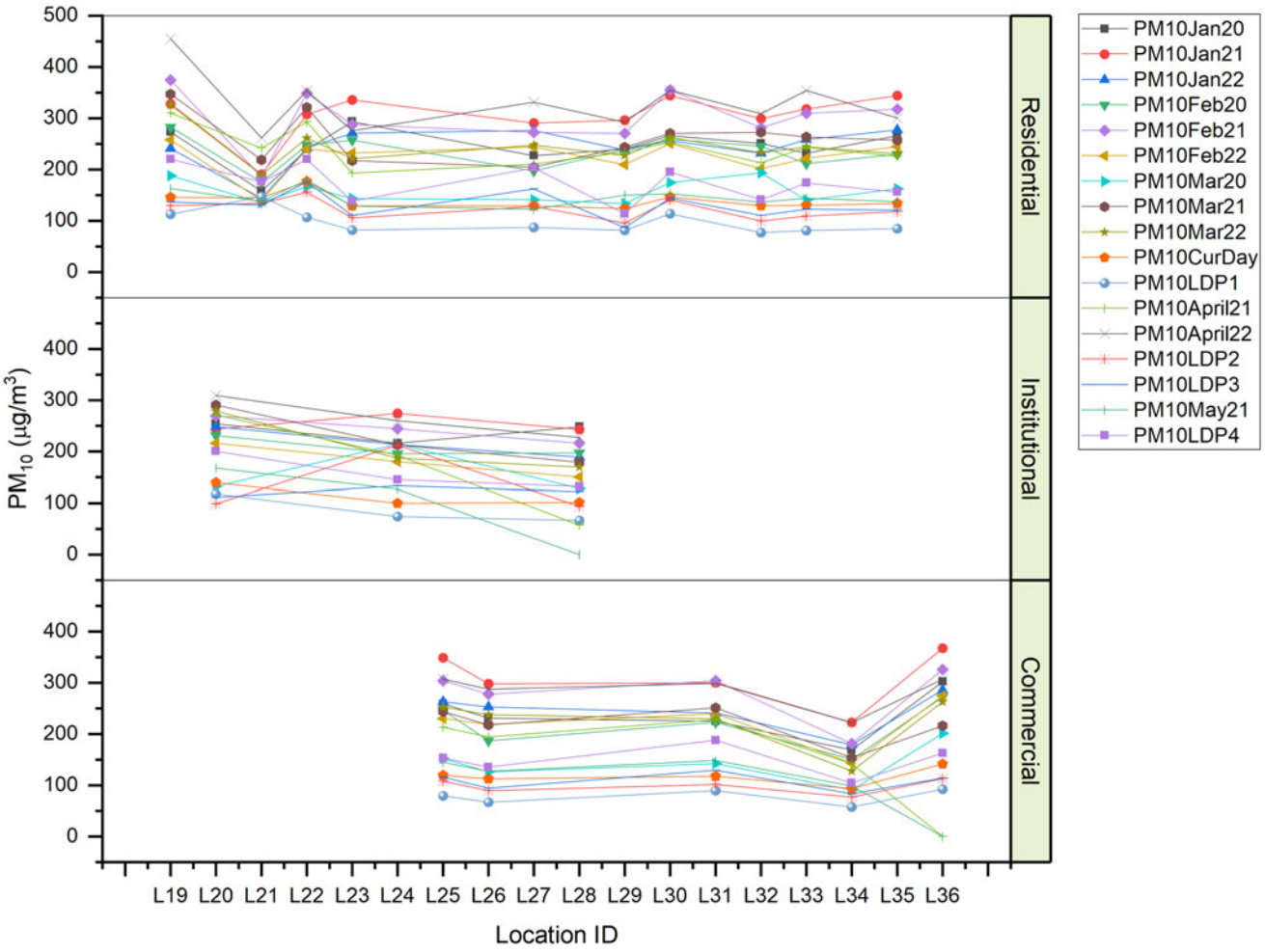



**FIG. 4.** Average $PM_{10}$ variations during pre- and lockdown phases and corresponding periods of 2021 and 2022 (L19–L36).

implementation by the Government (Central and State both) further supported by the citizens of Delhi. Only essential services and people needing medical attention along with those attending them were allowed during LDPs. Police, other enforcement and monitoring agencies together with certain Non-Governmental Organizations (NGOs) took up to the task and helped the cause.

The average reduction in $PM_S$ over corresponding periods from four different years (2018, 2019, 2021, and 2022) was also looked into vis-à-vis study locations and the period of 2020 in-question. The result of such analysis is depicted in Supplementary Figs. SM2 and SM3 in the Supplementary Data (for $PM_{2.5}$ and $PM_{10}$, respectively). PM ($PM_{2.5}$:$PM_{10}$) ratios were also examined at all the locations with a view to understand the proportion of $PM_{2.5}$ and $PM_{10}$. As the higher ratios (i.e., higher proportion of $PM_{2.5}$ in $PM_{10}$ or total PM) are considered to be indicator of presence of natural aerosols in total PM in the ambient air contributed by various natural phenomenon and the lower ratios are indicative of lesser proportion of $PM_{2.5}$ in total PM, that is, lesser degree of aerosols sourced from anthropogenic causes, such as vehicular exhaust, construction activities and so on (Zhao *et al.*, 2019; Biswas *et al.*, 2020). The location-wise PM ratios are presented in Supplementary Figs. SM4 and SM5 in the Supplementary Data through the years 2018–2022.

*Impact of different lockdown phases on spatial distribution of $PM_{2.5}$ and $PM_{10}$*

The GIS mapping of monthly average $PM_S$ during January, February, March 2020, curfew days, and LDP1, 2, 3, and 4 were also undertaken in the present study (as part of supplemental study) and spatial distribution results are presented in Figs. 5 and 6. The pollution level is assigned as five-class distribution namely low, moderately low, moderate, moderately high, and high. This is done so, so as to provide more scope in depicting the range of results in five classes. This methodology has also been adopted in previous studies (Followed by: Gupta and Sarma, 2016; Kumar *et al.*, 2020).

Average $PM_{2.5}$ spatial distribution over NCT of Delhi in various class range show low (79.80–96.68 $\mu g/m^3$), moderately low (96.96–113.95 $\mu g/m^3$), moderate (113.96–131.03 $\mu g/m^3$), moderately high (131.04–148.10 $\mu g/m^3$), and high (148.11–165.18 $\mu g/m^3$) (Table 4). On the contrary, the $PM_{10}$ concentration range is represented as low (150–177.48 $\mu g/m^3$), moderately low (177.49–204.43 $\mu g/m^3$), moderate (204.44–231.37 $\mu g/m^3$), moderately high (231.38–258.32 $\mu g/m^3$), and high (258.33–285.26 $\mu g/m^3$) (Table 5).

The January and February 2020 (prelockdown period) showed a higher value of $PM_{2.5}$ and $PM_{10}$ as all the activities

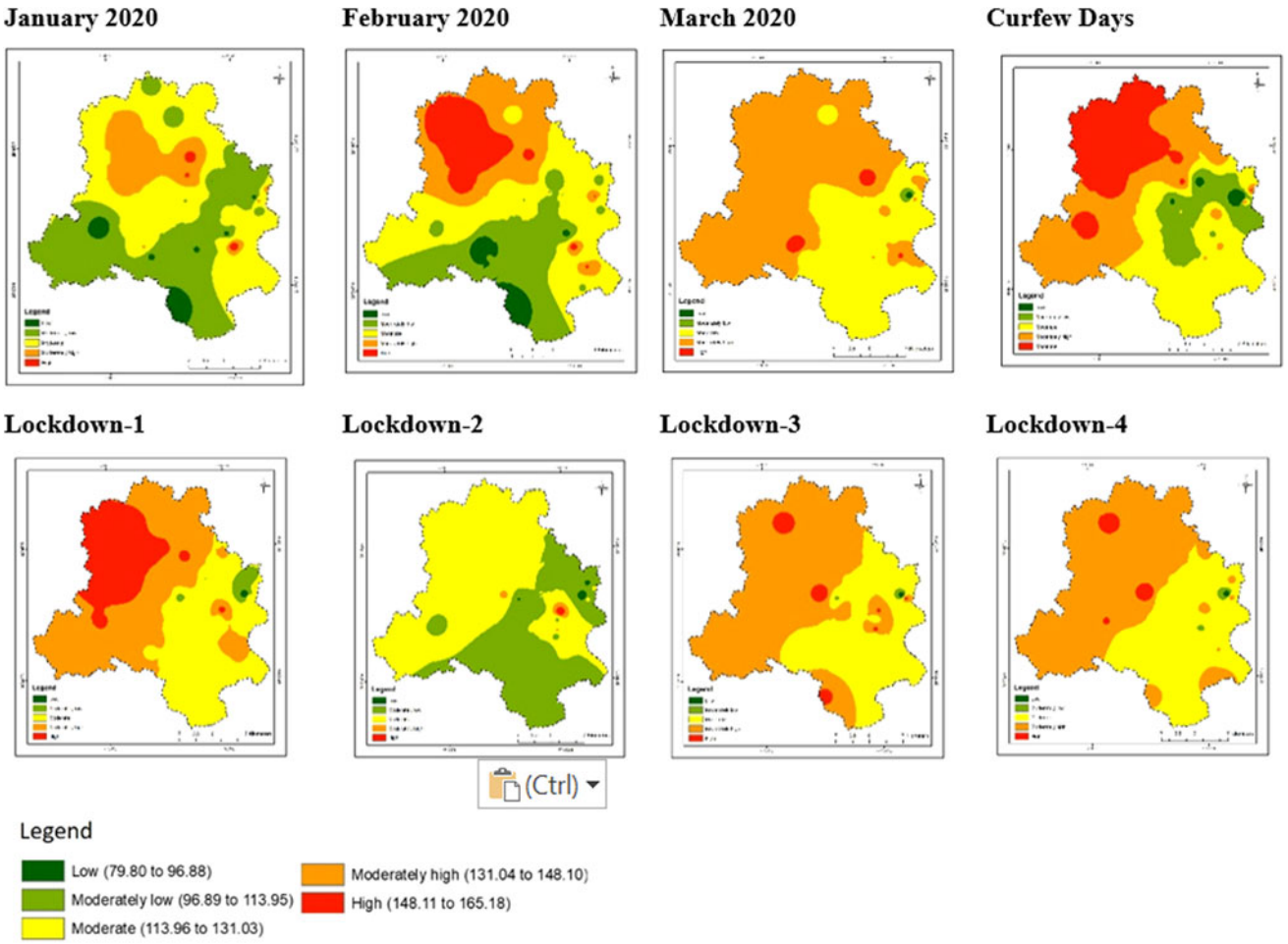



**FIG. 5.** The spatial distribution of $PM_{2.5}$ over Delhi city.

Downloaded by 14.139.251.98 from www.liebertpub.com at 07/31/22. For personal use only.

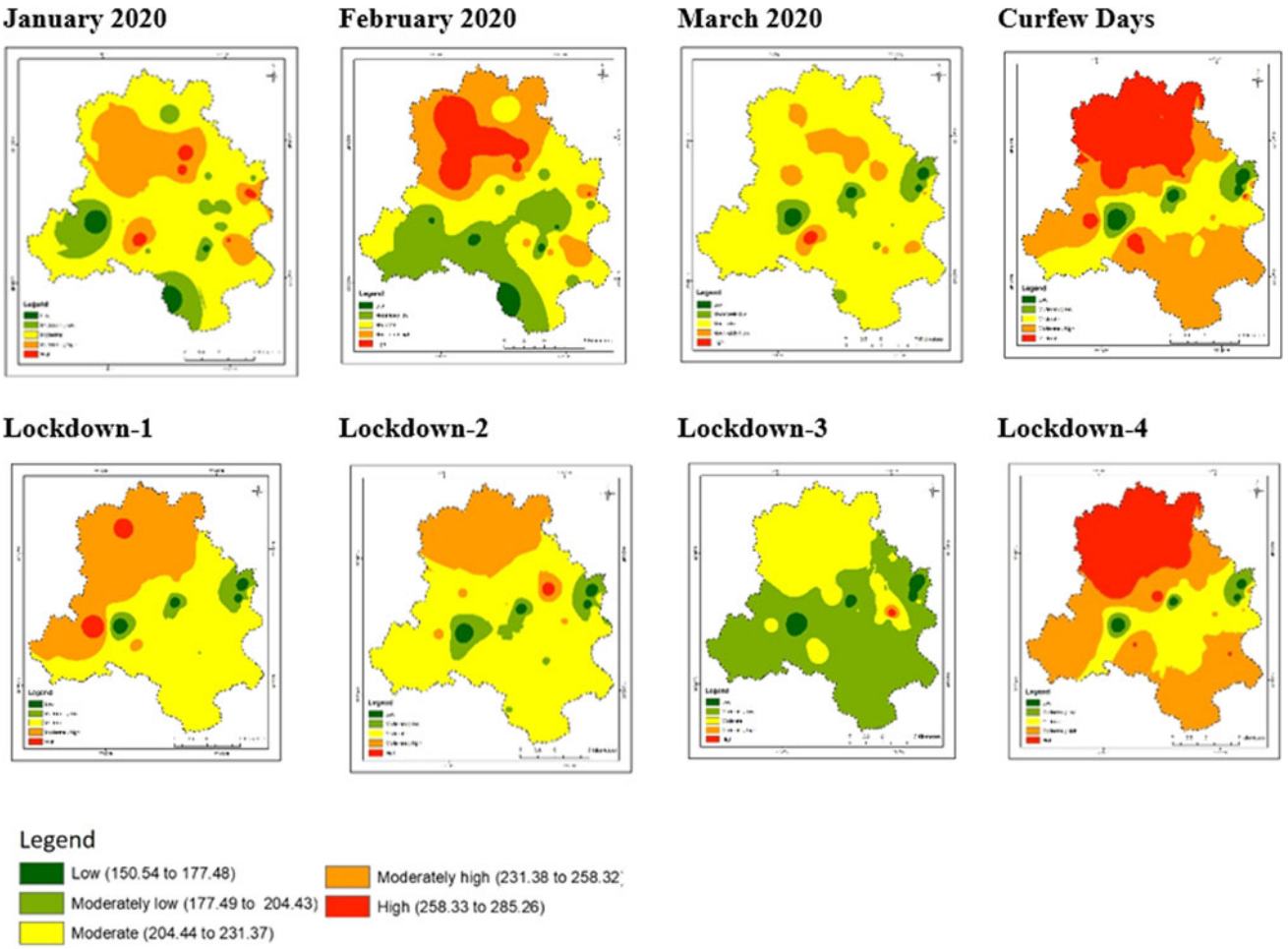



**FIG. 6.** The spatial distribution of $PM_{10}$ over Delhi city.

of industry, transport, commerce, and others were at normal paces. The maximum concentration level was observed in northwest Delhi. This is because of the presence of high vehicular and industrial activities in that area. There was a drastic reduction in the values of $PM_S$ in parts of central, east, south, and southwest Delhi because of restrictions during the lockdowns. The curfew day also showed a lower level of pollution than the prelockdown phase (Figs. 5 and 6).

Areas under these moderately high ranges prolonged during the early lockdown phase-1 period and gradually

TABLE 4. AREA IN SQUARE KILOMETER UNDER EACH DISTRIBUTION CLASS OF $PM_{2.5}$ VALUES IN DELHI

| *Class* | *Area in sq. km (%) January 2020* | *Area in sq. km (%) February 2020* | *Area in sq. km (%) March 2020* | *Area in sq. km (%) Curfew days* |
|---|---|---|---|---|
| Low | 45.53 (3.07) | 56.51 (3.81) | 5.55 (0.37) | 12.35 (0.83) |
| Moderately low | 592.57 (39.96) | 390.59 (26.34) | 335.63 (22.63) | 184.58 (12.45) |
| Moderate | 630.96 (42.55) | 522.50 (35.23) | 215.30 (14.52) | 452.29 (30.50) |
| Moderately high | 207.06 (13.96) | 320.83 (21.63) | 585.50 (39.48) | 489.32 (33.00) |
| High | 6.87 (0.46) | 192.57 (12.99) | 341.02 (23.00) | 344.46 (23.23) |
| | 1,483.00 (100) | 1,483.00 (100) | 1,483.00 (100) | 1,483.00 (100) |
| | *Lockdown 1* | *Lockdown 2* | *Lockdown 3* | *Lockdown 4* |
| Low | 1.87 (0.13) | 4.13 (0.28) | 0.92 (0.06) | 1.14 (0.08) |
| Moderately low | 28.93 (1.95) | 564.69 (38.08) | 4.13 (0.28) | 6.53 (0.44) |
| Moderate | 510.96 (34.45) | 902.73 (60.87) | 529.87 (35.73) | 583.69 (39.36) |
| Moderately high | 638.99 (43.09) | 9.04 (0.61) | 912.74 (61.55) | 863.17 (58.20) |
| High | 302.25 (20.38) | 2.40 (0.16) | 35.34 (2.38) | 28.46 (1.92) |
| | 1,483.00 (100) | 1,483.00 (100) | 1,483.00 (100) | 1,483.00 (100) |

PM, particulate matter.

Downloaded by 14.139.251.98 from www.liebertpub.com at 07/31/22. For personal use only.

Table 5. Area in Square Kilometer Under Each Distribution Class of $PM_{10}$ Values in Delhi

| Class | *Area in sq. km (%) January 2020* | *Area in sq. km (%) February 2020* | *Area in sq. km (%) March 2020* | *Area in sq. km (%) Curfew days* |
|---|---|---|---|---|
| Low | 34.59 (2.33) | 32.02 (2.16) | 46.94 (3.16) | 29.93 (2.02) |
| Moderately low | 182.22 (12.29) | 405.09 (27.32) | 204.09 (13.76) | 70.54 (4.76) |
| Moderate | 894.21 (60.30) | 517.24 (34.88) | 621.25 (41.89) | 331.31 (22.34) |
| Moderately high | 351.06 (23.67) | 356.86 (24.06) | 587.08 (39.59) | 622.65 (41.99) |
| High | 20.92 (1.41) | 171.80 (11.58) | 23.65 (1.59) | 428.58 (28.90) |
| | 1,483.00 (100) | 1,483.00 (100) | 1,483.00 (100) | 1,483.00 (100) |
| | *Lockdown 1* | *Lockdown 2* | *Lockdown 3* | *Lockdown 4* |
| Low | 19.66 (1.33) | 23.27 (1.57) | 36.85 (2.48) | 15.59 (1.05) |
| Moderately low | 73.58 (4.96) | 111.08 (7.49) | 900.61 (60.73) | 45.80 (3.09) |
| Moderate | 771.08 (51.99) | 992.90 (66.95) | 537.80 (36.26) | 343.06 (23.13) |
| Moderately high | 586.45 (39.55) | 350.28 (23.62) | 5.53 (0.37) | 652.52 (44.00) |
| High | 32.24 (2.17) | 5.47 (0.37) | 2.20 (0.15) | 426.04 (28.73) |
| | 1,483.00 (100) | 1,483.00 (100) | 1,483.00 (100) | 1,483.00 (100) |

started declining. In this phase, almost all the activities were strictly enforced to close down with strong vigilance from the Government. The outcome of these enforcements was observed during the lockdown phase-2 where the area under $PM_{2.5}$ was drastically reduced. Relaxations in many activities were given during the lockdown phase-2, and as a result, the area under $PM_{2.5}$ during lockdown-3 again increased in alarming rate. Similar was the case for lockdown-4. In addition, the high pollution level in early March 2019, just before the lockdown could possibly have kept the monthly mean values of March 2020 in few cities unaltered amid a blanket restriction over traffic flow, industries, and other businesses, and industries. A quick decline in pollutant levels in such cities due to lockdown is more peculiar (Shrestha *et al.*, 2020).

### *Meteorological aspects*

The meteorological parameters during the study period were also collected as secondary data from India Meteorological Department (IMD) web portal and analyzed. Four variables were studied, that is, atmospheric temperature (AT, unit °C), solar radiation (SR, unit $W/m^2$), wind direction (WD, unit °), and wind speed (WS, unit m/s). All the parameters used in the study correspond to 24 h average values over the study period, which spanned from January 1 to May 31, 2020 and pre- and postlockdown periods between 2019 and 2022. The meteorological parameters were plotted against the average PM values against all the study locations to assess any significant variation over the study locations (Supplementary Figs. SM6–SM13 in the Supplementary Data).

## Concluding Remarks

Although the COVID-19 lockdown in India and elsewhere in the world never had a goal of ambient air quality improvement as such, interestingly, more than 50% reduction in ambient PM (both $PM_{10}$ and $PM_{2.5}$) was recorded during lockdown 1 and 2 compared to lockdown 3 and 4. Lockdown 1 showed the maximum reduction in PM while the lockdown 4 depicted relatively the lowest reduction among the total lockdown period. During the lockdown, $PM_{10}$ and $PM_{2.5}$ levels were found well below the NAAQS (Singh *et al.*, 2020). The main reason for these conspicuous positive changes in initial lockdown phases is a strict restriction on the anthropogenic activities giving rise to PM concentrations, primarily, vehicular and construction (Roy and Balling, 2021).

On the contrary, declination in such reduction may be attributed to the relaxations provided to the public in between and during the ending phases of lockdowns. Although we assessed that the meteorological conditions within the city boundaries have been largely flat during the study period, their contribution to such improved air quality cannot be denied (Singh *et al.*, 2020). Nevertheless, such reductions in PM levels were short-lived because the levels went up again in the years 2021 and 2022 (except similar lockdowns) as the situation returned to normalcy postlockdown. (Vasudevan *et al.*, 2021). In other words, COVID-19 had a significant effect on the air quality during full lockdown implemented for few months and afterward, however, there was unprecedented growth of poor air quality, which has a seasonal effect compared to previous years (Tiwari *et al.*, 2021).

In the purview of air quality, the effectiveness and cost are always top factors for policymakers to decide upon preferred measures for abating air pollution. This study found that lockdown shows a substantial positive change in the air quality of NCT of Delhi. Interestingly, the resident of the Delhi city witnessed a rare blue sky after decades during the lockdown period. It is, therefore, for the immediate benefit toward urban quality that Government imposes lockdown in case of severe ambient air pollution in a densely populated megacity like Delhi; it still remains a temporary or quick-fix solution, to be looked as a last line of defense.

## Acknowledgments

The authors thank the CPCB, DPCC, Amity University Haryana (AUH), and DTU for the extended support during research activity and secondary data sources for completed research objectives of this article.

Downloaded by 14.139.251.98 from www.liebertpub.com at 07/31/22. For personal use only.

### Authors' Contributions

Conceptualization and methodology development: A.K.; formal analysis and investigation: A.K., K.S., and A.P.; writing—original draft preparation: A.K. and A.P.; writing—review and editing: A.P.; supervision: critical inputs: R.K.M. and A.P.; approval of article: A.K., K.S., A.P., R.K.M., and P.C.S.D.

### Author Disclosure Statement

The authors declare that they have no conflict of interest or competing interest to declare.

### Funding Information

The present research did not receive any funding or funding is not applicable.

### Supplementary Material

Supplementary Data

Downloaded by 14.139.251.98 from www.liebertpub.com at 07/31/22. For personal use only.

### References

Adams, M.D., and Kanaroglou, P.S. (2016). Mapping real-time air pollution health risk for environmental management: Combining mobile and stationary air pollution monitoring with neural network models. *J. Environ. Manage*. 168, 133.

Adar, S., Sheppard, L., Vedal, S., Polak, J.F., Sampson, P.D., Diez Roux, A.V., Budoff, M., Jacobs, D.R., Jr., Graham Barr, R., Watson, K., and Kaufman, J.D. (2013). Fine particulate air pollution and the progression of carotid intima-medial thickness: A prospective cohort study from the multi-ethnic study of atherosclerosis and air pollution. *PLoS Med.* 10, 1.

Agarwal, A., Kaushik, A., Kumar, S., and Mishra, R.K. (2020). Comparative study on air quality status in Indian and Chinese cities before and during the COVID-19 lockdown period. *Air Qual. Atmos. Health* 13, 1.

Biswas, K., Chatterjee, A., and Chakraborty, J. (2020). Amplified ozone pollution in cities during the COVID-19 lockdown. *J. Geovis. Spat. Anal.* 4, 25.

Briz-Redón, A., Belenguer-Sapiña, C., and Serrano-Aroca, A. (2020). Changes in air pollution during COVID-19 lockdown in Spain: A multi-city study. *J. Environ. Sci. (China)* 101, 16.

Burrough, P.A., McDonnell, R.A., and Lloyd, C.D. (2015). *Principles of Geographical Information Systems*, 3rd ed. UK: Oxford University Press, 352 p.

Cao, J., Lee, S., Chow, J., Watson, J.G., Ho, K.F., Zhang, R.J., Jin, Z.D., Shen, Z.X., Chen, G.C., Kang, Y.M., Zou, S.C., Zhang, L.Z., Qi, S.H., Dai, M.H., Cheng, Y., and Hu, K. (2007). Spatial and seasonal distributions of carbonaceous aerosols over China. *J. Geophy. Res.* 112, D22.

Central Pollution Control Board (CPCB). (2011). Air quality monitoring emission inventory and source apportionment study for Indian cities: National summary report. Delhi: Central Pollution Control Board.

Central Pollution Control Board (CPCB). (2012). National ambient air quality status & trends in India-2010 (No. NAAQMS/35/2011–2012). Delhi: Central Pollution Control Board, Ministry of Environment, Forest and Climate Change (MoEF&CC).

Chu, B., Zhang, S., Liu, J., Ma, Q., and He, H. (2020). Significant concurrent decrease in $PM_{2.5}$ and $NO_2$ concentrations in China during COVID-19 epidemic. *J. Environ. Sci. (China)* 99, 346.

Dandona, L., Dandona, R., Kumar G.A., *et al.* (2017). Nations within a nation: Variations in epidemiological transition across the states of India 1990–2016 in the global burden of disease study. *Lancet.* 390, 2437. DOI: 10.1016/S0140-6736(17)32804-0.

Devara, P.C.S., Kumar, A., Sharma, P.B., Banerjee, P., Khan, A.A., Tripathi, A., Tiwari, S., and Beig, G. (2020). Influence of air pollution on coronavirus (COVID-19): Some evidences from studies at AUH. Gurugram: SSRN.

Dutheil, F., Baker, S.J., and Navel, V. (2020). COVID-19 as a factor influencing air pollution? *Environ. Pollut.* 263, 114.

Fann, N., Lamson, A.D., Anenberg, S.C., Wesson, K., Risley, D., and Hubbell, B.J. (2012). Estimating the national public health burden associated with exposure to ambient $PM_{2.5}$ and ozone. *Risk Anal.* 32, 81.

Filonchyk, M., and Peterson, M. (2020). Air quality changes in Shanghai, China, and the surrounding urban agglomeration during the COVID-19 lockdown. *J. Geovis. Spat. Anal.* 4, 22.

Gautam, S., and Hens, L. (2020). COVID-19: Impact by and on the environment, health and economy. *Environ. Dev. Sustain*. 22, 4953.

Gupta, P., and Sarma, K. (2016). Spatial distribution of various parameters in groundwater of Delhi, India. *Cogent Eng*. 3, 1.

Jain, S., and Sharma, T. (2020). Social and travel lockdown impact considering coronavirus disease (COVID-19) on air quality in megacities of India: Present benefits, future challenges and way forward. *Aerosol Air Qual. Res.* 20, 1222.

Jiang, L., and Bai, L. (2018). Spatio-temporal characteristics of urban air pollutions and their causal relationships: Evidence from Beijing and its neighboring cities. *Sci. Rep.* 8, 1279.

Kumar, A., Mishra, R.K., and Sarma, K. (2020). Mapping spatial distribution of traffic induced criteria pollutants and associated health risks using kriging interpolation tool in Delhi. *J. Transp. Health* 18, 100879.

Li, J., Han, X., Li, X., Yang, J., and Li, X. (2018). Spatiotemporal patterns of ground monitored $PM_{2.5}$ concentrations in China in recent years. *Int. J. Environ. Res. Public Health* 15, 114.

Mahato, S., Pal, S., and Ghosh, K.G. (2020). Effect of lockdown amid COVID-19 pandemic on air quality of the megacity Delhi, India. *Sci. Total Environ.* 730, 139086.

Mishra, R.K., Pandey, A., Pandey, G., and Kumar, A. (2019). The effect of odd-even driving scheme on $PM_{2.5}$ and $PM_{1.0}$ emission. *Transp. Res. D Trans. Environ*. 67, 541.

Mohan, M., and Kandya, A. (2007). An analysis of the annual and seasonal trends of air quality index of Delhi. *Environ. Monit. Assess*. 131, 267.

Qu, W., Arimoto, R., Zhang, X., Zhao, C.H., Wang, Y.Q., Sheng, L.F., and Fu, G. (2010). Spatial distribution and inter annual variation of surface $PM_{10}$ concentrations over eighty-six Chinese cities. *Atmos. Chem. Phys.* 10, 5641.

Roy, S.S., and Balling, R.C. (2021). Impact of the COVID-19 lockdown on air quality in the Delhi Metropolitan Region. *Appl. Geogr.* 128, 102418.

Selvi, S.M., Ravikumar, K., Rajendran, A.D., Bagavathi, A.B., Narayanan, N., and Mangottiri, V. (2020). Assessment of air quality index in major cities of India–lessons from lockdown. *IOP Conf. Ser. Mater. Sci. Eng.* 955, 012079.

Shrestha, A.M., Shrestha, U.B., Sharma, R., Bhattarai, S., Tran, H.N.T., and Rupakheti, M. (2020). *Lockdown Caused by COVID-19 Pandemic Reduces Air Pollution in Cities Worldwide.* Available at: https://eartharxiv.org/repository/view/304 (accessed December 10, 2021).

Sicard, P., De Marco, A., Agathokleous, E., Feng, Z., Xu, X., Paoletti, E., Rodriguez, J.J.D., and Calatayud, V. (2020). Amplified ozone pollution in cities during the COVID-19 lockdown. *Sci. Total Environ.* 735, 139542.

Singh, V., Singh, S., Biswal, A., Kesarkar, A.P., Mor, S., and Ravindra, K. (2020). Diurnal and temporal changes in air pollution during COVID-19 strict lockdown over different regions of India. *Environ. Pollut.* 266, 115368.

Tiwari, A., Gupta, R., and Chandra, R. (2021). Delhi air quality prediction using LSTM deep learning models with a focus on COVID-19 lockdown. arxiv.2102.10551

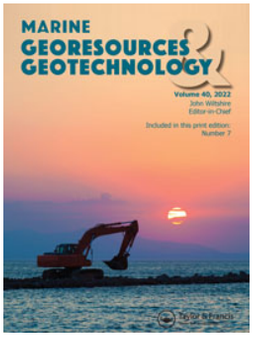
Tyagi, S., and Sarma, K. (2020). Qualitative assessment, geochemical characterization and corrosion-scaling potential of groundwater resources in Ghaziabad district of Uttar Pradesh, India. *Groundw. Sustain. Dev.* 10, 100370.

Vasudevan, M, Narayanan, N., Selvi, S.M., Ravikumar, K., Rajendran, A.D., and Bagavathi, A.B. (2021). Correlating the trends of COVID-19 spread and air quality during lockdowns in Tier-I and Tier-II cities of India—Lessons learnt and futuristic strategies. *Environ. Sci. Pollut. Res. Int.* Available at https://doi: 10.1007/s11356-021-16028-1.

World Health Organization (WHO). (2021). *Coronavirus Disease (Covid-2019) Situation Reports—173 (WHO, 2020)*. Data as received by WHO from national authorities by 10:00 CEST. Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200711-covid-19-sitrep-173.pdf (accessed July 11, 2020).

Zhao, D., Chen, H., Yu, E., and Luo, T. (2019). $PM_{2.5}/PM_{10}$ ratios in eight economic regions and their relationship with meteorology in China. *Adv. Meteorol.* 2019, 5295726.

Downloaded by 14.139.251.98 from www.liebertpub.com at 07/31/22. For personal use only.



Marine Georesources & Geotechnology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/umgt20

# Cyclic behaviour of stratified soil under liquefied states

Arpit Jain, Satyendra Mittal & Sanjay Kumar Shukla

**To cite this article:** Arpit Jain, Satyendra Mittal & Sanjay Kumar Shukla (2022): Cyclic behaviour of stratified soil under liquefied states, Marine Georesources & Geotechnology, DOI: 10.1080/1064119X.2022.2095946

**To link to this article:** https://doi.org/10.1080/1064119X.2022.2095946

Published online: 11 Jul 2022.

Submit your article to this journal

Article views: 191

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=umgt20

MARINE GEORESOURCES & GEOTECHNOLOGY
https://doi.org/10.1080/1064119X.2022.2095946

Taylor & Francis
Taylor & Francis Group

Check for updates

# Cyclic behaviour of stratified soil under liquefied states

Arpit Jain[a], Satyendra Mittal[a] and Sanjay Kumar Shukla[b,c]

[a]Department of Civil Engineering, Indian Institute of Technology, Roorkee, India; [b]School of Engineering, Edith Cowan University, Joondalup, Australia; [c]Department of Civil Engineering, Delhi Technological University, Delhi, India

**ABSTRACT**

A substantial amount of work on the cyclic behaviour of cohesionless soil has been performed after neglecting the reality of stratification existing nearby marine structures or alluvial plains. This complicated mechanism of stratified soil failure on earthquake-induced soil liquefaction is affected significantly by the presence of a less permeable soil layer. Several past studies investigated that under the action of dynamic loads, the presence of a silt layer in homogeneous sandy soil crucially modifies the pore pressure dissipation characteristics, axial strain rate and stress–strain behaviour of the whole stratified specimen. In the present study, cyclic behaviour of sand specimen interlayered with silt layer has been presented in respect of its pre-and post-liquefaction behaviour. Undrained stress-controlled cyclic triaxial tests have been performed on homogeneous sand specimens and on specimens interlayered with a silt layer, placed at $H/4$, $H/2$ and $3H/4$ (where $H$ is the height of the specimen from the top) to signify the role of placement depth. The instability and failure of such specimens have been compared using the failure line and phase transformation line. Few key observations suggested that the number of liquefaction failure cycles enhanced two times and secant modulus (under post liquefied state) improved by 4.8 times when the placement depth of the silt layer increased from $H/4$ to $3H/4$ under similar loading conditions. Additionally, the role of initial static shear stress has been investigated at three different confining pressures in a few stratified specimens. It has been observed that failure of the stratified specimen is due to the obstruction of pore water movement caused by the presence of the silt layer. A comparison of the present work has been made with significant studies using liquefaction resistant curves to ensure the efficacy of this study.

**ARTICLE HISTORY**
Received 19 January 2022
Accepted 20 May 2022

**KEYWORDS**
Cyclic triaxial; stratified soil; silt; pre-liquefaction; post-liquefaction; initial stresses

## 1. Introduction

The failure of soil mass due to earthquake-induced liquefaction was noted for the first time in the Niigata earthquake, in 1964. A condition where the particle contact stress becomes zero due to the generation of high excess pore water pressure is termed liquefaction (Kramer 1996). One of the deadliest earthquakes occurred in Christchurch, New Zealand, in 2011 where lateral spreading, landslides, and liquefaction played a key role in destruction (Kaiser et al. 2012). Also, the occurrence of liquefaction nearby the marine environment has severely affected the stability of offshore structures (Jeng 2013). Soil strata around marine structures exist in a combination of sandy, silty or clayey strata (Liu and Jeng 2016). The authors performed a series of experiments with nearly 3000 wave cycles for each test under several conditions to study the influence of clay content on wave-induced liquefaction in marine sediments. Therefore, a simple study using homogeneous soil specimens will be either underestimating or overestimating the failure mechanism (Ecemis 2021). There exists a scarcity of knowledge and research work on the liquefaction behaviour of stratified soil strata. It is essential to understand the behaviour of stratified soil liquefaction to encounter the design problems and safety of geotechnical structures located in the vicinity of marine or alluvial soil deposits. In a study performed by Liu and Jeng (2016), wave-induced liquefaction was investigated in marine sediments using one-dimensional equipment. Authors in this study emphasised the usage of multilayers rather than single sandy strata, therefore, they utilised sand-clay strata and observed critical clay content value. Previously, Liu and Qiao (1984) investigated the liquefaction behaviour of soil strata underlying the structure's foundation. The authors observed a clear difference in failure mechanism in homogeneous and stratified soil systems due to the generation of water-film and sand boils. Elgamal, Dobry, and Adalıer (1989) also conducted shaking table tests that focussed on the liquefaction of multilayered sandy soil systems. Fiegel and Kutter (1994) performed experiments on layered soil models consisting of fine sand overlaid by a layer of less permeable silica flour (silt) using a centrifuge. The excess pore water pressure generated in the fine sand vented out through the thinnest zone of the overlying silt layer.

Balakrishnan and Kutter (1999) performed centrifuge tests on the stratified system of sand and clay to understand the problem of settlement and lateral sliding in the case of a

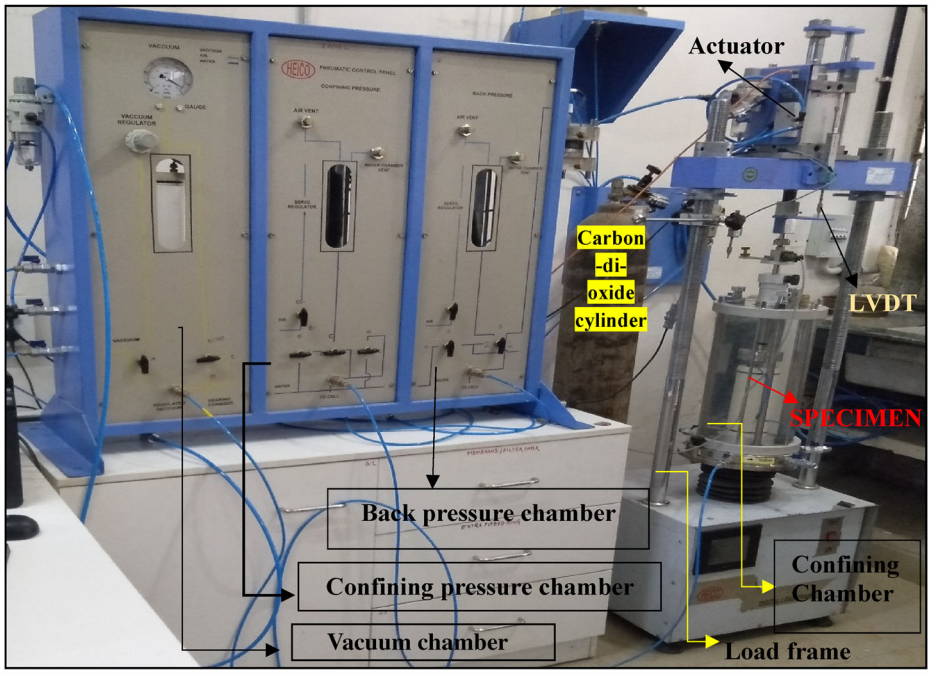
CONTACT Arpit Jain ajain4@ce.iitr.ac.in Department of Civil Engineering, Indian Institute of Technology, Roorkee, Uttarakhand, India
© 2022 Informa UK Limited, trading as Taylor & Francis Group

bridge. The authors suggested the improvement of the sand-clay interface through anchoring and nailing to reduce lateral deformations. Arulanandam and Scott (1993) performed experimental studies to investigate the effect of stratification and inhomogeneity of soil mass on the liquefaction-induced lateral spreading. Using the seams of rock flour paste in liquefiable sand deposits, Brennan and Madabhushi (2005) performed a series of dynamic centrifuge tests on capped soil strata (less permeable soil on the top) and other stratified soil systems to replicate the field conditions nearby marine soil, alluvial plains and recent fills. Here, due to the development of high-pressure fluid (water film), destabilising effect on the overlying layer resulted in an early failure. Maharjan and Takahashi (2013) performed centrifuge tests on stratified soil using patches of silt layers and observed that there was the generation of pore pressure due to the differences in permeability which restricts the upward movement resulting from the uneven settlements. Ozener, Ozaydin, and Berilgen (2009) gave experimental results to demonstrate the effects of relative density, input acceleration and presence of a silt seam on the generation of excess pore water pressure in sand deposits using a shake table subjected to dynamic loading. Recently, several researchers have utilised cyclic triaxial facilities to closely observe the liquefaction failures and role of silt seams in the stratified soil system (Jia and Wang 2012; Konrad and Dubeau 2003; Tasiopoulou et al. 2019; Xiu et al. 2020). Konrad and Dubeau (2003) performed undrained cyclic triaxial tests on stratified silt specimens where the presence of silt seam induced a much lower cyclic resistance. Jia and Wang (2012) performed strain-controlled cyclic triaxial tests on sand interlayered with silt and obtained a critical thickness of the silt seam corresponding to the critical value of shear strain. Krim et al. (2019) investigated the liquefaction resistance of sand-clay mixtures after carefully observing the role of clay content and gradation characteristic using triaxial facility. Xiu et al. (2020) depicted the importance of stratification by performing several cyclic triaxial experiments on the stratified soil specimens prepared by inserting a powdery layer into the fine grained soil. Using microbially induced desaturation and precipitation (MIDP), Stallings Young et al. (2021) used an innovative approach of liquefaction mitigation in the stratified soils to lower the risk of liquefaction nearby embankment along the lower Fraser river. The understanding of liquefaction behaviour of silt is also very important to maintain stability of several structures existing nearby silty regions. Karakan, Tanrinian, and Sezer (2019) investigated the liquefaction and post-liquefaction behaviour of non-plastic silt using cyclic triaxial testing and observed the pore pressure response and axial strains responsible for significant failures. Additionally, Karakan and Altun (2018) studied the liquefaction behaviour of silty sand after performing a series of stress-controlled tests on specimens with different silt contents and different loading conditions and environments.

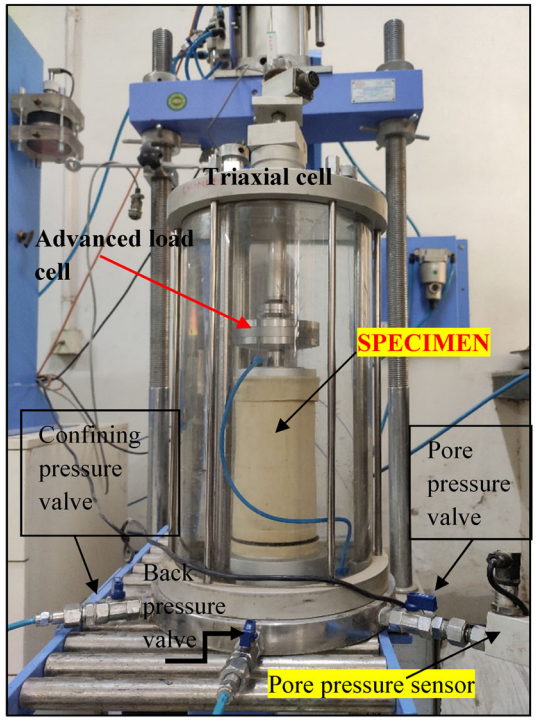
Most of these studies have provided a clear insight into the failure mechanism of stratified soil systems during the occurences of earthquakes or liquefaction. Several researchers such as Toyota (1995), Dash (2008), Porcino, Marcianò, and Nicola Ghionna (2009), and Sriskandakumar, Wijewickreme, and Byrne (2012) investigated the shear behaviour of post-liquefied undrained sandy soil. Additionally, many researchers have noticed prominent liquefaction failures in the stratified soil strata even after the occurrence of earthquakes as a post-liquefaction study (Kokusho 1999; Kokusho and Fujita 2001; Kokusho and Kojima 2002). The flow failures in Niigata city during the 1964 earthquake were studied by Kokusho and Fujita (2001) where the water film acted as a sliding surface. A similar type of film was observed by Kokusho and Kojima (2002) where the water film beneath a less permeable layer triggered the post-liquefaction failures. Using a 1 D lucite tube, Kokusho and Kojima (2002) gave the mechanism of post-liquefaction water film generation in layered sand. As reported by Amini and Trandafir (2008), liquefied soil deposits exhibit the post-liquefaction shear strength response. Sitharam, Vinod, and Ravishankar (2009) obtained a monotonic stress–strain response of post-liquefied soil where excess pore water pressure at the end of cyclic loading was allowed to dissipate. Zhang, Yang, and Yu (2021) also applied monotonic undrained shearing in order to investigate the post-liquefaction behaviour of soil specimens and noticed the importance of cyclic shear strain amplitude in governing post-liquefaction shear strength.

Most of the marine structures have to encounter intense wind forces, ocean thrusts and severe deformation under the action of earthquakes or liquefaction failures (if presented nearby significant earthquake zone). The research work on the liquefaction of stratified soil strata is very important but too scanty. Several sites nearby marine or intertidal deposits have a presence of a less permeable soil layer at varying depths. The prime aim of the present study is based on the replication of the soil conditions nearby alluvial plain, marine or similar deposits using a cyclic triaxial testing facility. Significant parameters like density, deviator stress, the effect of confinement and role of initial stresses have been considered for three different placement depths which is correctly useful for critical field projects nearby similar soil conditions. Also, the placement depth variation of the silt layer governed strength behaviour under post-liquefaction conditions too.

Cyclic loading plays a key role in replicating the nature and distribution of several waveforms or loading types due to the earthquakes, sea waves or wind forces (Kargar, Salehzadeh, and Shahnazari 2016). The influence of the presence of a silt layer or less permeable soil on the liquefaction characteristics of stratified specimens needs further investigation. Therefore, the present study aims to investigate the effect of the silt layer on the cyclic behaviour of fine sand for three different silt layer locations in an attempt to replicate the soil strata nearby alluvial plains, marine structures and recent fills. The role of initial stresses has also been explored at three different confining stresses on the selected stratified specimens. Using the variation in key parameters like density, cyclic stress ratio (*CSR*) and stress amplitude, pre and post-liquefaction studies have been carried out.




(a) Complete experimental setup




(b) Triaxial cell with specimen

**Figure 1.** Cyclic triaxial testing facility.

## 2. Experimental details

### 2.1. Cyclic triaxial testing

The cyclic triaxial testing apparatus used in this study consists of a load frame with an actuator to simulate cyclic loading, a data logger, pore pressure transducer, triaxial cell, and chamber for the application of confining pressure and backpressure, all of them connected to a data acquisition system. The schematic diagram of the testing facility is shown in Figure 1. The frequency range for the apparatus is 0.1 Hz to 10 Hz; however, for this study, the adopted value of frequency was 1 Hz as used by previous researchers to study the liquefaction phenomenon (Kargar, Salehzadeh, and Shahnazari 2016). Different sensors are provided in load cell, pressure transducer and linearly variable differential transducer (LVDT). A specially designed load cell is located

Table 1. Properties of materials used (As per IS:1498).

| S.N. | Particulars | Notations | Silt | Fine sand |
|---|---|---|---|---|
| 1. | Soil type | – | Low plastic silt (ML) | SP |
| 2. | Maximum particle size | $D_{max}$ | 0.075 mm | 0.42 mm |
| 3. | Mean particle size | $D_{50}$ | 0.017 mm | 0.20 mm |
| 4. | Specific gravity | $G_s$ | 2.68 | 2.65 |
| 5. | Maximum void ratio | ($e_{max}$) | 1.27 | 0.88 |
| 6. | Minimum void ratio | ($e_{min}$) | 0.34 | 0.57 |
| 7. | Liquid limit | $w_l$ | 15% | – |
| 8. | Plastic limit | $w_p$ | 12% | – |
| 9. | Permeability | $k$ | $5.8 \times 10^{-5}$ cm/s | $7.2 \times 10^{-3}$ cm/s |
| 10. | Coefficient of uniformity | $C_u$ | – | 1.95 |

inside the triaxial cell which avoids the friction coming on the connecting rod and thus provides exact load value on the specimen. The sensitivity of the load cell is 0.1% of the total capacity (5 kN). The variations in strain, deformations and deflections can be measured with the help of LVDT (±50 mm range). Also, the pore pressure transducer (PPT) is a pressure measuring device with a maximum capacity of 2000 kPa located at the top position. The volume measurement device used in the study has a capacity of 200 ccs and a sensitivity of ±0.01 cc. All the testing procedures were adopted as per ASTM 5311. The triaxial cell used can accommodate a specimen of diameter 70 mm, with a length ($L$) to diameter ($D$) ratio of 2:1.

### 2.2. Materials used

Fine sand (SP) and silt (ML) were used in this study. The fine sand was collected from the local Solani river, Roorkee, India, which was further oven-dried before being used for the experiments. The material for silt was collected from a canal located at Solanipuram park, Roorkee, India. This collected material was oven-dried and then passed through a 75 µm sieve to obtain desired silty material. The classification of fine sand and silt was carried out and their properties are given in Table 1 as per IS: 1498. Figure 2(a) and (b) show the photographic and SEM images of sand and silt. It can be noticed that sand particles have a sub-rounded shape while silty particles have angular shapes. The magnification of 250 x and 500 x was used while capturing SEM images. The particle size distribution curve of fine sand and silt used in the study has been presented in Figure 2(c).

### 2.3. Specimen preparation and test procedure

The cylindrical specimens (diameter = 70 mm, height = 140 mm) were prepared using the oven-dried sand and silt through the split mould. Fine sand and silt were poured inside as per the air pluviation technique to simulate natural soil deposits (Ladd 1974, Jain, Mittal, and Shukla 2021). Air pluviation was adopted to create clearly segregated and uniform specimens like that of the natural profiles. Homogeneous specimens were prepared in five layers using a conical funnel whereas in five to seven layers in the case of stratified specimens. In order to prepare stratified specimens, the location of the silt layer was fixed beforehand through colour markings, provided over the membrane. The amount of sand or silt to be used was pre-calculated to ensure the attainment of desirable relative density within the colour markings. Firstly, the fine sand was poured inside, followed by less permeable silt and finally, the top layer of fine sand was provided using a conical funnel.

The images provided could be considered as evidence of the steps adopted by the authors for specimen preparation. One of the cases from the test series has been considered (the silt layer was located at 35 mm or H/4 from the top of the specimen) to depict the detailed specimen procedure through captured images. The significant steps followed during specimen preparation have been depicted using Figure 3(a)–(d). The cross-sectional view can be seen in Figure 3(a). The suitable markings provided inside the membrane for the pre-calculated amount of soil (sand or silt) are visible in Figure 3(a). It shows the preparation of a stratified specimen where the silt layer of 8 mm thickness was placed at 35 mm from the top of the specimen. It can be seen in Figure 3(b) that sand was poured until it reached a designated marking to maintain desired density. In the further steps, sand was placed followed by silt placement at 35 mm from the top as shown in Figure 3(c) and (d). Finally, the top sand layer was placed at a pre-calculated amount. To facilitate a better and more accurate specimen preparation, markings were also provided on the scale used.

The attainment of desired density in the air pluviation (AP) technique depends on two main parameters, i.e., drop height and flow rate of the material used (Miura and Toki 1982, Cresswell, Barton, and Brown 1999, Wijewickreme, Sriskandakumar, and Byrne 2005). The flow rate is defined as the pluviated soil mass falling per unit time. To achieve lower density, smaller drop heights and higher flow rates are advisable (Vaid and Negussey 1984, Zhao et al. 2006, Choi et al. 2010). In this study, the flow rate was adopted as 70.4 N/min and 36.8 N/min for low and medium dense specimens, respectively. Also, the constant drop height of 3 cm was maintained for the preparation of the specimens. Secondly, due to the technical difficulties in other specimen preparation of the stratified specimens, the air-pluviation technique was adopted. Initially, few check tests were performed with the water sedimentation technique (WS) too. Here, due to the improper flow of carbon dioxide and difficulty in the application of vacuum in the stratified specimens, this technique was not adopted. Meanwhile, there was no such problem with the AP technique. Therefore, due to the adopted flow rate, drop height and markings; it was possible to achieve low to medium density for the stratified specimens using the air-pluviation technique. Three different varieties of the stratified patterns were adopted in relation to the location of the silt layer as 35 mm ($H$/4), 70 mm ($H$/2), and 105 mm (3$H$/4) from the top of the specimen, where $H$ is the height of the specimen, i.e., 140 mm.

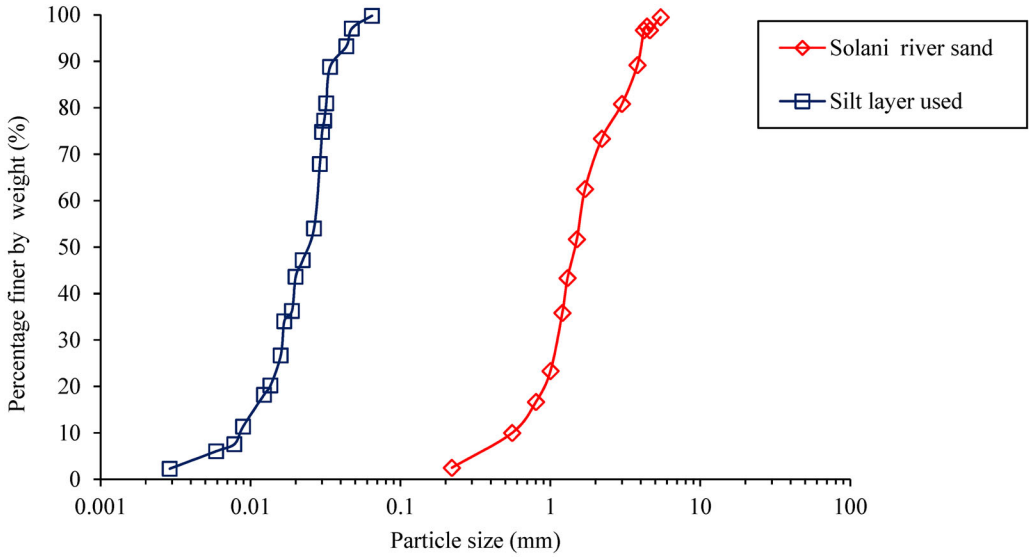
The specimen was finally constituted using an applied vacuum due to the inability of the sand specimen to stand vertical on its own. Taking solubility advantage of $CO_2$ molecules in water, $CO_2$ was applied for 30 minutes where air molecules are replaced, which resulted in optimal saturation.

(i) Photographic image (ii) SEM image (500X)

(a) Fine sand

(i) Photographic image (ii) SEM image (500X)

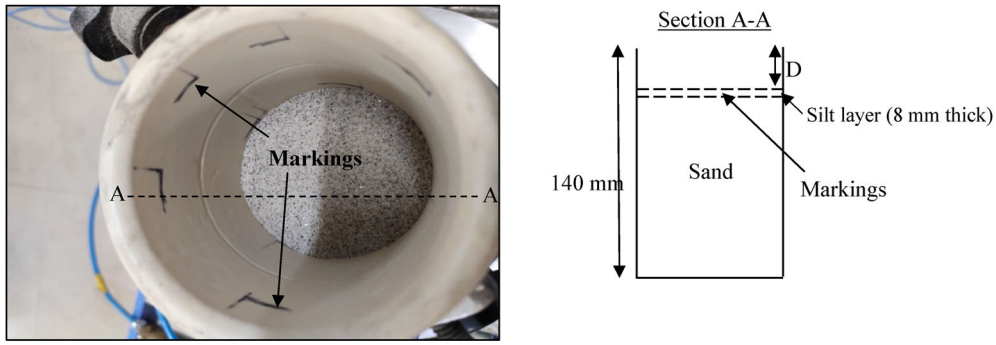
(b) Silt



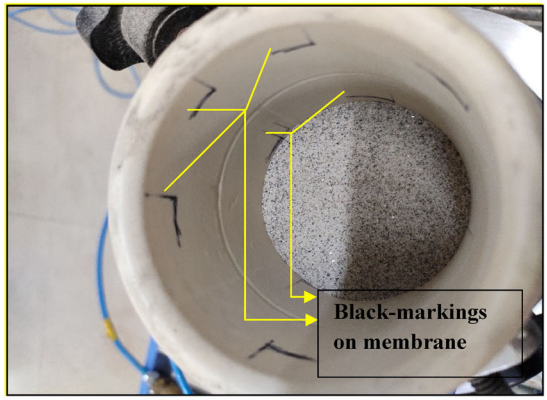

(c) Particle size distribution curves of sand and silt

**Figure 2.** Materials used in the study.

Then, the deaired water was circulated from bottom to top of the specimen using the backpressure valve in order to saturate the material inside. In this study, as the *B* value reached 0.98 it was considered as the complete saturation of the whole stratified specimen including the silt layer. This was due to the unavailability of Skempton's coefficient *B* for

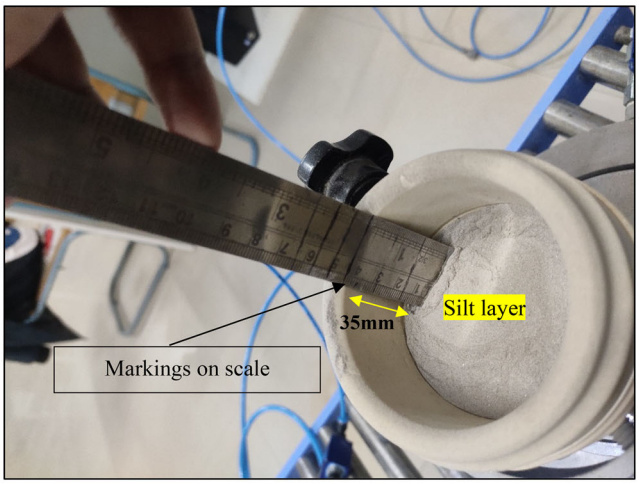



(a) Markings used in the preparation of stratified specimen along with the cross-sectional view.




(b) Image showing first stage of specimen preparation

(c) Image showing placement of soil mass up to the required height




(d) Placement of silt and pre-check before sand placement

**Figure 3.** Step-wise description of specimen preparation.

Table 2. Details of cyclic test performed (As per ASTM 5311).

| Specimen | $D_{rsand,1}$ | $D_{rsilt}$ | $D_{rsand,2}$ | $D_{r,m}$ | $D$ | $CSR$ | $q_{cyc}$ (kPa) | $\sigma'_c$ (kPa) | $CP$ (kPa) | $BP$ (kPa) | $N_f$ | $B$ | $\alpha$ | $r_u$ | $\varepsilon$ (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S1 (NS) | 29.92 | – | – | 29.92 | No silt | 0.16 | 32.5 | 100 | 200 | 100 | 25 | 0.99 | 0 | 1 | 2.52 |
| S2 | 27.24 | 71.25 | 27.62 | 36.10 | *H*/2 | 0.16 | 32.5 | 100 | 220 | 120 | 38 | 0.98 | 0 | 0.99 | 2.51 |
| S3 | 26.28 | 70.24 | 27.55 | 33.20 | *H*/4 | 0.16 | 32.5 | 100 | 240 | 140 | 30 | 0.98 | 0 | 0.99 | 2.49 |
| S4 | 28.26 | 72.24 | 28.51 | 35.00 | 3*H*/4 | 0.16 | 32.5 | 100 | 240 | 140 | 61 | 0.98 | 0 | 1 | 2.53 |
| S5 | 27.24 | 71.25 | 29.24 | 34.80 | *H*/2 | 0.20 | 32.5 | 100 | 240 | 140 | 19 | 0.98 | 0 | 0.99 | 2.52 |
| S6 | 28.26 | 72.24 | 29.56 | 35.10 | 3*H*/4 | 0.20 | 32.5 | 100 | 260 | 160 | 22 | 0.98 | 0 | 1 | 2.51 |
| S7 | 26.28 | 70.24 | 27.81 | 35.32 | *H*/4 | 0.20 | 32.5 | 100 | 240 | 140 | 10 | 0.98 | 0 | 1 | 2.50 |
| S8 | 48.72 | 78.22 | 49.25 | 53.43 | *H*/2 | 0.16 | 32.5 | 100 | 220 | 120 | 348 | 0.98 | 0 | 0.98 | 0.3 |
| S9 | 45 | 75.02 | 46.55 | 50.15 | *H*/4 | 0.16 | 32.5 | 100 | 220 | 120 | 81 | 0.98 | 0 | 1 | 2.51 |
| S10 | 48.8 | 77.23 | 50.22 | 53.80 | 3*H*/4 | 0.16 | 32.5 | 100 | 240 | 140 | 187 | 0.98 | 0 | 0.99 | 2.52 |
| S11 (NS) | 48.87 | – | – | 48.87 | No silt | 0.16 | 32.5 | 100 | 200 | 100 | 75 | 0.99 | 0 | 1 | 2.53 |
| S12* | 27.24 | 71.25 | 28.52 | 36.14 | 0 | 0.16 | 32.5 | 100 | 220 | 120 | 28 | 0.98 | 0 | 1 | 2.50 |
| S13** | 27.24 | 71.25 | 29.52 | 36.15 | $H$-($t_{silt}$/2) | 0.16 | 32.5 | 100 | 240 | 140 | 78 | 0.98 | 0 | 1 | 2.52 |
| S14 (NS) | 30.02 | – | – | 30.02 | No silt | 0.16 | 32.5 | 100 | 200 | 100 | 55 | 0.99 | 0.09 | 1 | 2.53 |
| S15 | 27.28 | 72.24 | 28.96 | 27.42 | *H*/2 | 0.16 | 32.5 | 100 | 200 | 100 | 151 | 0.98 | 0.09 | 0.98 | 2.01 |
| S16 | 28.27 | 71.24 | 29.66 | 36.10 | *H*/2 | 0.27 | 32.5 | 50 | 150 | 100 | 13 | 0.98 | 0.166 | 1 | 10.02 |
| S17 | 27.24 | 71.22 | 28.56 | 36.30 | *H*/2 | 0.1 | 32.5 | 150 | 220 | 120 | 535 | 0.99 | 0.062 | 0.97 | 0.2 |
| S18 | 28.27 | 72.54 | 29.62 | 36.20 | *H*/4 | 0.20 | 40 | 100 | 220 | 120 | 28 | 0.98 | 0.09 | 1 | 2.52 |
| S19 | 29.17 | 71.55 | 30.21 | 36.20 | *H*/2 | 0.20 | 40 | 100 | 240 | 140 | 69 | 0.98 | 0.09 | 1 | 2.53 |
| S20 | 28.89 | 72.54 | 30.22 | 36.20 | 3*H*/4 | 0.20 | 40 | 100 | 220 | 100 | 95 | 0.98 | 0.09 | 1 | 2.50 |
| S21 | 27.25 | 71.89 | 29.85 | 36.20 | *H*/4 | 0.27 | 32.5 | 50 | 170 | 120 | 8 | 0.98 | 0.166 | 1 | 2.53 |
| S22 | 27.52 | 72.85 | 28.52 | 34.42 | 3*H*/4 | 0.27 | 32.5 | 50 | 150 | 100 | 13 | 0.98 | 0.166 | 1 | 2.49 |
| S23(NS) | 28.52 | 72.48 | 29.65 | 28.52 | No silt | 0.27 | 32.5 | 50 | 170 | 120 | 4 | 0.99 | 0.166 | 1 | 2.53 |
| S24(FS) | 30.02 | – | – | 30.02 | Full silt | 0.14 | 28 | 100 | 260 | 160 | 31 | 0.98 | 0 | 1 | 2.51 |
| S25(FS) | 30.54 | – | – | 30.54 | Full silt | 0.16 | 32.5 | 100 | 260 | 160 | 10 | 0.98 | 0 | 1 | 2.49 |
| S26(FS) | 30.55 | – | – | 30.55 | Full silt | 0.20 | 40 | 100 | 280 | 180 | 5 | 0.98 | 0 | 1 | 2.53 |

Note: CP = confining pressure, BP = Back pressure, S12* = Specimen have silt layer located on the top of the specimen, S13** = Specimen have silt layer located at the bottom of the specimen. All the stress values provided above are in kPa. The full form of abbreviation used has been provided in the notations list.

the stratified conditions. In all of the specimens, Skempton's coefficient $B$ achieved was more than 0.98. The attainment of optimal saturation was with the increment of 20 kPa in both confining and backpressure. Then, the specimens were consolidated in isotropic and anisotropic states at a desired effective confining pressure as shown in the test series under undrained conditions (Table 2). Few monotonic tests were also performed on both homogeneous and stratified specimens in order to locate the failure line (FL) and phase transformation line (PTL).

### 2.4. Calculation of relative density for the stratified specimens

The calculation of relative density in the case of the stratified specimens does not follow a similar concept as that of the homogeneous soil specimen. Therefore, in order to resolve this difficulty, Yoshimine and Koike (2005) suggested a method to calculate the relative density of the stratified specimens. This method was also adopted by earlier researchers (Jia and Wang 2012; Xiu et al. 2020). The following expressions have been utilised:

$$D_{r,m} = \frac{e_{max,m} - e_e}{e_{max,m} - e_{min,m}} \quad (1)$$

$$e_{max,m} = \sum_{i}^{n} [R_{m,i}(e_{max,\ i} + 1)] - 1 \quad (2)$$

$$e_{min,m} = \sum_{i}^{n} [R_{m,i}(e_{min,i} + 1)] - 1 \quad (3)$$

$$e_e = \sum_{i}^{n} [R_{m,i}(e_i + 1)] - 1 \quad (4)$$

In these equations $D_{r,m}$, $e_{\max,m}$, $e_{\min,m}$, $e_e$ and $R_{m,i}$ denote modified relative density, modified maximum void ratio, modified minimum void ratio, equivalent void ratio and percentage of the dry weight of $i$th layer in the stratified soils, respectively. The details of all the parameters are presented in Table 3.

As suggested in several papers (Belkhatir et al. 2010; Thevanayagam and Mohan 2000), for the calculation of intergranular and interfine void ratio the presence of fine content plays a key role. These fine particles enter into the matrix of coarser soil skeleton to form a stable or unstable structure depending on their content. In all of these previous studies, the fine particles were uniformly mixed with coarser particles at different contents where the concept of intergranular and interfine void ratio is more useful. In this study, the fine particles (silt) have been used as an obstruction in pore water flow rather than mixing with sand grains. Therefore, for different layering systems, the concept of equivalent or effective void ratio has been used for analysis purposes in correlation with previous research studies on stratified soil systems (Jia and Wang 2012; Yoshimine and Koike 2005; Xiu et al. 2020). Additionally, the effect of mixing of fine content on the stratified system will emerge as a useful future research work where the concept of intergranular and interfine void ratio can be utilised (Table 3).

### 2.5. Significant parameters utilised in the study

After successfully preparing the specimens, consolidation stage was initiated at isotropic and anisotropic states as per the requirement of the test under undrained conditions. Three different effective confining stresses, i.e., 50 kPa, 100 kPa and 150 kPa were applied at initial stress of 20 kPa

**Table 3.** Details of the parameters used in the calculation of relative density of the specimens.

| Specimen | $D_{rsand,1}$ | $W_{sand,1}$ (g) | $e_{sand,1}$ | $R_{m,1}$ | $D_{r,\ silt}$ (%) | $W_{silt}$ (g) | $e_{silt}$ | $R_{m,silt}$ | $D_{rsand,2}$ (%) | $W_{sand,2}$ (g) | $e_{sand,2}$ | $R_{m,2}$ | $D_{r,\ m}$ (%) | $e_e$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S1 (NS) | 29.92 | 798.41 | 0.78 | – | – | – | – | – | – | – | – | – | 29.92 | 0.78 |
| S2 | 27.24 | 374.64 | 0.79 | 0.46 | 71.25 | 50.71 | 0.60 | 0.06 | 27.62 | 376.9 | 0.79 | 0.47 | 36.10 | 0.77 |
| S3 | 26.28 | 175.69 | 0.80 | 0.22 | 70.24 | 50.42 | 0.61 | 0.06 | 27.55 | 573.66 | 0.80 | 0.72 | 33.20 | 0.79 |
| S4 | 28.26 | 574.36 | 0.79 | 0.71 | 72.24 | 50.95 | 0.60 | 0.06 | 28.51 | 176.36 | 0.79 | 0.22 | 35.00 | 0.77 |
| S5 | 27.24 | 374.66 | 0.80 | 0.46 | 71.25 | 50.71 | 0.60 | 0.06 | 29.24 | 375.96 | 0.79 | 0.47 | 34.80 | 0.78 |
| S6 | 28.26 | 574.30 | 0.79 | 0.71 | 72.24 | 50.95 | 0.60 | 0.06 | 29.56 | 176.68 | 0.79 | 0.22 | 35.10 | 0.77 |
| S7 | 26.28 | 175.69 | 0.80 | 0.22 | 70.24 | 50.42 | 0.61 | 0.06 | 27.81 | 573.92 | 0.79 | 0.72 | 35.32 | 0.77 |
| S8 | 48.72 | 389 | 0.73 | 0.46 | 78.22 | 52.8 | 0.545 | 0.06 | 49.25 | 389.47 | 0.73 | 0.47 | 53.43 | 0.73 |
| S9 | 45 | 181.54 | 0.74 | 0.22 | 75.02 | 51.85 | 0.57 | 0.06 | 46.55 | 593.13 | 0.74 | 0.72 | 50.15 | 0.73 |
| S10 | 48.80 | 595.5 | 0.73 | 0.71 | 77.23 | 52.47 | 0.55 | 0.06 | 50.22 | 183.25 | 0.72 | 0.22 | 53.80 | 0.71 |
| S11 (NS) | 48.87 | 825.58 | 0.73 | – | – | – | – | – | – | – | – | – | 48.87 | 0.73 |
| S12* | – | – | – | – | 71.25 | 50.71 | 0.60 | 0.06 | 28.52 | 751.00 | 0.79 | – | 36.14 | 0.77 |
| S13** | 27.24 | 749.33 | 0.79 | 0.93 | 71.25 | 50.71 | 0.60 | 0.06 | – | – | – | – | 36.15 | 0.77 |
| S14 (NS) | 30.02 | 826.3 | 0.79 | – | – | – | – | – | – | – | – | – | 30.02 | 0.79 |
| S15 | 27.28 | 374.69 | 0.80 | 0.5 | 72.24 | 50.95 | 0.60 | 0.06 | 28.96 | 375.78 | 0.79 | 0.50 | 27.42 | 0.80 |
| S16 | 28.27 | 375.33 | 0.79 | 0.46 | 71.24 | 50.42 | 0.60 | 0.06 | 29.66 | 376.24 | 0.79 | 0.47 | 36.10 | 0.77 |
| S17 | 27.24 | 347.66 | 0.79 | 0.45 | 71.22 | 50.71 | 0.60 | 0.06 | 28.56 | 375.52 | 0.79 | 0.48 | 36.30 | 0.77 |
| S18 | 28.27 | 176.30 | 0.79 | 0.22 | 72.54 | 51.10 | 0.59 | 0.06 | 29.62 | 575.72 | 0.79 | 0.72 | 36.20 | 0.77 |
| S19 | 29.17 | 375.60 | 0.79 | 0.46 | 71.55 | 50.80 | 0.60 | 0.06 | 30.21 | 376.60 | 0.79 | 0.47 | 36.20 | 0.77 |
| S20 | 28.89 | 574.94 | 0.79 | 0.71 | 72.54 | 51.10 | 0.59 | 0.06 | 30.22 | 176.90 | 0.79 | 0.22 | 36.20 | 0.77 |
| S21 | 27.25 | 176.00 | 0.79 | 0.22 | 71.89 | 51.00 | 0.60 | 0.06 | 29.85 | 575.95 | 0.79 | 0.71 | 36.20 | 0.77 |
| S22 | 27.52 | 573.63 | 0.80 | 0.71 | 72.85 | 51.24 | 0.59 | 0.06 | 28.52 | 176.37 | 0.79 | 0.22 | 34.42 | 0.78 |
| S23(NS) | 28.52 | 796.51 | 0.79 | – | – | – | – | – | – | – | – | – | 28.52 | 0.79 |
| S24(FS) | – | – | – | – | 30.02 | 717.37 | 0.99 | – | – | – | – | – | 30.02 | 0.99 |
| S25(FS) | – | – | – | – | 30.54 | 718.54 | 0.98 | – | – | – | – | – | 30.54 | 0.98 |
| S26(FS) | – | – | – | – | 30.55 | 718.58 | 0.98 | – | – | – | – | – | 30.55 | 0.98 |

Note: $D_{rsand,1}$ = Relative density of the upper sand layer, $W_{sand,1}$ = Weight of the upper sand layer $R_{m,1}$ = percentage of the dry weight of the upper sand layer in the stratified soil strata $D_{rsilt}$ = Relative density of the silt layer, $W_{sand,1}$ = Weight of the silt layer $R_{m,1}$ = percentage of the dry weight of the silt layer in the stratified soils $D_{rsand,2}$ = Relative density of the lower sand layer, $W_{sand,1}$ = Weight of the lower sand layer $R_{m,1}$ = percentage of the dry weight of the lower sand layer in the stratified soils $e_{sand,1}$ = void ratio of the upper sand layer, $e_{silt}$ = void ratio of the silt layer, $e_{sand,2}$ = void ratio of the lower sand layer, $e_e$ = equivalent void ratio of the specimen S12* = Specimen have silt layer located on the top of the specimen, S13** = Specimen have silt layer located at the bottom of the specimen.

(Table 2). The isotropic tests were also performed at confining stress of 50 kPa, 100 kPa and 150 kPa. Tests were performed at different combinations of relative density ranges from 27.42 (loose) to 53.80% (medium dense). The silt seam was prepared at a relative density ranging between 70.24 to 78.22%, in order to obstruct the maximum drainage possibility. This obstruction plays a key role in early failures as observed by Butterfield and Bolton (2003). The role of anisotropic consolidation was to create static shear stress on the plane of maximum shear stress in the soil specimen (Yang and Sze 2014). The initial shear stress ratio (α) has been given by Yang and Sze (2011 a) as:

$$\sigma'_{nc} = \frac{(\sigma'_{1c} + \sigma'_{3c})}{2} \tag{5}$$

$$\alpha = \frac{q_s}{2 \times \sigma'_{nc}} \tag{6}$$

where $q_s$ = initial deviator stress, $\sigma'_{1c}$ = major principal stress, $\sigma'_{3c}$ = minor principal stress, and $\sigma'_{nc}$ = effective normal stress

Seed and Lee (1966) showed that static shear stress $q_s/2$ on the 45° planes simulates field stress conditions, here $\sigma'_{nc}$ rather than $\sigma'_{3c}$ is used as it is the effective normal stress $\sigma'_{nc}$. Yang and Sze (2014) showed that the parameter "α" indicates a special condition of isotropic consolidation when equal to zero. At the isotropic state, there is an absence of initial static shear stress in soil specimens. In this case, the value of "α" varies due to changes in confining stresses as shown in Table 2, whereas $q_s$ is taken as 20 kPa for all the anisotropic cases. After the required isotropic or anisotropic consolidated state, cyclic loading was applied to the specimen. All the cyclic inputs were provided through a stress-controlled undrained mechanism at two different cyclic stress ratios (*CSR*) at a frequency of 1 Hz.

$$CSR = \frac{q_{cyc}}{2 \times \sigma'_c} \tag{7}$$

where $q_{cyc}$ = cyclic deviator stress, $\sigma'_c$ = effective confining pressure

The stratified specimen had unique strain development and pore pressure dissipation characteristics compared to that of the homogeneous ones. The variation in deformation and pore water pressure characteristics have been observed for the key parameters like relative density, cyclic stress ratio and location of silt layer as shown in Table 2. A total of 26 cyclic and 16 monotonic tests were considered in this study. Each test was performed two times in order to maintain the uniformity of results. Tables 2 and 4 summarise the details of all the cyclic and monotonic tests performed, respectively.

## 3. Results and discussion

A total of 26 cyclic triaxial tests were performed on both the stratified and non-stratified specimens using stress-controlled mechanism. Seed and Lee (1966) observed initial liquefaction in the isotropically consolidated specimens when the pore pressure ratio ($r_u$) reaches unity at which effective stress becomes zero and provided the expression for $r_u$ as

$$r_u = \frac{\Delta u}{\sigma'_{3c}} \tag{8}$$

**Table 4.** Details of monotonic tests performed.

| Specimen | Type | $D_{rsand,1}$ | $D_{rsilt}$ | $D_{rsand,2}$ | $D_{r,m}$ | $D$ | Strain rate (mm/min) | $\sigma'_3$ (kPa) | $CP$ (kPa) | $BP$ (kPa) | $B$ | $r_u$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S5 | Post-liquefaction | 27.24 | 71.25 | 29.24 | 34.80 | $H/2$ | 1.2 | 100 | 240 | 140 | – | 0.99 |
| S6 | Post-liquefaction | 28.26 | 72.24 | 29.56 | 35.10 | $3H/4$ | 1.2 | 100 | 260 | 160 | – | 1 |
| S7 | Post-liquefaction | 26.28 | 70.24 | 27.81 | 35.32 | $H/4$ | 1.2 | 100 | 240 | 140 | – | 1 |
| S8 | Post-liquefaction | 48.72 | 78.22 | 49.25 | 53.43 | $H/2$ | 1.2 | 100 | 220 | 120 | – | 0.98 |
| S9 | Post-liquefaction | 45 | 75.02 | 46.55 | 50.15 | $H/4$ | 1.2 | 100 | 220 | 120 | – | 1 |
| S10 | Post-liquefaction | 48.8 | 77.23 | 50.22 | 53.80 | $3H/4$ | 1.2 | 100 | 240 | 140 | – | 0.99 |
| S11 (NS) | Post-liquefaction | 48.87 | – | – | 48.87 | No silt | 1.2 | 100 | 200 | 100 | – | 1 |
| S27 | Static | 27.28 | 72.24 | 28.96 | 27.42 | $H/2$ | 1.2 | 100 | 200 | 100 | 0.97 | 0.98 |
| S28 | Static | 28.27 | 71.24 | 29.66 | 36.10 | $H/2$ | 1.2 | 50 | 150 | 100 | 0.96 | 1 |
| S29 | Static | 27.24 | 71.22 | 28.56 | 36.30 | $H/2$ | 1.2 | 150 | 270 | 120 | 0.95 | 0.97 |
| S30(NS) | Static | 28.52 | – | – | 28.52 | No silt | 1.2 | 50 | 170 | 120 | 0.98 | 1 |
| S31(NS) | Static | 28.96 | – | – | 28.96 | No silt | 1.2 | 100 | 220 | 120 | 0.98 | 1 |
| S32(NS) | Static | 29.52 | – | – | 29.52 | No silt | 1.2 | 150 | 270 | 120 | 0.98 | 1 |
| S33(FS) | Static | 30.02 | – | – | 30.02 | Full silt | 1.2 | 25 | 185 | 160 | 0.95 | 1 |
| S34(FS) | Static | 30.54 | – | – | 30.54 | Full silt | 1.2 | 50 | 210 | 160 | 0.96 | 1 |
| S35(FS) | Static | 30.55 | – | – | 30.55 | Full silt | 1.2 | 100 | 280 | 180 | 0.95 | 1 |

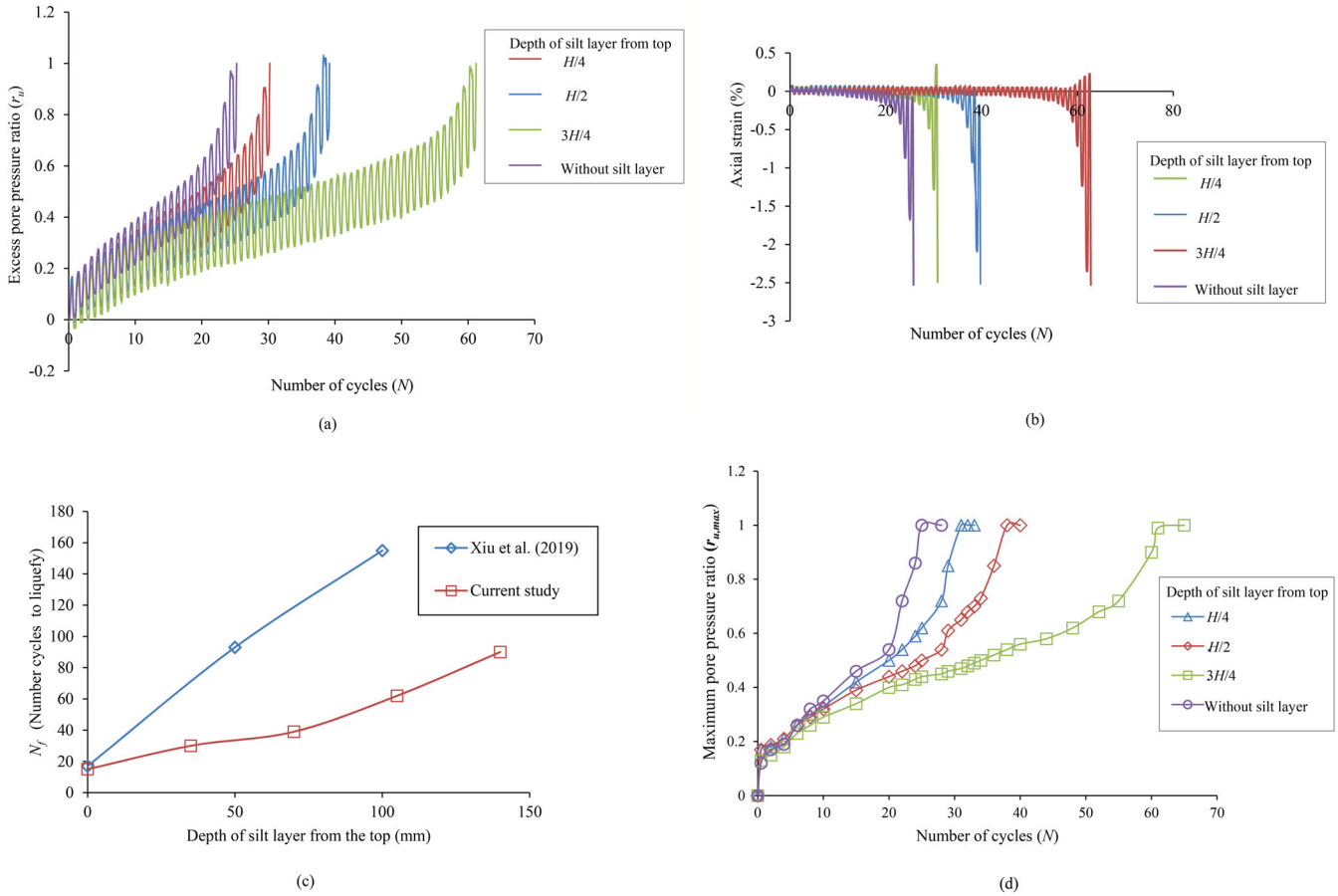



**Figure 4.** Specimens S1, S2, S3 and S4 at an effective confining pressure of 100 kPa: (a) Excess pore pressure ratio (b) axial strain (c) Effect of locations of silt layer for two different studies (d) variation of maximum pore pressure ratio.

where is $u$ = pore pressure and $\sigma'_{3c}$ = effective confining stress.

This criterion was used in this study for the liquefaction failures.

### 3.1. Significance of the location of the silt layer

The silt layer was placed at three different locations, i.e., $H/4$, $H/2$ and $3H/4$ (from the top of the specimen) which was responsible to develop significant changes in the permeability of the overall specimen. Figure 4 shows the pore pressure ratio and axial strain behaviour obtained after cyclically loading the specimens at a confining pressure of 100 kPa. In most of the specimens, it was observed that the initiation of liquefaction was earliest when the silt layer was placed at $H/4$ from the top and longest when at $3H/4$ from the top. The permeability of sand is 2 to 3 times that of the silt layer. As the placement depth of the silt layer increased from $H/4$ to $3H/4$, the pore water flow from the lower sandy strata became lesser. This caused more cycles to liquefy the specimen and break of silt layer. The exact scenario was observed by Xiu et al. (2020) at a

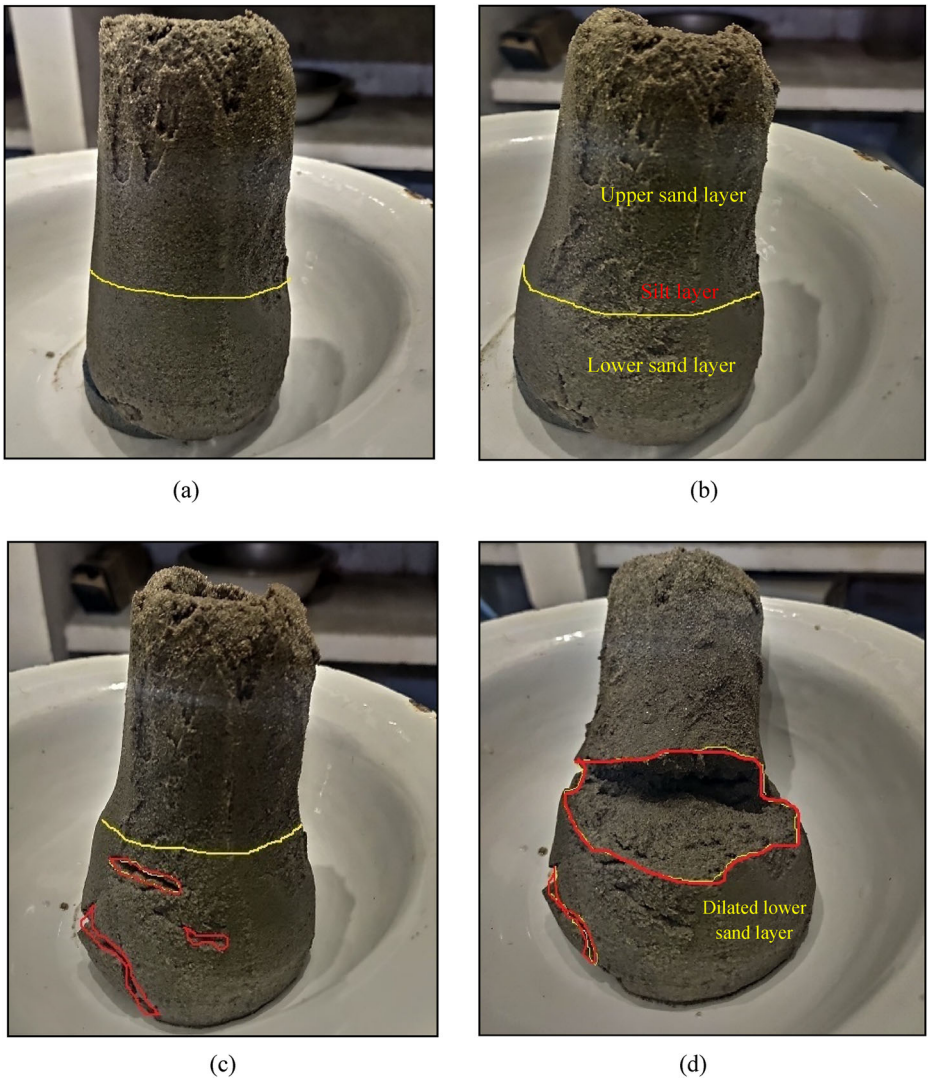
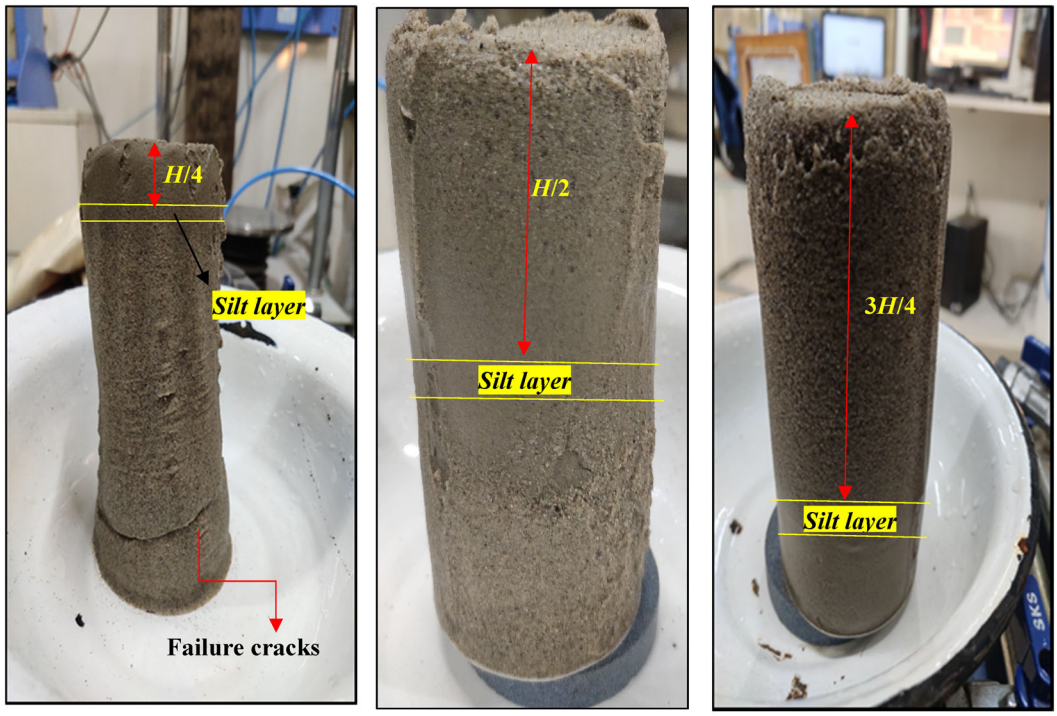



**Figure 5.** Stages of failure of the specimen S2 where yellow line depicts the location of silt layer, red line shows the originating failure surfaces.

confining pressure of 50 kPa. In the stratified specimen S2, the attainment of liquefaction was observed in 38 cycles as shown in Figure 4(a). The silt layer caused hindrance in the path of pore pressure movement, resulting in a different failure mechanism when compared with the homogeneous specimen. At this chosen density, the silt layer provided a limited resistance against liquefaction which causes an early rise of pore pressure in the homogeneous specimen (Figure 4(a)). When the placement depth of the silt layer increased to $3H/4$ the amount of pore-water flow from the lower sand strata becomes much less, causing a stress reduction from water molecules which develops dilation type characteristics. In Figure 4(a), at the beginning of the second loading cycle, the dilation of specimen S4 occurs which is responsible to generate negative pore pressure in the early cycle. Therefore, a negative $r_u$ value of 0.2 was observed. Similarly, the deformation characteristics varied significantly as the 2.5% strain rate was firstly achieved for the nearer location of the silt layer (from the top) as shown in Figure 4(b).

The location of the silt layer had emerged as one of the critical parameters in the liquefaction of stratified sand specimens. In Figure 4(c), the comparative analysis of the current study with the findings of Xiu et al. (2020) has been provided. Both the curve depicts an increasing trend in both the studies. The thickness of the silt layer used by Xiu et al. (2020) was 2.5 times that of the current study which developed significant resistance against liquefaction. Therefore, the number of cycles to liquefy ($N_f$) was higher for Xiu et al. (2020) than that of the present study. It can be observed that there was an 83.3% increase in $N_f$ value when "$D$" increased from 0 to 140 mm. The reason behind this difference is due to the increment in hindrance towards pore flow movement. The comparison of both studies is useful because in both the studies a non-plastic soil mass of nearly similar gradation and with a particle size smaller than 75 microns was used in the interlayers. Figure 4(d) shows the variation of maximum pore pressure ratio ($r_{u,max}$) with a number of cycles where it can be seen that the homogeneous specimen was most vulnerable to liquefaction whereas specimen S4 with silt layer at $3H/4$ was the most stable. The rate of pore pressure generation was highest in specimen S1.

When the silt layer was placed at $H/2$ from the top, the dilation was observed in the sand layer at the bottom, this can be easily visualised in Figure 5, where a separate line (yellow in colour) depicts the location of silt layer and separate lines




(a) Silt layer at H/4 from the top (b) Silt layer at H/2 from the top (c) Silt layer at 3H/4 from the top

**Figure 6.** Images of tested specimens having silt layer at different placement depths.

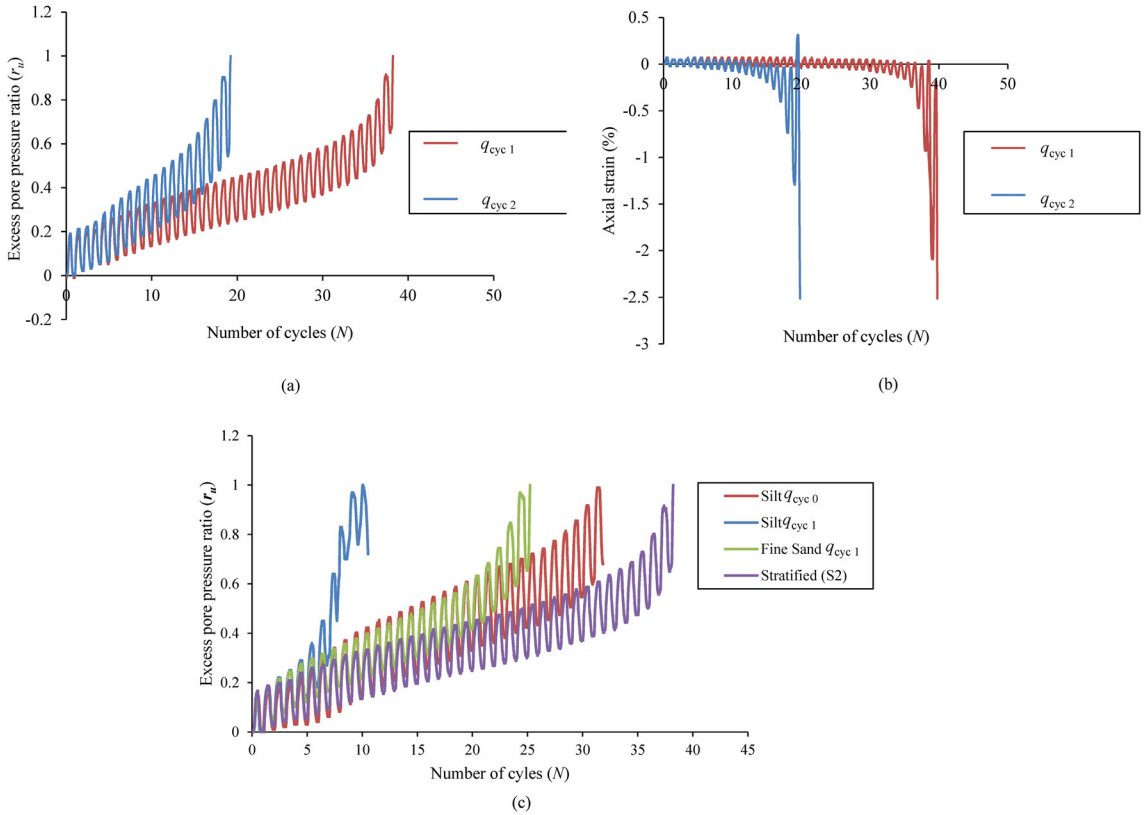



**Figure 7.** Effect of deviator stresses ($q_{cyc0}$ = 28 kPa, $q_{cyc1}$ = 32.5 kPa and $q_{cyc2}$ = 40 kPa ) on the specimens at a confining pressure of 100 kPa where silt layer located at H/2 from the top: (a) Excess pore pressure ratio vs number of cycles ($N$), and (b) Axial strain vs number of cycles ($N$) (c) Comparison with silt specimens.

(red in colour) show generated failure surfaces. In Figure 6, the tested specimens are presented where the silt layer is located in between the yellow lines. This can be confirmed after a closer look at the specimen. The colour of the soil becomes dark and particles look closely spaced at the silt layer location. Also, as the placement depth of silt increases from $H/4$ to $3H/4$, the specimen becomes more stiff and stable. Therefore, cracks can be seen in Figure 6(a) and an intact

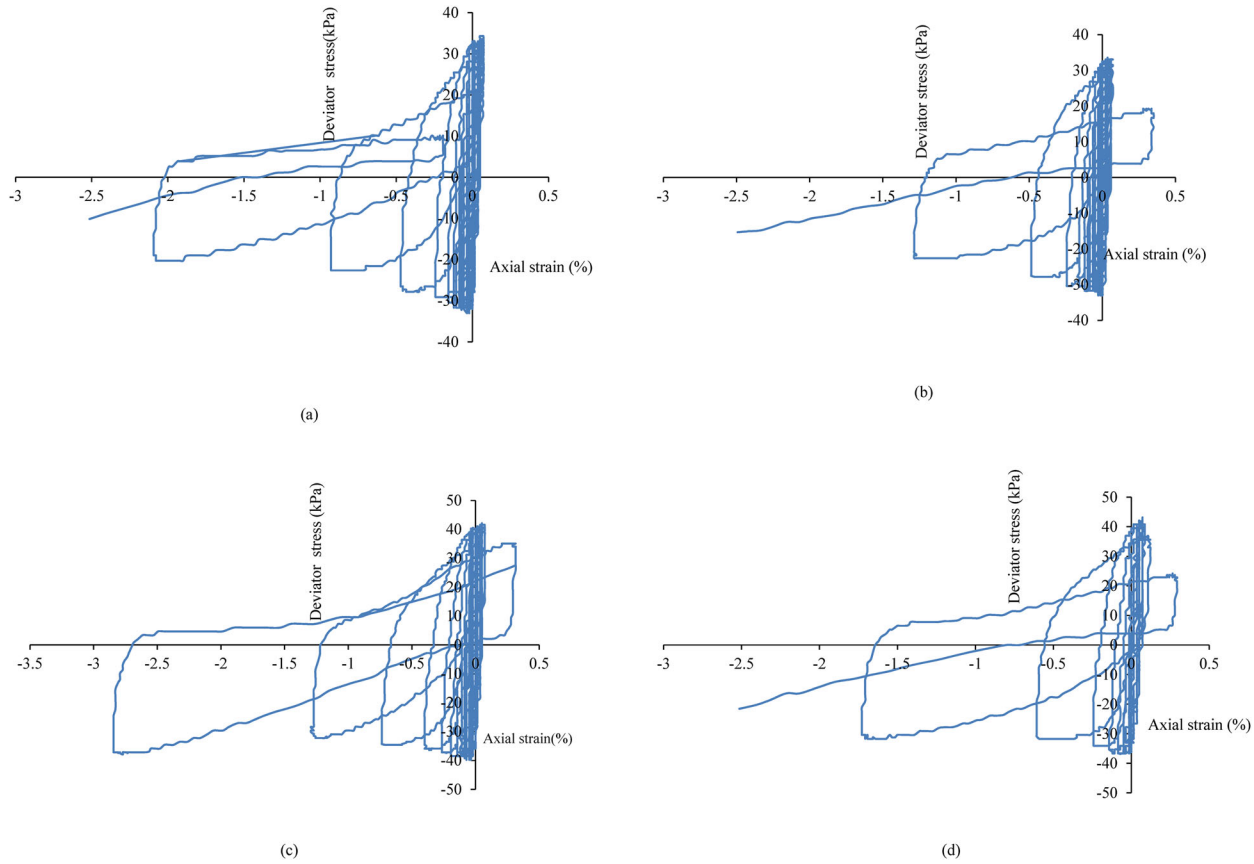



**Figure 8.** Deviator stress vs axial strain plot for specimens isotropically consolidated at a confining pressure of 100 kPa: (a) $q_{cyc\ 1}$, silt layer at $H/2$ from the top (b) $q_{cyc\ 1}$, silt layer at $H/4$ from the top (c) $q_{cyc\ 2}$, silt layer at $H/2$ from the top (d) $q_{cyc\ 2}$, silt layer at $H/4$ from the top.

specimen in Figure 6(c). Silty material spreads due to the application of compressive and extension loading, therefore, dark particles are also visible in the periphery of yellow lines.

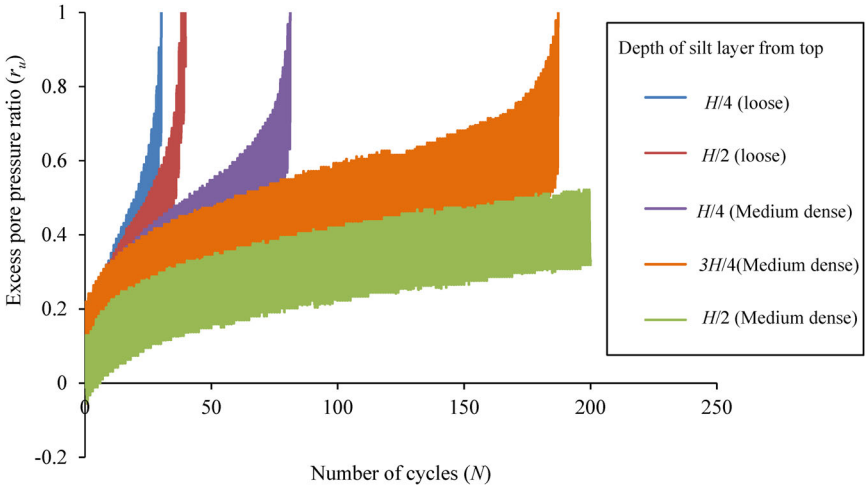
### 3.2. Effect of deviator stresses on the stratified specimen

The higher deviator stress values generate larger excess pore water pressure causing instability and early failures. With the increment in cyclic loading, the development of excess pore water pressure became quick as expected, along with larger strain values. Specimens S2 (silt layer at $H/2$, *CSR* 0.162) and S5 (silt layer at $H/2$, *CSR* 0.2) were considered to understand the effect of two different deviator stresses $q_{cyc1}$ (32.5 kPa) and $q_{cyc2}$ (40 kPa), as shown in Figure 7(a) and (b). In these specimens, the silt layer was located at $H/2$ from the top. The amount of pore pressure developed below the silt layer will be higher for the larger load applied. Here, the increment in pore pressure flow was due to the increment in loading rather than the amount of permeable sand present below the silt layer. At deviator stress of 32.5 kPa, specimen S2 (silt layer at $H/2$, CSR 0.162) reached a strain rate of 2.5% in 39 cycles. After increasing this value of $q_{cyc}$ to 40 kPa, specimen S5 (silt layer at $H/2$, CSR 0.2) touched a 2.5% strain rate in just 19 cycles pointing towards the significant changes in deformation behaviour of the specimen.

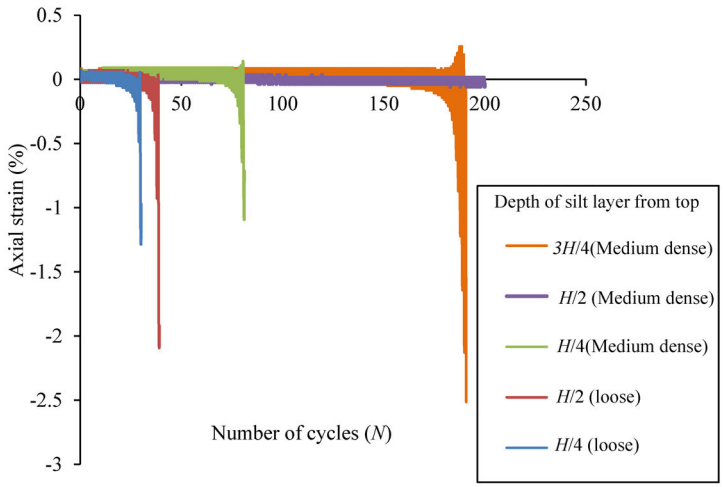
Figure 7(c) comprises of the pore pressure curves of the pure silt specimens, fine sand and stratified specimens at different deviator stresses (Table 2). Under the action of cyclic loads silt specimen S25 liquefied in 10 cycles at deviator stress of 32.5 kPa. Similar results were also obtained by Singh (1996). This might be due to the sudden development of excess pore water pressure due to the absence of any plastic bonds. Meanwhile, in all of the specimens, (S1, S2, S24 and S25), stratified specimen S2 liquefied in the highest number of 38 cycles due to a limited resistance provided by the silt layer.

### 3.3. Stress–strain relationships

The variety of axial strain patterns generated after the application of deviator stresses can be well understood through stress–strain relationships. Figure 8 depicts the stress–strain plots of the stratified specimens (S2, S3, S5, and S7) isotropically consolidated at a confining pressure of 100 kPa. Two different locations for the silt layer at $H/4$ and $H/2$ from the top, were chosen. The specimens S2 and S3 were cyclically loaded at a deviator stress of 32.5 kPa. It took 38 and 30 cycles for samples S2 and S3 respectively to reach the liquefied state. When the deviator stress increased to 40 kPa (Specimen S5 and S7), it resulted in early failures compared to the previous case when $q_{cyc}$ was 32.5 kPa. Due to the increased value of deviator stresses, the increased rate of pore pressure caused $r_u$ to reach unity. Specimens S5 and S7 reached a liquefied state in 19 and 10 cycles, respectively. In Figure 8, it can be observed that initially, the development




(a) Excess pore pressure ratio vs number of cycles ($N$)




(b) Axial strain vs number of cycles ($N$)

**Figure 9.** Effect of density on the isotropically consolidated specimen at a confining pressure of 100 kPa.

of strains was uniformly distributed on both the compressive and extension side. This ensures the correct calibration of the apparatus. With the increase in loading cycles, the instability is triggered on the tension side as the soil is weak in tension rather than compression. When the specimen reaches the liquefied state, most of the strains develop on the tension/extension side and a little portion on the compressive side. Therefore, the axial strain accumulates on the negative (tension/extension) side.

### 3.4. Density effect on the stratified specimen due to silt layer location

In order to look upon the influence of silt layer thickness on medium dense specimens S8, S9, S10, and S11 were chosen (Table 3). Specimen S9 (silt layer at $H$/4 from the top) liquefied after 81 s which was most vulnerable as compared to S8 and S10 (Figure 7(a)). There was a larger development of pore pressure when the silt layer was placed at $H$/4 from the top. The homogenous specimen was least prone to liquefaction failure which can be seen in Figure 7 due to lesser pore pressure development. Specimen S9 has a significant strain rate higher than 2.5% which was absent in the specimens S8, S10 and S11 as shown in Figure 7(b). Meanwhile, specimen S10 showed inconsistent behaviour as it contradicts the suggested theory where specimen with a silt layer nearer to the top becomes more vulnerable.

In Figure 8, the specimens with two different relative densities having a silt layer at a depth of $H$/4 and $H$/2 (from the top) were considered for comparison. For specimen S3 (silt layer at $H$/4), it took only 30 s to reach a liquefied state. Meanwhile, specimen S9 having a silt layer at $H$/4, liquefied in 81 s. After increasing the silt layer depth to $H$/2, there was no sign of liquefaction even after 100 s. As shown in Figure 8(b), no initiation of liquefaction was observed in specimen S8. The typical response of axial strain known as cyclic mobility (Castro 1975) was noticeable in three specimens except specimen no. S8. It was observed that liquefaction occurred in both the homogeneous and stratified loose sand specimens. It was further observed that vulnerability towards liquefaction failure increases at lower relative density. Also, a specimen with a silt layer nearer to the top, has

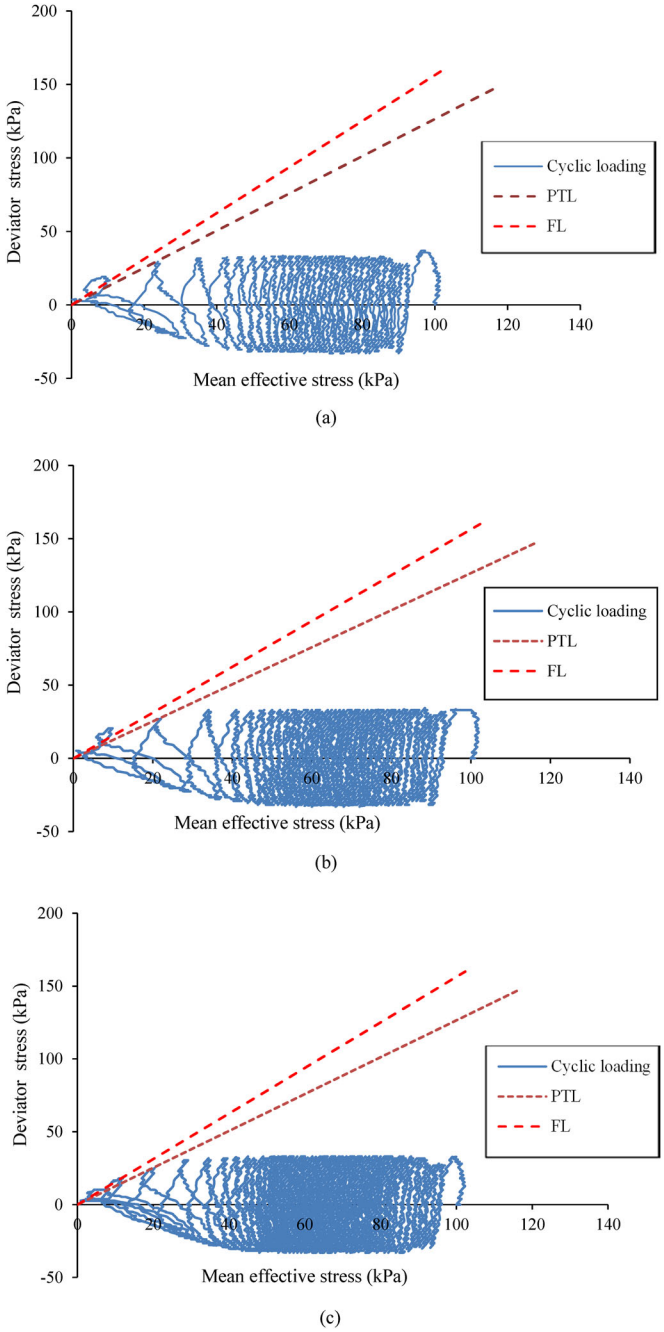



Figure 10. $q$ vs $p'$ plots for the isotropically consolidated specimens at 100 kPa at $q_{cyc\ 1}$, the specimens have silt seam located at depth of (a) $H/4$, (b) $H/2$, and (c) $3H/4$ from the top.

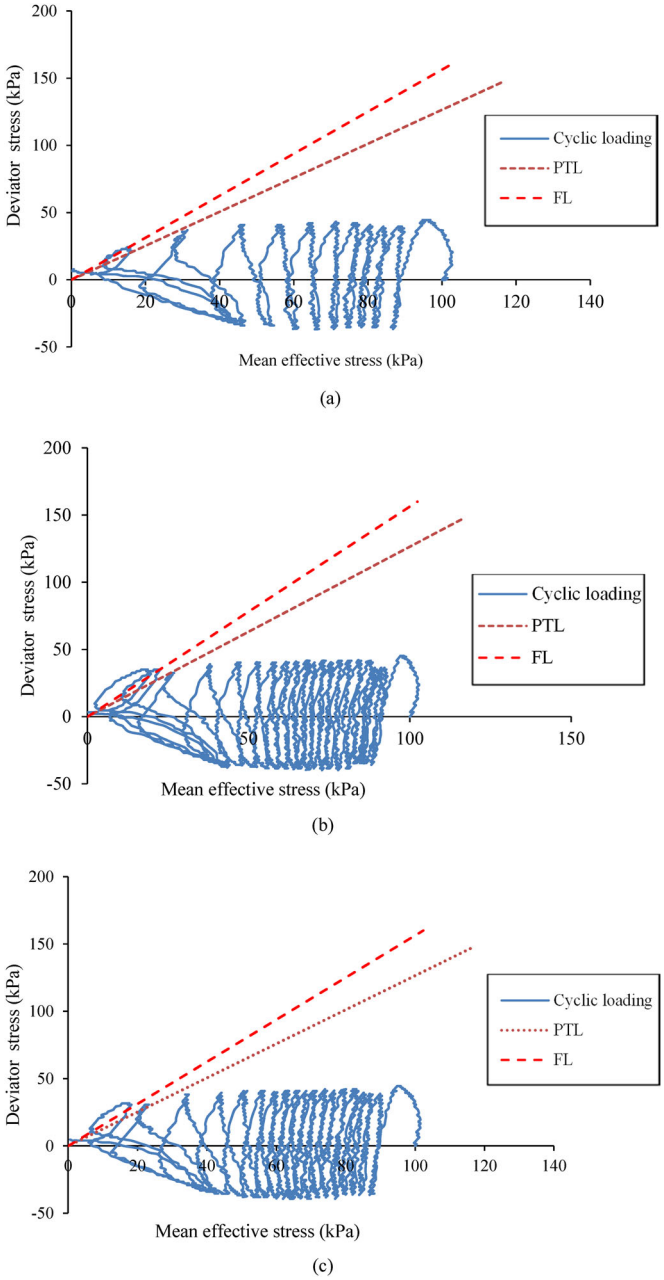
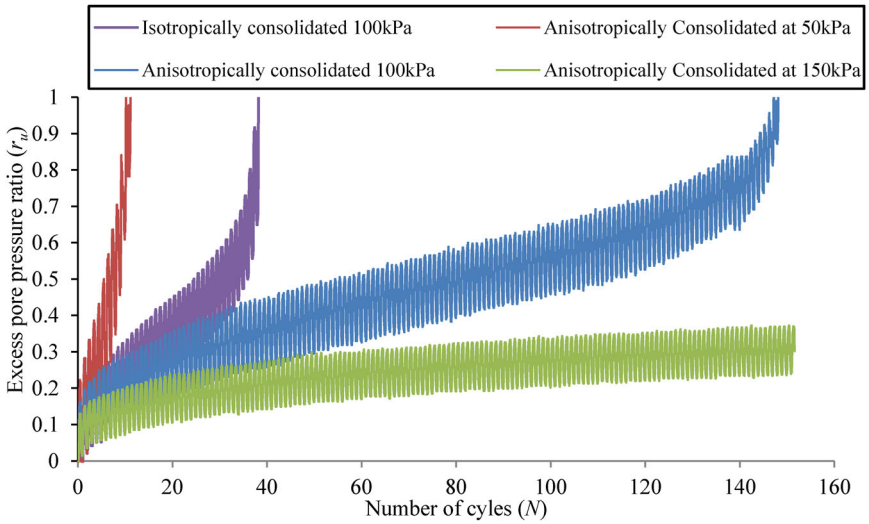


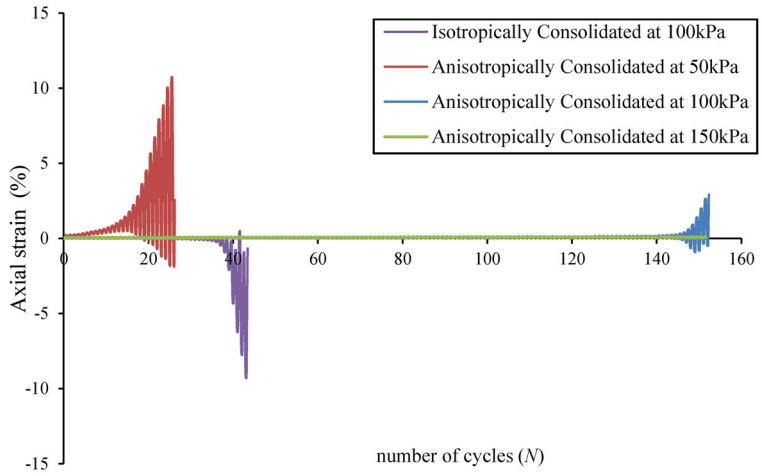

Figure 11. $q$ vs $p'$ plots for the isotropically consolidated specimens at 100 kPa at $q_{cyc\ 2}$, the specimens have silt seam located at depth of (a) $H/4$ (b) $H/2$ (c) $3H/4$ from the top.

a greater chance of early failure as explained earlier. When the packing of soil particles increases, the increased grain to grain contact will counter the development of crucial pore water flow inside the specimen. Therefore, the amount of developed excess pore pressure will be minimal which amounts to delayed liquefaction, i.e., more number of stress cycles to liquefy. The denseness of sand particles enhanced the time to reach the 2.5% strain rate from 39 to 82 s (compared to specimens S3 and S9).

In order to investigate the influence of silt layer presence with changes in density, specimens S2, S3, S8, S9 and S10 were chosen (Table 2). It can be observed that vulnerability towards liquefaction failure increases at lower relative density. Therefore, in loose specimen S3 (silt layer at $H/4$), it took only 30 cycles to reach a liquefied state (Figure 9). Meanwhile, medium dense specimen S9 having a silt layer at $H/4$, liquefied in 81 cycles. With an increase in density of the specimen, enhancement in grain to grain contact recedes the rate of excess pore pressure development inside the specimen. Therefore, the amount of developed excess pore pressure will be minimal which amounts to delayed liquefaction, i.e., more cycles to liquefy (Figure 9(a)). After increasing the silt layer depth to $H/2$ (specimen S8), there was no sign of liquefaction even after 100 s. The specimen having a silt layer nearer to the top has a greater chance of early failure which was also observed in loose conditions. Additionally, specimen S10 showed inconsistent behaviour which liquefied earlier than that of specimen S8. This contradicts the




(a) Excess pore pressure ratio vs number of cycles ($N$)




(b) Axial strain vs number of cycles ($N$)

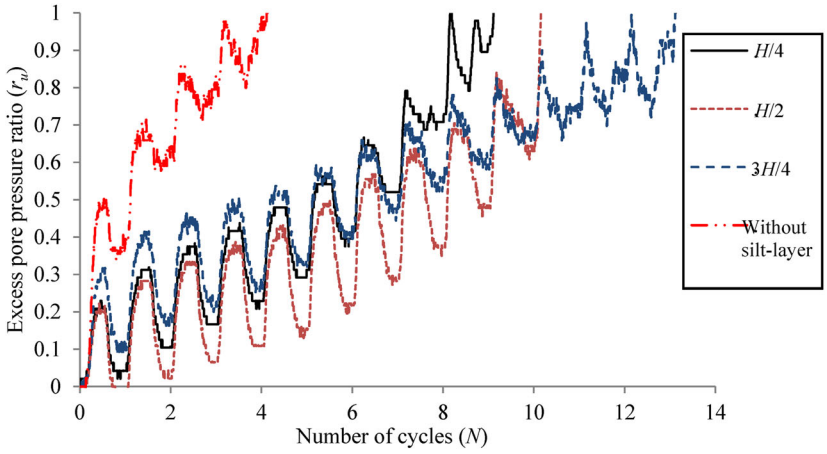
**Figure 12.** Influence of effective confining stresses on the isotropic and anisotropically consolidated specimens when silt layer located at 70 mm from the top.

previous observation in a loose state where the specimen has a silt layer nearer to the top liquefied in early cycles. The reason behind this unique observation is attributed to the denseness of lower and upper sandy strata which causes lower pore pressure generation in the whole specimen S8 than that of S10. The typical response of axial strain known as cyclic mobility (Castro 1975) was noticeable in three specimens except specimen no. S8. The denseness of sand particles delayed the cycles to reach 2.5% single strain amplitude from 39 to 82 cycles (specimens S3 and S9). Specimen S9 (silt layer at $H/4$ from the top) liquefied after 81 cycles which were most vulnerable as compared to S8 and S10 (Figure 9(a)). Specimens S9 and S10 have significant strain development which was absent in the specimens S8 (Figure 9(b)).

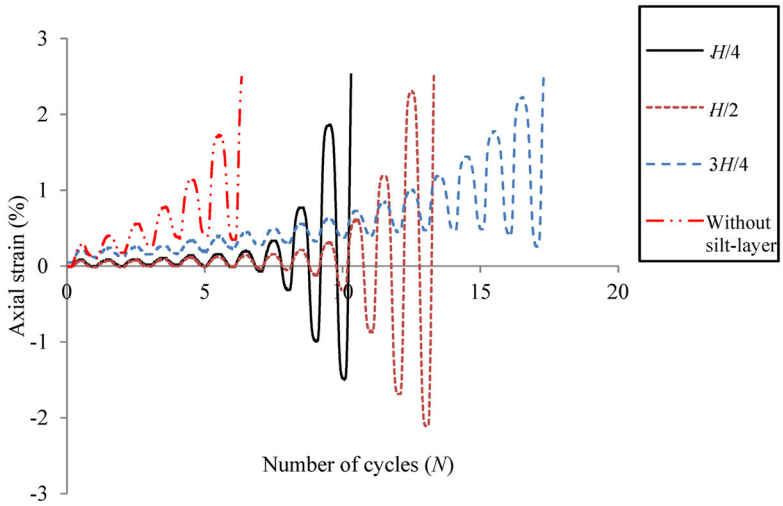
### 3.5. Deviator stress (q) vs. mean effective stress (p′) variation in stratified specimens

Several researchers (e.g., Vaid and Chern 1983, Vaid, Stedman, and Sivathayalan 2001) considered phase transformation line (PTL) and failure line (FL) in their work to analyse the failure mechanism from the critical state perspective. In Figure 10, the cyclic loading is represented using PTL and FL obtained from the monotonic tests (Ishihara, Tatsuoka, and Yasuda 1975). The importance of PTL was significant for dilative specimens as suggested by Vaid, Stedman, and Sivathayalan (2001). In the present work, two significant lines PTL and FL were obtained through the monotonic tests performed on the stratified specimens where the silt layer was located at a depth of $H/2$ from the top under the effective confining pressures of 50 kPa, 100 kPa and 150 kPa. Figure 10 shows the deviator stress ($q$) vs mean effective stress ($p'$) plots for three different locations of the silt layer at $H/4$, $H/2$ and $3H/4$ from the top of the specimen. The specimens S2, S3, and S4 have been considered in Figure 10. It can be easily visualised from Figure 10 that when the silt layer was nearer to the top, failure occurred in fewer cycles as explained earlier.

As the specimen reaches the PTL, the instability triggers then after touching FL, sudden failure occurred within a few cycles. This behaviour can easily be recognised from




(a) Excess pore pressure ratio vs number of cyles ($N$)




(b) Strain rate vs time

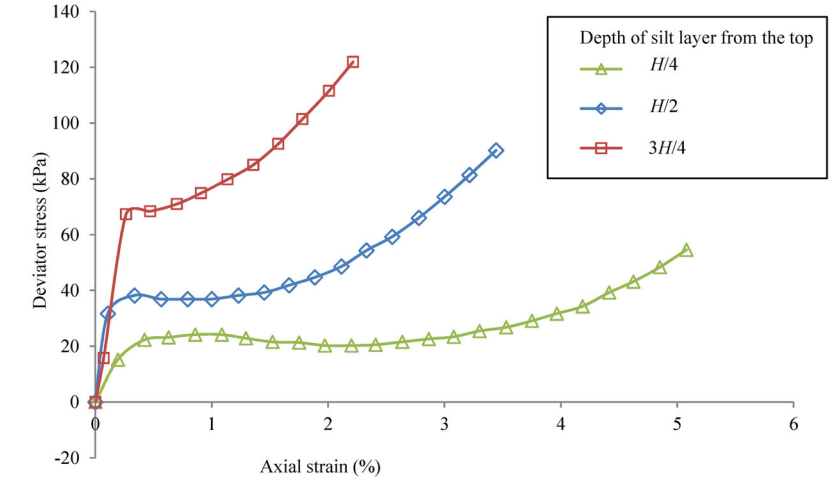
**Figure 13.** A comparison between homogeneous and stratified specimens at a confining pressure of 50 kPa.

Figure 10. In the previous sections, it was clearly observed that after the application of higher $q_{cyc}$, the failure occurs in fewer cycles. Figure 11 shows the $q$ vs $p'$ plots for the specimens S5, S6, and S7. The $q_{cyc}$ value provided was 40 kPa which caused the development of excess pore water pressure at a higher rate than that of the earlier case where $q_{cyc1}$ was 32.5 kPa.

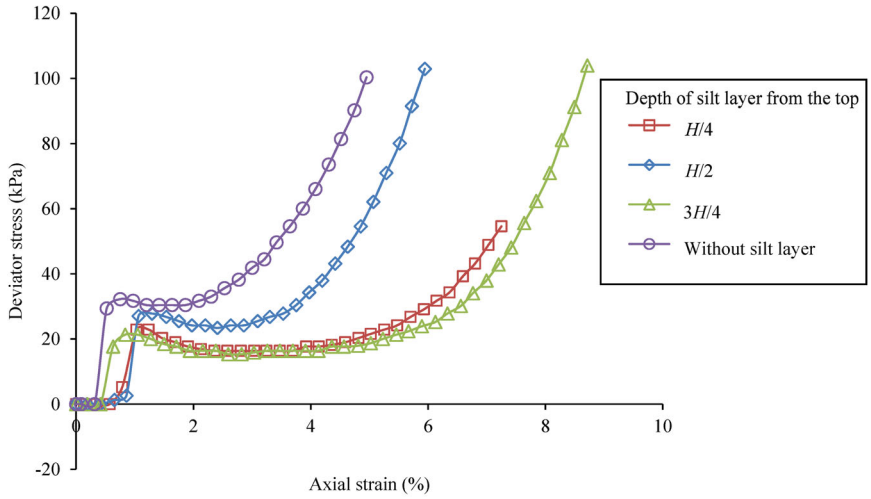
### 3.6. Influence of initial stress on the liquefaction characteristics of stratified soils

The isotropically consolidated specimens are incapable of simulating the presence of initial stresses on any struture. Therefore, few tests were performed under anisotropically consolidated state with an initial static stress ($q_s$) of 20 kPa under different confining pressures, i.e., 50 kPa, 100 kPa and 150 kPa as shown in Table 2.

The confining stresses are largely responsible for either mitigating or triggering failure in a soil mass. Specimens S2, S14, S15, S16, and S17 were considered for the comparative study of pore pressure ratio and strain rate as shown in Figure 12. In Figure 12(a) and (b), the effect of different effective confining pressures on both isotropic and anisotropically consolidated specimens can be clearly seen when the silt layer was located at $H/2$ from the top. As noticed in previous sections, in the absence of initial stresses, specimen S2 attained the pore pressure ratio as unity in just 38 cycles. The liquefaction failure in an anisotropically consolidated specimen is considered when the axial strain exceeds 2.5% as suggested by NRC (1985). The failure was earliest in specimen S16 in 13 cycles, due to low confining stresses. This shows that the presence of initial static shear stress was also unable to provide any mitigating effect against liquefaction at low confining stress. Specimen S15 reached liquefaction in 151 cycles but specimen S17 had not shown any sign of liquefaction even up to 500 cycles. Thus, it can be said that high confining stresses were good enough to stop the generation of sudden dissipated pore water pressure and showed minimal strain development. After the application of initial stress (specimen S15) of about 20 kPa, liquefied state was achieved in 151 cycles. With an increase in cyclic deviator stress to 40 kPa, even after the application of initial stress, there was a rapid generation of pore water pressure which caused liquefaction of the specimen S19 in 69 cycles (Table




(a) Loose specimen




(b) Medium dense specimen

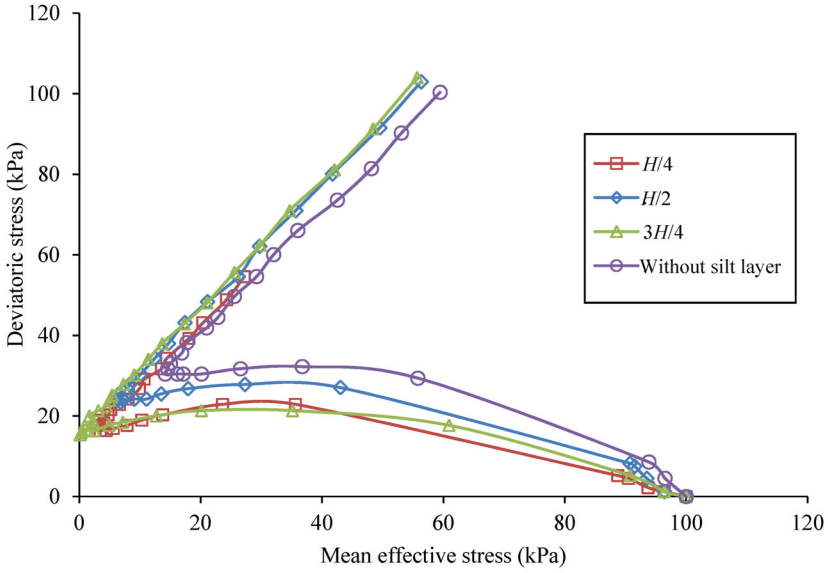
**Figure 14.** Stress vs strain plot of the post-liquefied stratified specimens.

2). The higher amount of excess pore pressure generation leads to the early failures at higher deviator stress values.

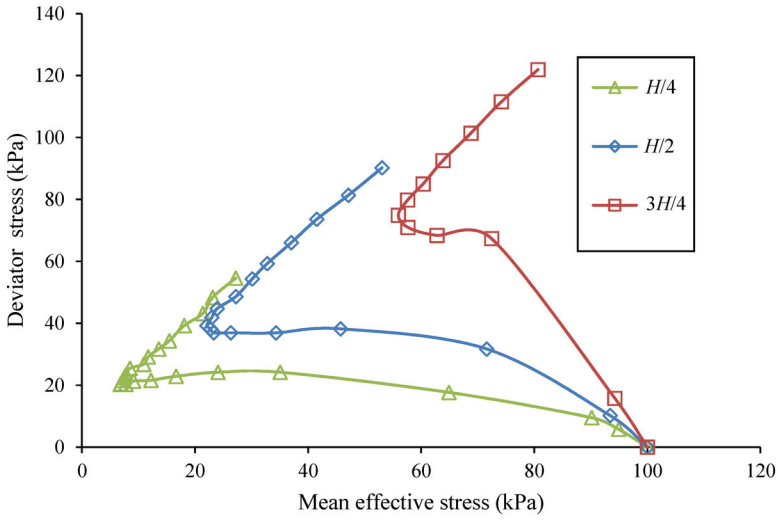
In order to understand the effect of low confining stresses on soil stratification, cyclic loading was applied on the specimens S16, S21, S22, and S23 at an effective confining stress of 50 kPa (Table 2). The liquefaction failure criteria were considered based on the axial strain theory suggested by NRC. It can be observed in Figure 13 that due to lower confinement most of the specimens liquefied within 18 cycles at cyclic deviator stress of 12.5 kPa. There was only a 63% increase in $N_f$ value when placement depth increased from $H/4$ to $3H/4$ which shows that the placement depth of the silt layer played a less insignificant role in governing liquefaction initiation and related failures. The amount of porewater flow in the lower strata may have increased due to low confining stresses which results in the early failure of all the specimens (Figure 13(a)). The rate of strain development for all the specimens is presented in (Figure 13(b)). Due to a certain amount of resistance developed by the silt layer, the stratified specimens attained 2.5% axial strain in higher cycles than that of the homogeneous specimen S23. A similar phenomenon was also observed at confining stress of 100 kPa. These results were in agreement with that of the shaking table tests on silt interlayered sand, performed by Ozener, Ozaydin, and Berilgen (2009) where the conclusions were made about the lesser effect of the silt layer on the development of stable water film at low confining stresses under strong ground shaking.

### 3.7. Post-liquefaction behaviour of the stratified specimen

Stress–strain behaviour of soil mass after the occurrence of liquefaction gives a clear understanding of the earthquake-induced settlements and failure mechanism of post-liquefied soil (Kokusho and Kojima 2002). In this study, the post-liquefaction behaviour of stratified soil mass gave significant results corresponding to the changes in silt layer location which will help in the assessment of liquefaction phenomenon nearby marine soil or alluvial plains. In the present study, after dissipating the generated excess pore water pressure after the completion of the undrained cyclic loading, monotonic loading was applied to the stratified specimens




(a) Medium dense specimen




(b) loose specimen

**Figure 15.** $q$ vs $p'$ plot in post-liquefied state for the stratified specimen.

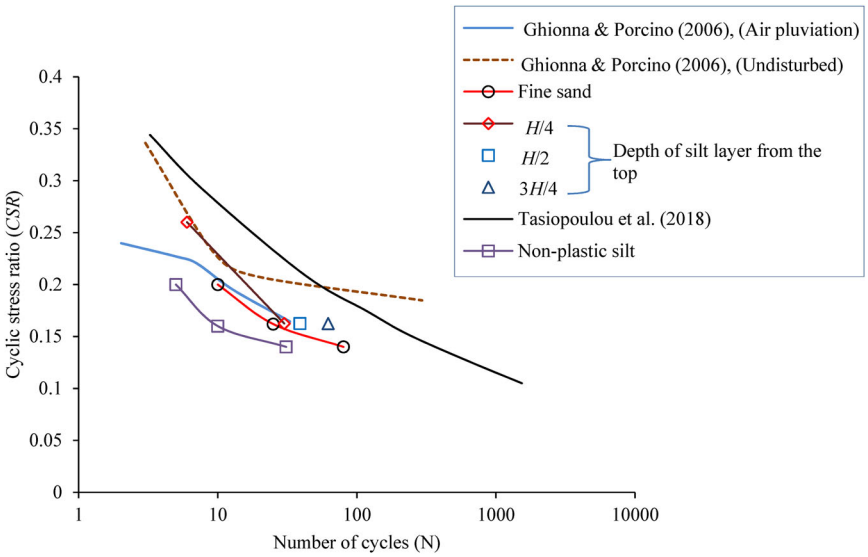



**Figure 16.** Cyclic liquefaction resistance curves for the present study and their comparison with previous studies.

which replicate the post liquefaction phenomenon (Sitharam, Vinod, and Ravishankar 2009).

The stress–strain behaviour of the post-liquefied stratified specimens was obtained after applying the monotonic loading on the liquefied specimens at a strain rate of 1.2 mm/min, as shown in Figure 14. In the present study, specimens S5, S6, S7, S8, S9, S10, and S11 were selected for the stress–-strain analysis of post-liquefied stratified soil specimens. All the specimens were isotropically consolidated at a confining pressure of 100 kPa. Specimens S5, S6 and S7 were provided with deviator stress of 40 kPa in the pre-liquefaction stage. The effect of the location of a less permeable silt layer on the stress–strain behaviour of post-liquefied specimens is visible in Figure 14. The amount of pore pressure developed under the silt layer depended on its location or depth from the top. The deformation characteristics, due to the occurrence of liquefaction under various conditions, caused significant changes in the post-liquefaction states. As explained in earlier sections, due to large pore water flow, the permeable sand layer generated a higher amount of excess pore pressure as the depth of the silt layer decreased. From Figure 14(a), it can be observed that in medium dense specimens, the location of the silt layer played an insignificant role in the variation in stress–strain behaviour of post-liquefied soil mass. But the presence of a silt layer was quite disastrous, which resulted in a lower stress value than that of the homogeneous specimens under post-liquefied state. Specimens S9 and S10 followed nearly the same trend of stress–strain plots as shown in Figure 14(a). At a particular axial strain rate of 1.2 mm/min, specimen S6 took the highest stress while specimen S7 took the lowest stress under post liquefied state (Figure 14(b)). The phenomenon was different when the specimens were analysed at higher relative density with a $q_{cyc1}$ 32.5 kPa as shown in Figure 14(a). Here specimen S11, due to the absence of a silt layer, took a higher amount of stress for a particular strain value than the other stratified specimens. Inconsistently, specimen S8 carried a higher amount of stress value than S10. This was not in agreement with previous results, as here the silt layer was nearer to the top. This might be due to the rearrangement of soil particles during the loading or excess pore water pressure dissipation.

The deviator stress ($q$) vs mean effective stress ($p'$) plots of the specimens S8, S9, S10, and S11 were aligned in nearly the same manner as shown in Figure 15(a). Meanwhile, as shown in Figure 15(b), specimens with a $D_r$ = 30% with deviator stress of 40 kPa at pre-liquefaction, showed a noticeable difference in $q$ vs $p'$ plot. Also, the plot of specimens S9 and S10 are superimposing each other. From the obtained results, it can be concluded that under a post-liquefied state, the location of the silt layer has an insignificant effect on $q$ vs $p'$ plots but the presence of the silt layer played an important role in medium dense specimens.

### 3.8. Comparison of liquefaction resistance curves

Tasiopoulou et al. (2019) performed several cyclic triaxial tests and numerically simulated the soil stratification located nearby intertidal structures, marine conditions and alluvial sites containing clay laminations. The liquefaction resistant curve of the present work has also been plotted along with the study of Tasiopoulou et al. (2019) as shown in Figure 16 where it can be observed that the trendline variation of the present work is following nearly a similar pattern as that of Tasiopoulou et al. (2019) which ensures the efficacy of the present study. The cyclic liquefaction resistance is observed as *CSR* vs the number of loading cycles to develop a 2.5% maximum axial strain. Figure 16 shows the cyclic liquefaction resistance curves for air pluviation (AP) and undisturbed specimens prepared by Ghionna and Porcino (2006), where it can be clearly seen that the AP method leads to lower liquefaction resistance values. Here, the specimens were isotropically consolidated at an effective confining pressure of 40 kPa. In the present study, a consolidation stress ratio ($K_c$) of stratified specimens was considered as 0.83 at an effective confining pressure of 100 kPa where consolidation stress ratio, $K_c = \sigma'_{3c}/\sigma'_{1c}$. The liquefaction resistance curve of pure silt specimen has also been compared with the other specimens of the present study and previous research work. It can be observed that the non-plastic silty specimens are more vulnerable to liquefaction than fine sand. A similar observation was also suggested by Singh (1996).

Meanwhile, the loading conditions, effective confining stresses, and soil type were different from Ghionna and Porcino (2006) but the suitable locations of points were obtained for cyclic resistant values in the current study. From Figure 16, it can easily be understood that the silt layer presented near the top has a lower value of liquefaction resistance. For a constant *CSR* value, the number of cycles required is lower for the specimen having a silt layer at a nearer location from the top. The investigations made by Tasiopoulou et al. (2018) on the presence of clay laminates considered in-situ specimens with clay laminates of about 1-2 cm in a sandy specimen of size 140 mm. In the present study, a silt layer of nearly the same thickness as that of the study by Tasiopoulou et al. (2018) was considered in the preparation of reconstituted stratified specimens with a same height of 140 mm Therefore, the comparison of the present study provided an insight into the results of reconstituted and in-situ soil specimens. Additionally, Ghionna and Porcino (2006) performed stress-controlled cyclic triaxial tests on the reconstituted specimens using different modes of specimen preparation, air pluviation and water sedimentation technique. A comparison with their study will be helpful in observing a deviation or agreement in results corresponding to the specimen preparation technique.

## 4. Conclusions

The significance of the presence of an impermeable silt layer in stratified soil strata attracted several researchers due to its devastating effects in marine slopes, foundations nearby alluvial plains and in other intertidal sites. In the present study, the cyclic behaviour of stratified soil mass was thoroughly

investigated in both pre and post-liquefaction conditions. Some of the important outcomes of this study are presented below:

1. After varying the location of the less permeable silt layer, strong pieces of evidence supporting the usefulness of the silt layer present in the stratified specimens have been observed under the action of cyclic loading. A specimen with a silt layer nearer to the top, i.e., at $H/4$ from the top was most vulnerable in most of the stratified cases. Based on the results obtained, it was observed the homogeneous specimen liquefied earlier due to a certain amount of resistance provided by the presence of silt layer.
2. Due to the increment in the amplitude of deviator stresses from 32.5 kPa to 40 kPa, pore pressure generated at a higher rate resulted in the early failure of the specimens. It was observed that with an increase in relative density, the number of failure cycles enhanced from 30 cycles to 81 cycles under the same loading conditions.
3. The overlapping of cyclic loading, phase transformation line (PTL) and failure line (FL) suggested a clear cut point where the instability triggers resulted in the liquefaction failure. The presence of a silt layer and its location emerged as a critical parameter in governing the liquefaction failures.
4. The liquefaction resistance of stratified specimens was significantly influenced by the presence of the silt layer even under the action of initial static shear stresses. Resistance towards liquefaction increased with an increment in confinement from 50 kPa to 100 kPa.
5. Post-liquefaction study also revealed that the presence of silt layer/less permeable soil also leads to the early failure under the post-liquefied state. In all of the specimens considered, homogeneous specimens carried the highest stress of 32.5 kPa. In a medium dense specimen, the location of the silt layer was insignificant but its presence played a key role in deciding the stress–strain behaviour in post-liquefaction.
6. The present work considers the use of a single silt layer to replicate stratified conditions nearby alluvial plains or marine deposits. In the future, stratified specimens with multiple layered systems can be utilised.

## List of symbols

| | |
|---|---|
| $H$ | Height of the specimen |
| $D$ | Placement depth of silt layer from the top |
| $B$ | Skempton's pore pressure coefficient |
| $N_f$ | Number of loading cycles to reach liquefaction stage |
| $e_e$ | Equivalent void ratio |
| $D_r$ | Relative density |
| $CSR$ | Cyclic stress ratio |
| $FL$ | Failure line |
| $PTL$ | Phase transformation line |
| $D_{r,m}$ | Modified relative density |
| $D_{r,sand,1}$ | Relative density of upper sand layer |
| $D_{r,sand,2}$ | Relative density of lower sand layer |
| $D_{r,silt}$ | Relative density of silt layer |
| $q$ | Deviator stress |
| $p'$ | Mean effective stress |
| $e_{max,m}$ | Modified maximum void ratio |
| $e_{min,m}$ | Modified minimum void ratio |
| $R_{m,i}$ | Percentage of the dry weight of $i_{th}$ layer in the stratified soils |
| $q_s$ | Initial static deviator stress |
| $q_{cyc}$ | Cyclic deviator stress |
| $\varepsilon$ | Axial strain |
| $\alpha$ | Initial shear stress ratio |
| $r_u$ | Excess pore pressure ratio |
| $\sigma_c'$ | Effective confining stress |
| $\sigma'_{1c}$ | Major principal stress |
| $\sigma'_{3c}$ | Minor principal stress |
| $\sigma'_{nc}$ | Effective normal stress |
| $u$ | Pore pressure |
| $K_c$ | Consolidation stress ratio |
| $q_{cyc}$ | Cyclic deviator stress |

## Acknowledgments

All the experiments in this study have been performed in the Geotechnical Engineering Laboratory, Indian Institute of Technology, Roorkee, India.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the Ministry of Human Resource Development, Government of India, New Delhi (grant no. MHR 002).

## ORCID

Arpit Jain http://orcid.org/0000-0002-7220-4040
Satyendra Mittal http://orcid.org/0000-0001-6216-9205
Sanjay Kumar Shukla http://orcid.org/0000-0002-4685-5560

## References

Amini, Z. A., and A. C. Trandafir. 2008. Post-Liquefaction Shear Behavior of Bonneville Silty-Sand. *In Geotechnical Earthquake Engineering and Soil Dynamics*, vol. IV, 1–9. USA: Geotechnical Special Publication. doi:10.1061/40975(318)95.

Arulanandan, K., and R. F. Scott. 1993. Verification of Liquefaction Analysis by Centrifuge Studies. In Proceedings of the International conference on the verification of Numerical Procedures for the analysis of Soil Liquefaction problems, University of Cailfornia, Davis, vol.1, October. Davis, CA: Balkema.

ASTM. 2013. *Standard Test Method for Load Controlled Cyclic Triaxial Strength of Soil. ASTM D5311/D5311*. M. West Conshohocken, PA: ASTM.

Balakrishnan, A., and Kutter. B. L. 1999. Settlement, Sliding and Liquefaction Remediation of Layered Soil. *Journal of Geotechnical Engineering* 125 (11): 968–978.

Belkhatir, M., A. Arab, N. Della, H. Missoum, and T. Schanz. 2010. Influence of Inter-Granular Void Ratio on Monotonic and Cyclic Undrained Shear Response of Sandy Soils. *Comptes Rendus Mécanique* 338 (5): 290–303. doi:10.1016/j.crme.2010.04.002.

Brennan, A. J., and S. P. G. Madabhushi. 2005. Liquefaction and Drainage in Stratified Soil. *Journal of Geotechnical and Geoenvironmental Engineering* 131 (7): 876–884. doi:10.1061/(ASCE)1090-0241(2005)131:7(876).

Butterfield, K., and M. D. Bolton. 2003. Modelling Pore Fluid Migration in Layered Liquefied Soils. In *Proceedings of the Pacific Conference on Earthquake Engineering*. Christchurch, New Zealand: New Zealand Society for Earthquake Engineering.

Castro, G. 1975. Liquefaction and Cyclic Mobility of Saturated Sands. *Journal of the Geotechnical Engineering Division* 101 (6): 551–569. doi:10.1061/AJGEB6.0000173.

Choi, S. K., M. J. Lee, H. Choo, M. T. Tumay, and W. Lee. 2010. Preparation of a Large Size Granular Specimen Using a Rainer System with a Porous Plate. *Geotechnical Testing Journal* 33 (1): 45–54.

Cresswell, A., M. E. Barton, and R. Brown. 1999. Determining the Maximum Density of Sands by Pluviation. *Geotechnical Testing Journal* 22 (4): 324–328.

Dash, H. K. 2008. Undrained Cyclic and Monotonic Response of Sand-Silt Mixtures. Ph.D. thesis, Indian Institute of Science, Bangalore.

Ecemis, N. 2021. Experimental and Numerical Modeling on the Liquefaction Potential and Ground Settlement of Silt-Interlayered Stratified Sands. *Soil Dynamics and Earthquake Engineering* 144: 106691. doi:10.1016/j.soildyn.2021.106691.

Elgamal, A. W., R. Dobry, and K. Adalıer. 1989. Small Scale Shaking Table Tests of Saturated Layered Sand-Silt Deposits. *In 2nd U.S. Japan Workshop on Soil Liquefaction*, Rep. No. 89-0032, 233–245. Buffalo: NCEER.

Fiegel, G. L., and B. L. Kutter. 1994. Liquefaction Mechanism for Layered Soils. *Journal of Geotechnical Engineering* 120 (4): 737–755. doi:10.1061/(ASCE)0733-9410(1994)120:4(737).

Ghionna, V. N., and D. Porcino. 2006. Liquefaction Resistance of Undisturbed and Reconstituted Samples of a Natural Coarse Sand from Undrained Cyclictriaxial Tests. *Journal of Geotechnical and Geoenvironmental Engineering* 132 (2): 194–202. doi:10.1061/(ASCE)1090-0241(2006)132:2(194).

Indian Standard Code. 2002. Classification and identification of soils for general engineering purposes. IS 1498-1970.

Ishihara, K., F. Tatsuoka, and S. Yasuda. 1975. Undrained Deformation and Liquefaction of Sand under Cyclic Stresses. *Soils and Foundations* 15 (1): 29–44. doi:10.3208/sandf1972.15.29.

Jain, A., S. Mittal, and S. K. Shukla. 2021. Effect of Specimen Preparation Technique on the Post-Liquefaction Behaviour of Fine Sand under Cyclic Loading. *Indian Geotechnical Journal* 51 (5): 1026–1037. doi:10.1007/s40098-021-00505-y.

Jeng, D. S. 2013. *Porous Models for Wave-Seabed Interactions*, 289. Heidelberg: Springer.

Jia, M., and B. Wang. 2012. Liquefaction Testing of Stratified Sands Interlayered with Silt. *Applied Mechanics and Materials* 256–259: 116–119. doi:10.4028/www.scientific.net/AMM.256-259.116.

Kaiser, A., C. Holden, J. Beavan, D. Beetham, R. Benites, A. Celentano, D. Collett, et al. 2012. The Mw 6.2 Christchurch Earthquake of February 2011: Preliminary Report. *New Zealand Journal of Geology and Geophysics* 55 (1): 67–90. doi:10.1080/00288306.2011.641182.

Karakan, E., and S. Altun. 2018. Determination of the Cyclic Properties of Silty Sands. Chap. 012 in *Handbook of Research on Trends and Digital Advances in Engineering Geology*. Hershey PA, USA: United States of America by IGI Global Engineering Science Reference. doi:10.4018/978-1-5225-2709-1.ch01.

Karakan, E., T. Tanrinian, and A. Sezer. 2019. Cyclic Undrained Behavior and Post Liquefaction Settlement of a Nonplastic Silt. *Soil Dynamics and Earthquake Engineering* 120: 214–227. doi:10.1016/j.soildyn.2019.01.040.

Kargar, S. H. R., H. Salehzadeh, and H. Shahnazari. 2016. Post-Cyclic Behavior of Carbonate Sand of the Northern Coast of the Persian Gulf. *Marine Georesources & Geotechnology* 34 (2): 169–180. doi:10.1080/1064119X.2014.987890.

Kokusho, T. 1999. Water Film in Liquefied Sand and Its Effects on Lateral Spread. *Journal of Geotechnical and Geoenvironmental Engineering* 125 (10): 817–826. doi:10.1061/(ASCE)1090-0241(1999)125:10(817).

Kokusho, T., and K. Fujita. 2001. Water Films Involved in Postliquefaction Flow Failure in Niigata City during the 1964 Niigata Earthquake. In *Proceedings of the Fourth International Conference on Recent Advances in Geotechnical Earthquake Engineering and Soil Dynamics* Paper No. 5: 38. Rolla, MO: University of Missouri-Rolla.

Kokusho, T., and T. Kojima. 2002. Mechanism for Postliquefaction Water Film Generation in Layered Sand. *Journal of Geotechnical and Geoenvironmental Engineering* 128 (2): 129–137. doi:10.1061/(ASCE)1090-0241(2002)128:2(129).

Konrad, J. M., and S. Dubeau. 2003. Cyclic Strength of Stratified Soil Samples. In *Submarine Mass Movements and Their Consequences. Advances in Natural and Technological Hazards Research*, ed. J. Locat, J. Mienert, L. Boisvert, vol 19, 47–57. Dordrecht: Springer.

Kramer, S. L. 1996. *Geotechnical Earthquake Engineering*. Upper Saddle River, NJ: Prentice-Hall.

Krim, A., A. Arab, M. Chemam, A. Brahim, M. Sadek, and I. Shahrour. 2019. Experimental Study on the Liquefaction Resistance of Sand–Clay Mixtures: Effect of Clay Content and Grading Characteristics. *Marine Georesources & Geotechnology* 37 (2): 129–141. doi:10.1080/1064119X.2017.1407974.

Ladd, R. S. 1974. Specimen Preparation and Liquefaction of Sands. *Journal of the Geotechnical Engineering Division* 100 (GT10): 1180–1184. doi:10.1061/AJGEB6.0000117.

Liu, B., and D. S. Jeng. 2016. Laboratory Study for Influence of Clay Content (CC) on Wave-Induced Liquefaction in Marine Sediments. *Marine Georesources & Geotechnology* 34 (3): 280–292. doi:10.1080/1064119X.2015.1005322.

Liu, H., and T. Qiao. 1984. Liquefaction Potential of Saturated Sand Deposits Underlying Foundation of the Structure. *In Proceedings of the 8th World Conference on Earthquake Engineering,* vol. 3, 199–206. Englewood Cliffs, N.J.: Prentice-Hall.

Maharjan, M., and A. Takahashi. 2013. Centrifuge Model Tests on Liquefaction-Induced Settlement and Pore Water Migration in Non-Homogeneous Soil Deposits. *Soil Dynamics and Earthquake Engineering* 55: 161–169. doi:10.1016/j.soildyn.2013.09.002.

Miura, S., and S. Toki. 1982. A Sample Preparation Method and Its Effect on Static and Cyclic Deformation-Strength Properties of Sand. *Soils and Foundations* 22 (1): 61–77. doi:10.3208/sandf1972.22.61.

National Research Council NRC. 1985. Liquefaction of Soils During Earthquakes. *National Research Council Rep. No. CETS-EE-001*, National Academy Press, Washington, D.C.

Ozener, P. T., K. Ozaydin, and M. M. Berilgen. 2009. Investigation of Liquefaction and Pore Water Pressure Development in Layered Sands. *Bulletin of Earthquake Engineering* 7 (1): 199–219. doi:10.1007/s10518-008-9076-3.

Porcino, D., V. Marcianò, and V. Nicola Ghionna. 2009. Influence of Cyclic Pre-Shearing on Undrained Behavior of Carbonate Sand in Simple Shear Tests. *Geomechanics and Geoengineering* 4 (2): 151–161. doi:10.1080/17486020902855662.

Seed, H. B., and K. L. Lee. 1966. Liquefaction of Saturated Sands during Cyclic Loading. *Journal of the Soil Mechanics and Foundations Division* 92 (6): 105–134. doi:10.1061/JSFEAQ.0000913.

Singh, S. 1996. Liquefaction Characteristics of Silts. *Geotechnical and Geological Engineering* 14 (1): 1–19. doi:10.1007/BF00431231.

Sitharam, T. G., J. S. Vinod, and B. V. Ravishankar. 2009. Post-Liquefaction Undrained Monotonic Behavior of Sands: Experiments and DEM Simulations. *Géotechnique* 59 (9): 739–749. doi:10.1680/geot.7.00040.

Sriskandakumar, S. D. Wijewickreme, and P. M. Byrne. 2012. Multiple Cyclic Loading Response of Loose Air-Pluviated Fraser River Sand. In Proceedings of the 15th World Conference on Earthquake Engineering, Lisbon, September 24–28.

Stallings Young, E. G., N. Mahabadi, C. E. Zapata, and L. A. Van Paassen. 2021. Microbial-Induced Desaturation in Stratified Soil Conditions. *International Journal of Geosynthetics and Ground Engineering* 7 (2): 37. doi:10.1007/s40891-021-00276-9.

Tasiopoulou, P., A. Giannakou, J. Chacko, and S. D. Wit. 2019. Liquefaction Triggering and Post-Liquefaction Deformation of Laminated Deposits. *Soil Dynamics and Earthquake Engineering* 124: 330–344. doi:10.1016/j.soildyn.2018.04.044.

Thevanayagam, S., and S. Mohan. 2000. Inter-Granular State Variables and Stress-Strain Behaviour of Silty Sands. *Géotechnique* 50 (1): 1–23. doi:10.1680/geot.2000.50.1.1.

Toyota, N. Post-Cyclic Triaxial Behavior of Toyoura Sand. In Proceedings of IS-Tokyo 1995, Tokyo, Japan, vol. 961, 189–195.

Vaid, Y. P., and J. C. Chern. 1983. Effect of Static Shear on-Resistance of Liquefaction. *Soils and Foundations* 23 (1): 47–60. doi:10.3208/sandf1972.23.47.

Vaid, Y. P., and D. Negussey. 1984. Relative Density of Pluviated Sand Samples. *Soils and Foundations* 24 (2): 101–105. doi:10.3208/sandf1972.24.2_101.

Vaid, Y. P., J. D. Stedman, and S. Sivathayalan. 2001. Confining Stress and Static Shear Stress in Cyclic Liquefaction." *Canadian Geotechnical Journal* 38 (3): 580–591. doi:10.1139/t00-120.

Wijewickreme, D., S. Sriskandakumar, and P. Byrne. 2005. Cyclic Loading Response of Loose Air-Pluviated Fraser River Sand for Validation of Numerical Models Simulating Centrifuge Tests. *Canadian Geotechnical Journal* 42 (2): 550–561. doi:10.1139/t04-119.

Xiu, Z., S. Wang, Y. Ji, F. Wang, and F. Ren. 2020. Experimental Investigation on Liquefaction and Post-Liquefaction Deformation of Stratified Saturated Sand under Cyclic Loading. *Bulletin of Engineering Geology and the Environment* 79 (5): 2313–2324. doi:10.1007/s10064-019-01696-8.

Yang, J., and H. Y. Sze. 2011. Cyclic Behavior and Resistance of Saturated Sand under Non-Symmetrical Loading Conditions. *Géotechnique* 61 (1): 59–73. doi:10.1680/geot.9.P.019.

Yang, J., and H. Y. Sze. 2014. Failure Modes of Sand in Undrained Cyclic Loading: Impact of Sample Preparation. *Journal of Geotechnical and Geoenvironmental Engineering* 140 (1): 152–169. doi:10.1061/(ASCE)GT.1943-5606.0000971.

Yoshimine, M., and R. Koike. 2005. Liquefaction of Clean Sand with Stratified Structure Due to Segregation of Particle Size. *Soils and Foundations* 45 (4): 89–98. doi:10.3208/sandf.45.4_89.

Zhao, Y., K. Gafar, M. Z. E. B. Elshafie, A. Deeks, J. Knappett, and S. Madabhushi. 2006. Calibration and Use of a New Automatic Sand Pourer. In Proceedings of the 6th International Conference on Physical Modelling in Geotechnics, Hong Kong, August.

Zhang, H., Y. Yang, and H. S. Yu. 2021. Liquefaction and Postliquefaction of Granular Material under Multi-Directional Cyclic Loading. *Marine Georesources & Geotechnology* 39 (11): 1261–1272. doi:10.1080/1064119X.2020.1823536.

*2022 3rd International Conference for Emerging Technology (INCET)*
*Belgaum, India. May 27-29, 2022*

# Deep Learning Framework for Macro Marine Litter Classification and Quantification

Dr. Shailender Kumar
*Department of Computer Science and Engineering*
*Delhi Technological University*
Delhi, India
shailenderkumar@dce.ac.in

Abhinav Gautam
*Department of Computer Science and Engineering*
*Delhi Technological University*
Delhi, India
abhinavgautam_2k18co012@dtu.ac.in

Khushbu Mehlawat
*Department of Computer Science and Engineering*
*Delhi Technological University*
Delhi, India
khushbumehlawat_2k18co174@dtu.ac.in

Kunal Rawat
*Department of Computer Science and Engineering*
*Delhi Technological University*
Delhi, India
kunalrawat_2k18co181@dtu.ac.in

2022 3rd International Conference for Emerging Technology (INCET) | 978-1-6654-9499-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/INCET54531.2022.9825334

***Abstract*—Marine Litter has become a serious concern in today's world. It is not only dangerous for marine life, but also has a negative socio-economic impact. Tourism being one of the major contributors to the economy, cleanliness around the water bodies becomes one of the major pointers to attract tourists. With present Governments investing in the marine litter management, it becomes essential to properly utilize the resources for litter management. Inadequate knowledge of quantity and type of litter leads to mismanagement and wastage of resources.**

**Therefore, we propose deep learning-based solution of Marine Litter detection. With the type and quantity of several types of marine waste, better management can be done. Along with this, we also propose a line sweep algorithm to find the approximate area cover of the marine litter using the scale of the images. Knowledge of area cover can help in litter management and monitoring.**

***Keywords—Macro Marine Litter (MML), data augmentation, Object Detection, YOLOv3, line sweep algorithms.***

## I. Introduction

Marine life has been getting demolished due to the accumulation of litter , which is causing a long-term threat on marine ecosystems. Much of this debris originates from land and thereafter breaks down into macro and micro elements due to thermal and oxidative degradation. This litter arises from various sectors majorly from aquaculture and fisheries. This persistent accumulation of litter is the result of poor waste management infrastructure. Marine litter can cause long term impact to sectors such as fishing , shipping and tourism. It can lead to a reduction in capture due to entanglement, ingestion and exposure to toxic material. It can damage the ship's vessels and equipment which can hamper the supply chain and also adds to the repair cost. Litter damage to shipping costs around USD 279 million per year.[1] Also polluted beaches can discourage tourists and can hamper jobs in the tourism sector. Social impact of MML includes fewer recreational activities and also it can impact the health of visitors. Consumption of contaminated marine products can lead to gastrointestinal diseases and deep wound infections in humans.

This issue has started gaining attention as this has become an environmental concern as it's killing the marine ecosystem and also putting human health at high risk. Methods have been introduced to keep a check on the increasing amount of garbage in the ocean. Some of the techniques are Trawling and counting, Monitoring via scuba drivers. These techniques require labor and are expensive, also the degree of litter detection depends upon the skills and experience of divers.

In this paper, we have proposed a deep learning based solution for the identification, classification, and quantification of the Macro Marine Litter.

## II. Terminologies

### A. MML (Marine Macro Litter)

Macro Marine litter is man-made waste that has been deliberately discharged into the sea, river and lakes by humans causing a broad spectrum of environmental , economic, health and cultural impacts.[2]

### B. Data Augmentation

Data Augmentation is a process of obtaining new images from the already available images by using several affine transformations such as rotation, flipping, random cropping etc. This helps in obtaining an enriched and more generalized dataset from a small dataset.[3]

### C. Bounding Box

Bounding Boxes are rectangles whose sides are parallel to the x and the y axes. These boxes are used for image localization, i.e. mark the objects of interest in an image to show their presence and location in the image.[4]

### D. Sweep Line Algorithms

Sweep line algorithms are used in solving computational geometry problems. In a sweep line algorithm an imaginary line is swept parallel to one of the axes in the XY- plane and a sub problem is solved or a task is performed at some positions of the line.[5]



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:26:20 UTC from IEEE Xplore. Restrictions apply.

## III. Previous works

With rising concern towards MML management and disposal along with advancement in AI and computer vision has allowed to come up with AI based solution for MML detection and quantification.

A deep learning framework using CNN[6] architecture for automatic detection of FMML and NON-FMML aerial images was proposed by. A photographic survey was done using drones and aircraft for the collection of aerial images. A web application was designed which uses two functions to classify the image. First the CNN model was trained and then the image is split into multiple cells, then the algorithm classifies each cell into FMML and Non-FMML.[7]

To tackle the problem of in-situ visual census which requires intensive human effort, researchers in developed an Aerial system known as Unmanned aerial system (UAS) framework. Their framework consists of 4 operational steps: (1) Collecting ultra-high images using UAS. (2) Processing the image using Multi Video Stereo and Structure from Motion to produce DSM(digital surface model) and mosaic. (3) Detecting MML using OOML and supervised classifiers which are KNN[8], SVM[9] and RF[10] .For identification of homogeneous objects with better accuracy, they did image segmentation for separating out homogeneous objects.. (4)Finally, computing the centroids of MML objects by the abundance maps which were created in previous steps.

After optimally tuning the 3 object-oriented machine learning (OOML) the results were 68% for SVM, 72% for RF and 65% for KNN. RF has the highest accuracy among all but it has more parameters to tune, KNN on the other hand has the worst accuracy among the other two with only 1 parameter to tune. With this they concluded that combination of Aerial Images with OOML classifier is a great step towards achieving cost-efficient monitoring for estimation of MML at sandy beaches. The results are described in Table I.[2].

Authors in discussed the classification of underwater marine litter using deep learning techniques. For model training various architectures are used such as 4-layer custom CNN architecture, VGG16[11], XCEPTION[12], ResNet [13]and YOLOv3. Dataset for the training was collected through the Internet and synthetic images were built by applying synthetic marine litter images on the underwater scenery. Results showed that peak validation accuracy ≈ 88%, testing accuracy ≈ 85% is achieved by VGG16 . Future improvements as suggested by the author can be done by using different architecture such as YOLOv4 and using more diverse data for training.[14] Fig.1 shows comparison between various classifiers.

| O-O Classifier | Parameters | Area | TP | FN | FP | P (%) | R (%) | F (%) |
|---|---|---|---|---|---|---|---|---|
| RF | ntree = 500; Dep = 5; Ns = 5 | A2 | 77 | 41 | 25 | 75 | 65 | 70 |
| | | A3 | 87 | 37 | 29 | 75 | 70 | 73 |
| SVM | Ϋ = 0.1; C = 5 | A2 | 74 | 44 | 21 | 78 | 63 | 69 |
| | | A3 | 76 | 48 | 26 | 75 | 61 | 67 |
| KNN | K = 10 | A2 | 72 | 46 | 34 | 68 | 61 | 64 |
| | | A3 | 78 | 46 | 38 | 67 | 63 | 65 |

For SVM and KNN the 12 bands are normalized.

Fig. 1. Performance of various classifiers

In a survey, various algorithms in the domain of underwater trash detection were tested and results were presented. Pros and cons of various traditional manual practices and AI based solutions were discussed.[15]

## IV. Our Approach

Our methodology consists of several steps, firstly using an object detection model, to detect and classify MML objects in the image. Then using a sweep line algorithm to find the total area cover of MML in the image and finally, obtain area on the ground from the area on the image.

### *A. MML detection*

In MML detection, we train and deep learning object detection model for detection of the following classes of MML- cardboard, glass, paper, plastic, metal and edible waste.

For object detection task we trained a YOLOv3 object detection model. Several algorithms are now available for object detection. YOLOv3 is the fastest of all state-of-art object detection algorithms.

YOLOv3 is the third version of YOLO [16] algorithms. Generally, in object detection there are 2 separate tasks, one for proposing or obtaining bounding boxes and then running a classifier on the bounding boxes. But in YOLO these two tasks are done in one forward pass. In YOLO, the image is divided into grid cells, and in one pass, a prediction is obtained for each grid cell. Thus, YOLO is faster than other algorithms.

Originally, YOLOv3 contains fifty-three CNN layers (Darknet-53), for detection tasks additional 53 layers are stacked giving one-hundred-twenty-six layers of YOLO. Fig. 2 depicts the architecture of YOLOv3.[17]

We re-configured the original model for our task of MML detection, since we have 6 classes, we adjusted the number of filters in the 59th, 67th and 75th layers to be 33. The reconfiguration of the model is depicted in Table I and Table II.

There are several advantages of YOLOv3 apart from speed. It has better precision on small and large objects. It does the detection at three layers, 82, 94 and 106 by down sampling the images, to detect small, medium and large objects. For down sampling, it uses 1x1 conv. rather than max pooling thus preserving the low-level features.[18]

The MML objects can be very closely packed and overlapping each other, thus detection becomes tough, since YOLOv3 provides 9 anchor boxes per grid cell, thus resolving the problem to some extent. With use of anchor boxes overlapping objects can be detected with better accuracy.

The customization in the YOLOv3 model is mentioned in Table I and Table II.

Number of classes in convolutional layers was decided on the equation in Fig. 3.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:26:20 UTC from IEEE Xplore. Restrictions apply.

*B. MML Quantification*

MML quantification is necessary for management of waste. Prior knowledge of the quantity of litter of different classes can enhance management and utilization of resources. Quantification can be done in two ways.

| | Type | Filters | Size | Output |
|---|---|---|---|---|
| | Convolutional | 32 | 3 × 3 | 256 × 256 |
| | Convolutional | 64 | 3 × 3 / 2 | 128 × 128 |
| 1× | Convolutional | 32 | 1 × 1 | |
| | Convolutional | 64 | 3 × 3 | |
| | Residual | | | 128 × 128 |
| | Convolutional | 128 | 3 × 3 / 2 | 64 × 64 |
| 2× | Convolutional | 64 | 1 × 1 | |
| | Convolutional | 128 | 3 × 3 | |
| | Residual | | | 64 × 64 |
| | Convolutional | 256 | 3 × 3 / 2 | 32 × 32 |
| 8× | Convolutional | 128 | 1 × 1 | |
| | Convolutional | 256 | 3 × 3 | |
| | Residual | | | 32 × 32 |
| | Convolutional | 512 | 3 × 3 / 2 | 16 × 16 |
| 8× | Convolutional | 256 | 1 × 1 | |
| | Convolutional | 512 | 3 × 3 | |
| | Residual | | | 16 × 16 |
| | Convolutional | 1024 | 3 × 3 / 2 | 8 × 8 |
| 4× | Convolutional | 512 | 1 × 1 | |
| | Convolutional | 1024 | 3 × 3 | |
| | Residual | | | 8 × 8 |
| | Avgpool | | Global | |
| | Connected | | 1000 | |
| | Softmax | | | |

Fig. 2. YOLOv3 original architecture.

TABLE I. CUSTOM CONFIGURATION OF YOLO LAYERS

| Layer | Classes |
|---|---|
| 1st YOLO layer | 6 |
| 2nd YOLO layer | 6 |
| 3rd YOLO layer | 6 |

TABLE II. CUSTOM CONFIGURATION OF CONV LAYERS

| Layer | filters |
|---|---|
| 59th Convolutional Layer | 33 |
| 67th Convolutional Layer | 33 |
| 75th Convolutional Layer | 33 |

$$\#\,filters = (\#\,classes + 5) * 11$$

Fig. 3. Number of filters in convolutional layers preceding the yolo layers

Firstly, the quantification of MML in terms of area cover.

For quantification in terms of area the litter covers, total area of bounding boxes of the litter is to be calculated.

Bounding boxes are rectangles, whose sides are parallel to both the axes as shown in Fig. 4. For defining rectangles whose sides are parallel to axes, we need coordinates of any of the two vertices. In our case, top-left and bottom-right vertices is selected. So, the rectangle shown in Fig. 4 is represented as (x1, y1, x2, y2). Axes in the system are defined as depicted in Fig.5.

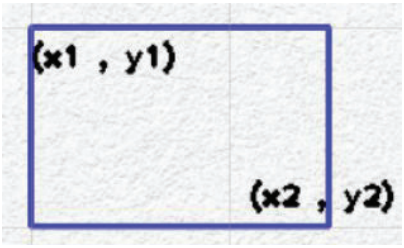



Fig. 4. Bouding Box represntation

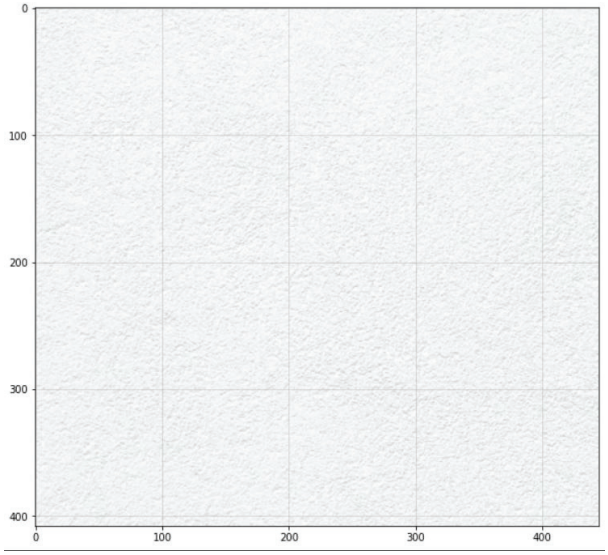



Fig. 5. Axes used in the system

But the bounding boxes may overlap (Fig. 6) each other, so simply adding the area of each box will not give the correct result. For obtaining the total area of the rectangles which may overlap, a line sweep algorithm was used to find the total area of the bounding boxes. The steps of the algorithm are as follows-obtain U, set of unique x values from the set of rectangles R in sorted order.

split all the rectangles whose area falls within the range of the x $\epsilon$ U and obtain the new set of rectangles R'. With this we will be able to separate rectangles along the x-axis. Now only those rectangles overlap which have the same x1 and x2.

Then for all rectangles in R', merge the rectangles that have the same minx (minimum values of x) and maxX (maximum value of x) and are overlapping, to obtain the set of non-overlapping rectangles Rc. We merge the overlapping rectangles, so that the intersecting area does not get counted twice in the area calculation.

Now, since rectangles in Rc are non-overlapping, total area simply becomes the total sum of areas of all the rectangles.

For example, in the given set of rectangles R = {(100,100,250,250),(150,150,280,280),(210,210,320,320),(180,80,300,340)} shown in Fig. 6.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:26:20 UTC from IEEE Xplore. Restrictions apply.

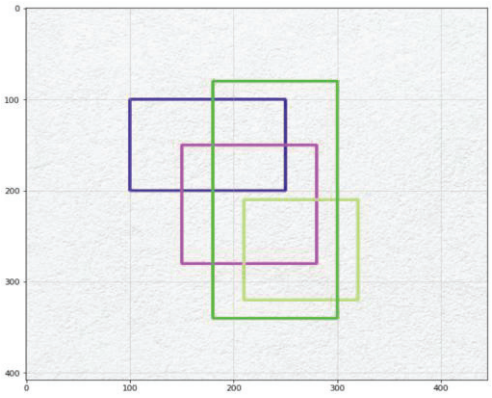


Fig. 6. Rectangles in R.

We obtain the set of unique sorted x values U = {100,150,180,210,250,280,300,320}.

Then we split the rectangles in R on x ϵ U to new set of rectangles R' = {( 100 , 150 , 100 , 200 ),( 150 , 180 , 100 , 200 ),( 180 , 210 , 100 , 200 ),( 210 , 250 , 100 , 200 ),( 150 , 180 , 150 , 280 ),( 180 , 210 , 150 , 280 ),( 210 , 250 , 150 , 280 ),( 250 , 280 , 150 , 280 ),( 210 , 250 , 210 , 320 ),( 250 , 280 , 210 , 320 ),( 280 , 300 , 210 , 320 ),( 300 , 320 , 210 , 320 ),( 180 , 210 , 80 , 340 ),( 210 , 250 , 80 , 340 ),( 250 , 280 , 80 , 340 ),( 280 , 300 , 80 , 340 )} (Fig. 7).

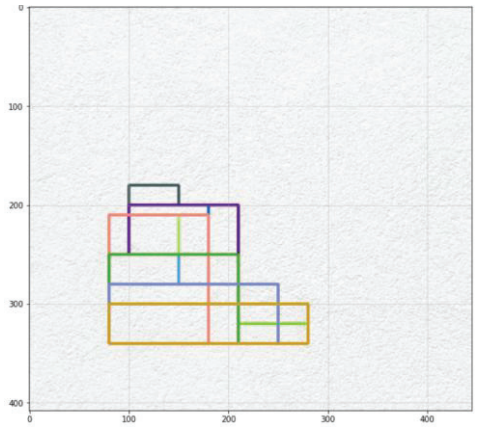


Fig. 7. Rectangles in R'

Then we merge the rectangles which are overlapping and have the same minX and maxX to obtain set of non-overlapping rectangles Rc = {( 100 , 100 , 150 , 200 ),( 150 , 100 , 180 , 280 ),( 180 , 80 , 210 , 340 ),( 210 , 80 , 250 , 340 ),( 250 , 80 , 280 , 340 ),( 280 , 80 , 300 , 340 ),( 300 , 210 , 320 , 320 )}.(Fig. 8).

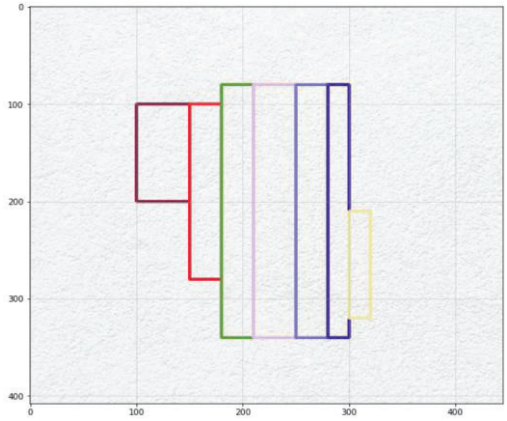


Fig. 8. Rectangles in $R_c$

Final area becomes 43800 pixel$^2$.

Secondly, we can also classify MML in terms of their amount or number. The YOLOv3 algorithm detects multiple objects in an image, rather than one. So along with obtaining the bounding boxes of the detected objects, we can get the count of each type. The accuracy of counting depends on the accuracy of the model to distinguish between different objects, many of which can be closely packed with each other.

### C. Conversion of area on image to area on ground

The area obtained using the line sweep algorithm is in pixel$^2$.

To convert the area on the image to the area on the ground, we need the scale of the image of the image. Scale of the image can be in m/pixel, which means actual length on ground (in metre) represented per pixel.

If the scale of the image is not known, it can be obtained using the method described in [19], by fixing two points on the ground and same two points on the image, the ratio can be used as the scale of the image.

After obtaining the scale of the image, let it be d m/pixel, the area on the ground can be obtained using the formula in Fig. 9.

$$\text{Area on ground} = (\text{Area on the image}) * d^2 \text{ m}$$

Fig. 9. Formula to obtain area on ground from area on image

If the scale of the image is not known, we can still calculate the approximate area of the MML, using the percentage area cover. For this method, we need the total area of the ground in the image, then using the ratio of MML in the image to obtain the area on the ground. The formula in Fig. 10 can be used to obtain the area on the ground.

$$\textit{MML area on the ground} = \frac{\textit{MML area in the image}}{\textit{Total area of image}} * \textit{total area on the ground}$$

Fig. 10. Formula to obtain area on ground using the area ratios.

## V. Experimentation

The experimentation pipeline is described in Fig. 11 starting from raw image data to model training and testing.

### A. Data Collection

Data was collected mainly from two sources, Anti-plasti [20] and Google Open-Source images. Dataset originally consisted of images of MML objects, single and multiple objects in an image. This dataset was highly imbalanced and very less in the number for training such a deep network.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:26:20 UTC from IEEE Xplore. Restrictions apply.

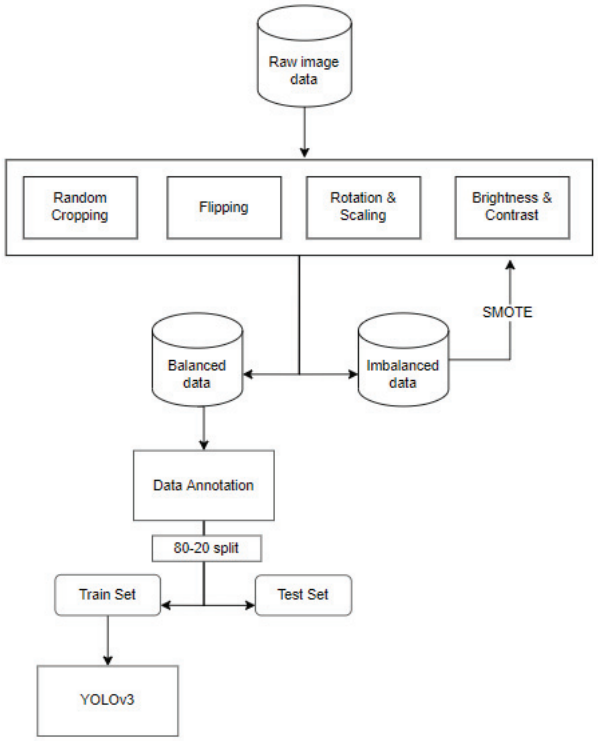



Fig. 11. Experimentation Framework

### B. Data Augmentation

The original dataset contained around 50 images. Data Augmentation techniques such as Random cropping, flipping, rotation, scaling, and brightness & contrast variations. Since the MML objects are of shape, which is invariant of the orientations, thus we were able to use a lot of techniques. We finally obtained a dataset containing 1000 images. Since the YOLOv3 model uses CNN layers and CNN(s) are invariant to transformed images, thus Augmentation can be freely used to obtain a bigger dataset.[3]

### C. Data Annotaion

The images obtained are then annotated for MML objects, we used the LabelImg tool for annotation.[21]

### D. Data balancing

Since the original dataset was imbalanced so the new dataset was also imbalanced, we used SMOTE which is an augmentation-based minority oversampling technique to obtain a balanced dataset. The balanced dataset is important, as it does not let the model get skewed towards data of one kind.[22]

### E. Model Training

Dataset was split 80-20 for training and testing, The model was trained for 1000 epochs and loss plots are as shown in Fig.12. and Fig. 13.

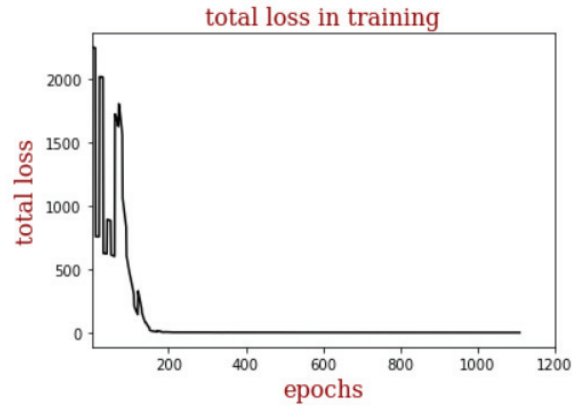



Fig. 12. Total Loss plot

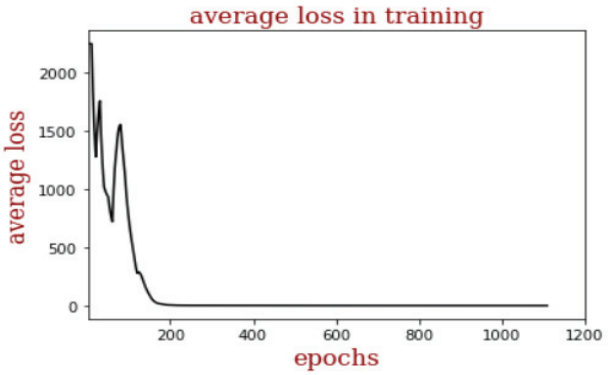



Fig. 13. Average Loss plot

### F. Testing

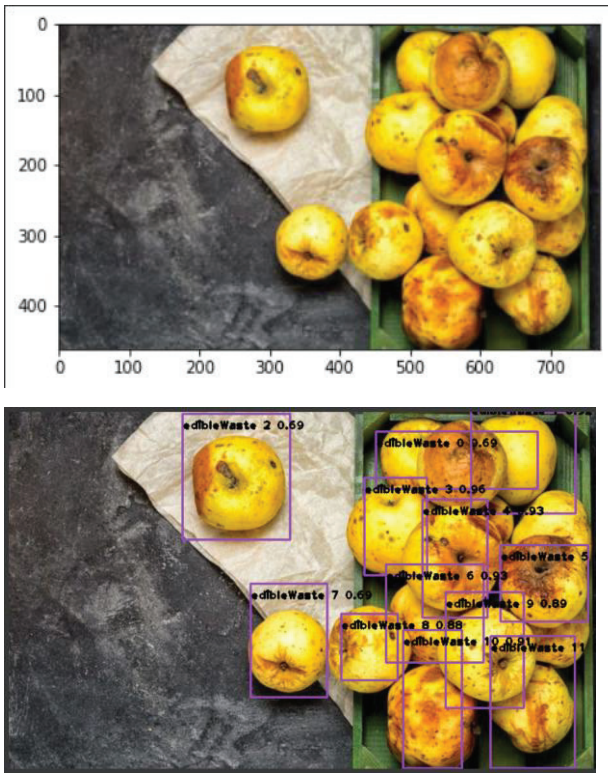



Fig. 14. Testing Sample



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:26:20 UTC from IEEE Xplore. Restrictions apply.

In this testing sample, we obtained the bounding boxes for the MML, as described in the Fig. 14.

We obtained the total area of the MML using the described sweep line algorithm as 144,777 pixel$^2$. The total area of the image is 400,000 pixel$^2$.

The obtained ratio is 0.362. Now using this ratio and the total ground area covered in the image, we can obtain the area covered by MML on the ground.

## VI. Conclusion & Future Work

YOLOv3 algorithm along with sweep line algorithm helped detection and quantification of MML.

Our methodology can be used for proper utilization of resources for waste cleaning, prior knowledge of amount, area cover and demography of the MML can help in proper management and utilization of resources.

This methodology can also be used for real-time monitoring of an area for MML or waste. With proper setups for live feed daily waste dumping can be estimated and cleaning process and resources can be planned based on it.

## Acknowledgment

The authors would like to thank Ministry of Foreign Affairs, India for providing the problem statement on detection of MML. The authors would also like to express their gratitude towards Delhi Technological University, Delhi, India for providing the resources and support during the research work.

## References

[1] "MARINE LITTER SOCIO-ECONOMIC STUDY".

[2] G. Gonçalves, U. Andriolo, L. Gonçalves, P. Sobral, and F. Bessa, "Quantifying Marine Macro Litter Abundance on a Sandy Beach Using Unmanned Aerial Systems and Object-Oriented Machine Learning Methods," *Remote Sensing 2020, Vol. 12, Page 2599*, vol. 12, no. 16, p. 2599, Aug. 2020, doi: 10.3390/RS12162599.

[3] "Image Augmentation for Deep Learning | by Suki Lau | Towards Data Science." https://towardsdatascience.com/image-augmentation-for-deep-learning-histogram-equalization-a71387f609b2 (accessed Jan. 08, 2022).

[4] "Basics of Bounding Boxes. What is a bounding box? | by Vineeth S Subramanyam | Analytics Vidhya | Medium." https://medium.com/analytics-vidhya/basics-of-bounding-boxes-94e583b5e16c (accessed Jan. 08, 2022).

[5] "Sweep line algorithm - Wikipedia." https://en.wikipedia.org/wiki/Sweep_line_algorithm (accessed Jan. 08, 2022).

[6] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," Mar. 2016, Accessed: Sep. 10, 2021. [Online]. Available: http://arxiv.org/abs/1603.07285

[7] O. Garcia-Garin *et al.*, "Automatic detection and quantification of floating marine macro-litter in aerial images: Introducing a novel deep learning approach connected to a web application in R," *Environmental Pollution*, vol. 273, p. 116490, Mar. 2021, doi: 10.1016/J.ENVPOL.2021.116490.

[8] "All about K-Nearest Neighbors. In 1951, Evelyn Fix and Joseph Hodges… | by Anoushka Nandan | GDSC KIIT | Medium." https://medium.com/dsckiit/all-about-k-nearest-neighbors-3027e60aad25 (accessed Jan. 09, 2022).

[9] "SVM | Support Vector Machine Algorithm in Machine Learning." https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/ (accessed Jan. 09, 2022).

[10] "Understanding Random Forest. How the Algorithm Works and Why it Is… | by Tony Yiu | Towards Data Science." https://towardsdatascience.com/understanding-random-forest-58381e0602d2 (accessed Jan. 09, 2022).

[11] K. Simonyan and A. Zisserman, "VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION," 2015. [Online]. Available: http://www.robots.ox.ac.uk/

[12] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions."

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition." [Online]. Available: http://image-net.org/challenges/LSVRC/2015/

[14] J. Musić, S. Kružic, I. Stančić, and F. Alexandrou, "Detecting Underwater Sea Litter Using Deep Neural Networks: An Initial Study," *2020 5th International Conference on Smart and Sustainable Technologies, SpliTech 2020*, Sep. 2020, doi: 10.23919/SPLITECH49282.2020.9243709.

[15] R. Bajaj, N. Kulkarni, R. Raut, and S. Garg, "Sea Debris: A Review of Marine Litter Detection Techniques," 2021.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 779–788, Dec. 2016, doi: 10.1109/CVPR.2016.91.

[17] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement", Accessed: Jan. 08, 2022. [Online]. Available: https://pjreddie.com/yolo/.

[18] "YOLOv3: Real-Time Object Detection Algorithm (What's New?) - viso.ai." https://viso.ai/deep-learning/yolov3-overview/ (accessed Jan. 08, 2022).

[19] M. Roy, A. Gautam, and A. Sugandhi, "A deep learning framework for enhancing maritime coastal security," *2021 2nd International Conference for Emerging Technology, INCET 2021*, May 2021, doi: 10.1109/INCET51464.2021.9456235.

[20] "GitHub - antiplasti/Plastic-Detection-Model: Image Recognition Model to detect plastics, glass, paper, rubbish, metal and cardboard. This is used to detect these pollution in the ocean to allow the eradication of these materials, helping marine life, fishermen, tourism and making the world resilient to climate change." https://github.com/antiplasti/Plastic-Detection-Model (accessed Jan. 08, 2022).

[21] "GitHub - tzutalin/labelImg:  LabelImg is a graphical image annotation tool and label object bounding boxes in images." https://github.com/tzutalin/labelImg (accessed Jan. 08, 2022).

[22] N. v. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/JAIR.953.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:26:20 UTC from IEEE Xplore. Restrictions apply.

Analog Integrated Circuits and Signal Processing
https://doi.org/10.1007/s10470-022-02069-0

# Design and implementation of OTA based fractional-order oscillator

**Garima Varshney**[1] · **Neeta Pandey**[1] · **Rajeshwari Pandey**[1]

Received: 10 February 2022 / Revised: 19 May 2022 / Accepted: 20 June 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

**Abstract**
This paper deals with design and implementation of fractional-order (FO) oscillator based upon two topologies which employ operational transconductance amplifier (OTA). The first topology is developed on a trans-admittance mode fractional-order all-pass filter (FAPF) cascaded with an trans-impedance mode integrator whereas the other one consist of a trans-admittance mode FAPF cascaded with a trans-impedance mode differentiator. The presence of OTA makes the proposed FO oscillator electronic tunable. Both topologies are verified through SPICE simulations using 180nm CMOS technology node. The ability of proposed FO oscillator, to generate very low frequency is also explored in this work. The stability analysis is done using *forlocus* function of MATLAB and the mathematical equations for sensitivity of the oscillation frequency with respect to various circuit parameters are also derived. Further, to ensure the robustness of the proposed circuits process-voltage-temperature (PVT) and Monte Carlo analyses have been performed.

**Keywords** Circuit design · Fractional-order oscillators · Monte Carlo analysis · PVT analysis · Stability analysis · VLF circuit · OTA

## 1 Introduction

These-days, fractional-order calculus is an expanding field for research, by virtue of its inherit capability to describe a real phenomenon more precisely when compared to conventional integer-order methods [1, 2].

Fractional-order systems are used in various fields of science in many applications, such as: control systems [3, 4], electrical engineering [5], analog signal processing [6, 7], biomedical [8, 9], communication [10, 11] and many more. Consequently, at interdisciplinary level fractional-order systems have huge opportunity for research.

In analog signal processing, oscillators are used for waveform generation and very crucial circuits. In general, oscillators are classified in two categories, namely, sinusoidal oscillators and relaxation oscillators (multivibrator). In [12], FO oscillator was first substantiated using well-known oscillator Wein oscillator using operational amplifier (OA). Further, in [12–30] various topologies of oscillators are generalized in fractional domain, such as: Wein's bridge, Colpitts, multiple integrator loop, RLC resonator etc. OTA is a very versatile active block, which provides electronic tuning of frequency of oscillation through the transconductance parameter. In [31, 32] and [33] a wide catalogue of integer-order OTA-C oscillators have been presented. FO oscillators are generalized version of integer-order oscillators and provide wider range of frequency of oscillation and controlled phase shift through $\alpha$. These features of FO oscillator can be used in variety of applications: such as, very low frequencies can be used in bio-medical applications, while very high frequencies are used in communication field. The oscillators with controlled phase shift find their applications in several areas such as: PSK modulation-demodulation scheme, diagnosing of Alzheimer disease and music synthesizing [34, 35] etc. OTA-C FO oscillators have not been much explored in literature; thus, this work proposes two topologies for OTA-C FO oscillator. A summary of FO oscillators available in open literature [12–30] based upon components count, topology used, electronic tunability and very low frequency (VLF) mode is given in Tab. 1 and following are the observations:

– Operational amplifier (Op-Amp) [12, 14, 16, 19, 24and 25], operational transresistance amplifier (OTRA) [13], second generation current conveyor (CCII) [14, 30], current feedback operational amplifier (CFOA) [15], opera-

✉ Garima Varshney
gvarshn1@gmail.com

[1] Department of Electronics and Communication Engineering, Delhi Technological University, 110042 Delhi, India

Published online: 07 July 2022

**Table 1** Comparison of proposed fractional-order oscillator based upon various features

| Ref. | ABBs used (count) | Passive components count | Topology used | FOC count (All grounded-Y/N) | Different orders of FOC used | Electronic tunability of $f_0$ | VLF mode |
|---|---|---|---|---|---|---|---|
| [12] | Opamp (1) | 4/7 | Wein's bridge | 2/3(N) | Yes | No | Not explored |
| [13] | OTRA (1/2) | 4/7 | Wein's bridge | 2(N) | Yes | No | Not explored |
| [14] | CC/Opamp (3/4) | 8 | Multiloop feedback | 4(Y) | Yes | No | Not explored |
| [15] | CFOA (7/8) | 7/8 | Multiloop feedback | 3(Y) | Yes | No | Not explored |
| [16] | Opamp (1/2) | 7 | Two port network | 2(N) | Yes | No | Not explored |
| [17] | OTA/IVB (4) | 3 | Colpitts | 2(N) | Yes | No | Not explored |
| [18] | OTA/IVB/CF (5) | 4 | Colpitts | 3(N) | Yes | No | Not explored |
| [19] | Opamp (1) | 6 | Wein's bridge | 2(N) | No | No | Not explored |
| [20] | DDCC (2) | 4 | HPF with integrator | 2(Y) | Yes | No | Not explored |
| [21] | VDIBA (1) | 4 | Wein's bridge | 2(N) | Yes | No | Not explored |
| [22] | DDCC (3) | 6 | Multiple integrator loop | 3(Y) | Yes | No | Not explored |
| [23] | MOS (1) | 4 | Colpitts | 3(N) | Yes | No | |
| [24] | Opamp (1) | 5/6 | Wein's bridge | 2(N) | Yes | No | Not explored |
| [25] | Opamp (2) | 6 | FAPF with integrator/differentiator | 2(N) | Yes | No | Not explored |
| [26] | MOS (12) | 3 | Multiple integrator loop | 3(N) | No | Yes | Not explored |
| [27] | MMCC (3) | 5 | Multiple integrator loop | 2(Y) | Yes | No | Not explored |
| [28] | MMCC (3) | 5 | Multiple integrator loop | 2(Y) | Yes | Yes | Not explored |
| [29] | DVCCTA/VGA (2) | 3 | Parallel RLC resonator | 2(Y) | No | Yes | Not explored |
| [30] | CCII+/DDA/VGA (4) | 5 | Parallel RLC resonator | 2(Y) | Yes | Yes | Not explored |
| Prop. | OTA (2/4) | 2 | FAPF with Integrator/Differentiator | 2(Y) | Yes | Yes | Explored |

tional transconductance amplifier (OTA) [17 and 18], differential difference current conveyor (DDCC) [20, 22], MOS transistors [23,26], voltage differencing inverting buffered amplifier (VDIBA) [21], multiplication-mode current conveyor (MMCC) [27–28] and differential voltage current conveyor transconductance amplifier (DVCCTA) [29] based FO oscillators are reported in respective references.

- [12–13, 19, 21,24] use Wein's bridge topology, while [14–15] use multiloop feedback topology. Colpitts topology is used in [17–18, 23], whereas [29–30] use parallel RLC resonator topology to build the FO oscillator. Multiple integrator loop topology is used in [22, 26–28], while integrator with fractional-order high-pass filter (FHPF) has been used in [20]. FAPF with integrator/differentiator topology is used in [25].
- The FO oscillators reported in [26, 28–30] possess electronic tunability of $f_0$.
- All tunable FO oscillators, available in open literature require resistor or floating FOC.
- The possibility of VLF mode has not been explored.

It is apparent from Table 1 that limited research is available for electronically tunable oscillators in fractional-domain which are resistor-less and use all grounded FOCs. In view of these issues the objectives of the proposed work are below:

- Resistor-less FO oscillator's structure with all grounded FOCs and electronic tunability of $f_0$.
- Lesser number of components count.
- Examine the suitability of the proposed FO oscillator for VLF mode.

This work proposes two topologies for FO oscillator, comprises of a trans-admittance mode FAPF and a trans-impedance mode integrator/differentiator using OTA. The proposed topologies are resistor-less and using grounded FOCs. Further, the component's count (ABB+passive) for FAPF with integrator topology is 4 (2+2) while the same for FAPF with differentiator topology is 6 (4+2). Moreover, the possibility to generate very low frequencies without changing the value of FOC is explored.

The rest of this manuscript is arranged as follows: Sect. 2 introduces the proposed FO oscillators' topologies, while sect. 3 talks about the simulation results. Stability and sensitivity analysis of the proposed structures are discussed in sect.4 followed by conclusion in section 5.

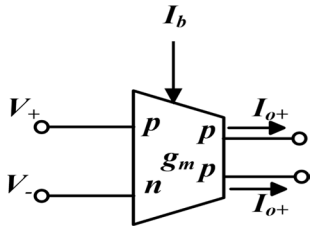



**Fig. 1** Block diagram of dual-output OTA

## 2 Proposed OTA based fractional-order oscillator

### 2.1 OTA

The OTA has many advantages, such as: ease of electronic tuning and cascading of trans-admittance and trans-impedance mode networks. The OTA processes the difference of input voltages and provides current at high impedance output node. The block diagram of dual-output OTA appears in Fig. 1 and the input–output port relationship is given by (1).

$$I_{o+} = g_m(V_+ - V_-) \tag{1}$$

where $g_m$ is depending upon bias current $I_b$ and called as transconductance gain of OTA. The dependence of $g_m$ upon $I_b$ is presented in (2)

$$g_m = \sqrt{\mu_n C_{ox} \frac{W}{L} I_b} \tag{2}$$

The symbols used in (2) have their usual meaning.

In this work, two topologies to construct a sinusoidal FO oscillator are proposed. The first topology consist of a trans-admittance mode fractional-order all-pass filter (FAPF) cascaded with a trans-impedance mode integrator, whereas in the other one the trans-admittance mode FAPF cascaded with a trans-impedance mode differentiator. Both topologies are presented below:

### 2.2 FAPF with integrator based FO oscillator

In this topology of proposed FO oscillator, show in Fig. 2, the two OTAs along with $C_\alpha$ constitute trans-admittance mode FAPF, while the trans-impedance mode integrator is composed using $C_\beta$. By doing regular analysis, characteristics equation of the proposed FO oscillator is given in as below:

$$s^{\alpha+\beta} - \frac{g_{m2}}{C_\beta} s^\alpha + \frac{g_{m2}}{C_\alpha} s^\beta + \frac{g_{m1} g_{m2}}{C_\alpha C_\beta} = 0 \tag{3}$$

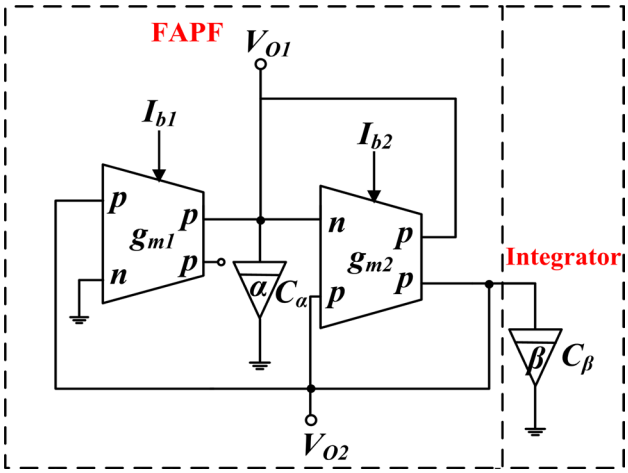



**Fig. 2** FAPF followed by integrator based fractional-order oscillator

The Barkhausen criterion is not sufficient but necessary condition for the sustained oscillations in a fractional-order oscillatory system [12]. Therefore, to make sure that the imaginary conjugate poles exist on $\pm j\omega$ axis, it is assumed that the imaginary conjugate poles are already on $\pm j\omega$ axis and they must satisfy the characteristics equation for a particular frequency of oscillation ($\omega_0$) of the oscillatory system. As a consequence, in complex s-plane, the pair of poles must occur at $\pm j\omega_0$ and both poles must satisfy the characteristics equation. Hence putting $+j\omega_0$ and $-j\omega_0$ in (3) gives (4).

$$\omega_0^{\alpha+\beta} cos\frac{(\alpha+\beta)\pi}{2} - \omega_0{}^{\alpha}\frac{g_{m2}}{C_\beta} cos\frac{\alpha\pi}{2} + \omega_0{}^{\beta}\frac{g_{m2}}{C_\alpha} cos\frac{\beta\pi}{2} + \frac{g_{m1}g_{m2}}{C_\alpha C_\beta} = 0 \tag{4a}$$

$$\omega_0^{\beta} sin\frac{(\alpha+\beta)\pi}{2} + \omega_0{}^{\beta-\alpha}\frac{g_{m2}}{C_\alpha} sin\frac{\beta\pi}{2} - \frac{g_{m2}}{C_\beta} sin\frac{\alpha\pi}{2} = 0 \tag{4b}$$

By solving eqs. (4a) and (4b), the condition of oscillation (CO) and frequency of oscillation (FO) is calculated and given in Table 2 for various cases of $\alpha$ and $\beta$.

### 2.3 FAPF with differentiator based FO oscillator

The second topology is based upon FAPF cascaded with a differentiator and shown in Fig. 3. The trans-admittance mode FAPF is same as in the previous topology, while the trans-impedance mode differentiator is constituted by $OTA_3$, $OTA_4$ and $C_\beta$. The characteristics equation of the proposed FO oscillator is given in (5).

**Table 2** Conditions of oscillation and frequencies of oscillation for various cases

| Topology | Case | CO | FO |
|---|---|---|---|
| FAPF+ | $\alpha = \beta \neq 1$ | $g_{m1} = \frac{g_{m2}}{C_\alpha C_\beta}\left[\frac{C_\alpha - C_\beta}{2cos\frac{\alpha\pi}{2}}\right]^2$ | $\omega_0 = \left[\frac{(g_{m2}/C_\beta)-(g_{m2}/C_\alpha)}{2cos\frac{\alpha\pi}{2}}\right]^{1/\alpha}$ |
| Integrator | $\alpha = 1;\beta \neq 1$ | $g_{m1} = \frac{C_\alpha C_\beta}{g_{m2}}\left[\omega_0^{1+\beta} sin\frac{\beta\pi}{2} - \frac{g_{m2}}{C_\alpha}\omega_0^\beta cos\frac{\beta\pi}{2}\right]$ | $\omega_0^\beta cos\frac{\beta\pi}{2} + \frac{g_{m2}}{C_\alpha}\omega_0^{\beta-1} sin\frac{\beta\pi}{2} - \frac{g_{m2}}{C_\beta} = 0$ |
| | $\alpha \neq 1;\beta = 1$ | $g_{m1} = \frac{C_\alpha C_\beta}{g_{m2}}\left[\omega_0^{1+\alpha} sin\frac{\alpha\pi}{2} + \frac{g_{m2}}{C_\alpha}\omega_0^\alpha cos\frac{\alpha\pi}{2}\right]$ | $\omega_0\, cos\frac{\alpha\pi}{2} + \frac{g_{m2}}{C_\alpha}\omega_0^{1-\alpha} - \frac{g_{m2}}{C_\beta} sin\frac{\alpha\pi}{2} = 0$ |
| FAPF+ | $\alpha = \beta \neq 1$ | $g_{m1} = C_\alpha\left[\left(2cos\frac{\alpha\pi}{2}\sqrt{\frac{g_{m3}g_{m4}}{C_\alpha C\beta}}\right) + \frac{g_{m3}g_{m4}}{g_{m2}C_\beta}\right]$ | $\omega_0 = \left[\frac{(g_{m1}/C_\alpha)-(g_{m3}g_{m4}/g_{m2}C_\beta)}{2cos\frac{\alpha\pi}{2}}\right]^{1/\alpha}$ |
| Differentiator | $\alpha = 1;\beta \neq 1$ | $g_{m2} = \frac{g_{m3}g_{m4}}{C_\beta}\left[1/(\omega_0^{1+\beta} sin\frac{\beta\pi}{2} + \frac{g_{m1}}{C_\alpha}\omega_0^\beta cos\frac{\beta\pi}{2})\right]$ | $\omega_0^\beta cos\frac{\beta\pi}{2} - \frac{g_{m1}}{C_\alpha}\omega_0^{\beta-1} sin\frac{\beta\pi}{2} + \frac{g_{m3}g_{m4}}{C_\alpha C_\beta} = 0$ |
| | $\alpha \neq 1;\beta = 1$ | $g_{m1} = \frac{C_\alpha}{\omega_0^{1-\alpha}}\left[\omega_0 cos\frac{\alpha\pi}{2} + \frac{g_{m3}g_{m4}}{g_{m2}C_\beta}\right]$ | $\omega_0^{1+\alpha} sin\frac{\alpha\pi}{2} - \frac{g_{m3}g_{m4}}{g_{m2}C_\beta}\omega_0^\alpha cos\frac{\alpha\pi}{2} - \frac{g_{m3}g_{m4}}{C_\alpha C_\beta} = 0$ |

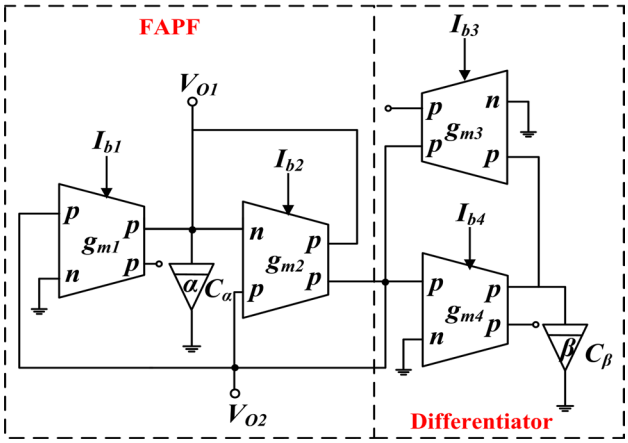



**Fig. 3** FAPF followed by differentiator based fractional-order oscillator

$$s^{\alpha+\beta} + \frac{g_{m3}g_{m4}}{g_{m2}C_\beta}s^\alpha - \frac{g_{m1}}{C_\alpha}s^\beta + \frac{g_{m3}g_{m4}}{C_\alpha C_\beta} = 0 \tag{5}$$

By doing routine analysis as discussed in section 3.1, to ensure the presence of $\pm j\omega_0$ poles, the following equations are obtained:

$$\omega_0^{\alpha+\beta} cos\frac{(\alpha+\beta)\pi}{2} + {\omega_0}^\alpha \frac{g_{m3}g_{m4}}{g_{m2}C_\beta} cos\frac{\alpha\pi}{2} - {\omega_0}^\beta \frac{g_{m1}}{C_\alpha} cos\frac{\beta\pi}{2} + \frac{g_{m3}g_{m4}}{C_\alpha C_\beta} = 0 \tag{6a}$$

$$\omega_0^\beta sin\frac{(\alpha+\beta)\pi}{2} - {\omega_0}^{\beta-\alpha}\frac{g_{m1}}{C_\alpha} sin\frac{\beta\pi}{2} + \frac{g_{m3}g_{m4}}{g_{m2}C_\beta} sin\frac{\alpha\pi}{2} = 0 \tag{6b}$$

**Table 3** Phase difference between $V_{O1}$ and $V_{O2}$

| Structure | Fractional-order case | Integer-order case |
|---|---|---|
| Fig.2 | $\tan^{-1}\left[\frac{C_\alpha\omega^\alpha \sin(\frac{\alpha\pi}{2})}{C_\alpha\omega^\alpha \cos(\frac{\alpha\pi}{2})+g_{m2}}\right]$ | $\tan^{-1}\left[\frac{C.\omega}{g_{m2}}\right]$ |
| Fig.3 | $\tan^{-1}\left[\frac{C_\alpha\omega^\alpha \sin(\frac{\alpha\pi}{2})}{C_\alpha\omega^\alpha \cos(\frac{\alpha\pi}{2})-g_{m1}}\right]$ | $\tan^{-1}\left[\frac{C.\omega}{g_{m1}}\right]$ |

On solving (6), the equation/expression for CO and FO are obtained and summarised for various cases of $\alpha$ and $\beta$ in Table2. The CO and FO are derived from each-other, hence cannot be controlled independently, that is not unusual in fractional-order oscillators. The phase difference between $V_{O1}$ and $V_{O2}$ for the present structures of FO oscillator and its integer-order counterpart is given in Table 3. It is observed that, phase difference in case of FO oscillator can be controlled with the help of $\alpha$.

## 3 Simulation results

### 3.1 Functional verification

To demonstrate functionality of the proposed FO oscillators, SPICE is used with 180nm CMOS technology model parameters. The OTA provided in [36] is used, for quick reference Fig. 4a is showing CMOS structure of the OTA, the W/L ratios of the transistors $M_{3-4}$ are taken as 5.76/0.72, $M_{1,2,5,7,9}$ are 2.16/0.72 and $M_{6,8,10}$ are 1.44/0.72 (all Ws and Ls are in μm). The supply voltages are used as $\pm 1.5V$. The proposed FO oscillators use FOCs approximated through RC ladder network method reported in [37]. The constructed RC network is shown in Fig. 4b, while specifications of the passive components used to realize FOCs can be found in

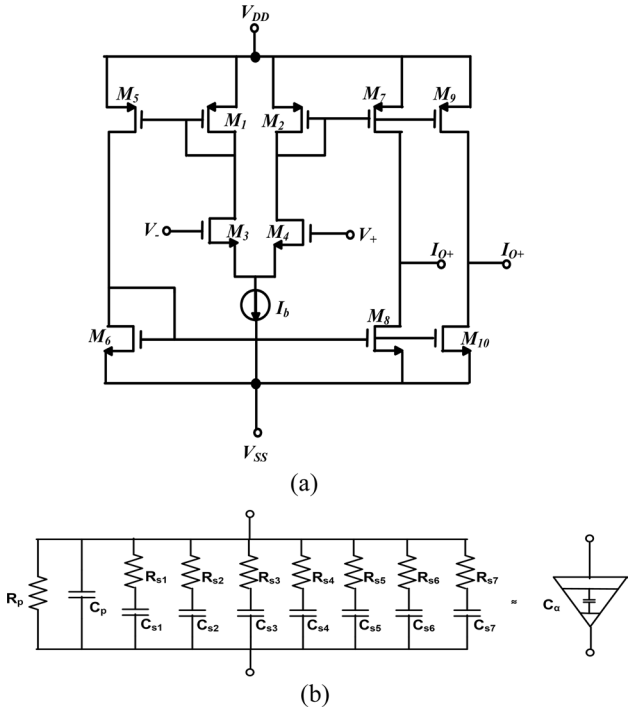



**Fig. 4** Sructures used for simulation **a** OTA's CMOS structure [36] **b** RC network to realize FOC [37]

Table 4. The bandwidth of the approximated FOC is between 1Hz to 1MHz with central frequency of 1kHz.

To design FO oscillator, first of all frequency of oscillation is chosen and by using FO equation for the particular case of $\alpha$ & $\beta$ (given in Table 2) other parameters are selected. Now using CO equation (given in Table 2) the remaining parameter is calculated.

The transient response and frequency spectrum for the proposed FO oscillators based upon FAPF with integrator topology and FAPF with differentiator topology are shown in Figs. 5 and 6 respectively, for different cases of $\alpha$ & $\beta$ and simulation settings for the same are summarized in Table 5. The theoretical and simulated frequency of oscillations for all the cases are also summarized in Table 5. The error between simulated and theoretical values of frequency of oscillation lies within 9% for both the topologies.

Moreover, to show the electronic tuning of frequency of oscillation for the proposed structures of FO oscillators, graphs between $f_0$, $g_{m1}$ and $g_{m2}$ are plotted and shown in Fig. 7, taking the values of $C_\alpha$, $C_\beta$, $g_{m3}$ and $g_{m4}$ from Table 5 for $\alpha$=$\beta$=0.5 case. It is observed from Fig. 7 that by electronically adjusting $g_{m1}$ and $g_{m2}$ simultaneously, the frequency of oscillation can be tuned.

Generally, to generate very low frequencies, a high value of $C_\alpha$ or a low value of OTA's transconductance is required. In the present work, the possibility to generate very low frequency without changing the value of $C_\alpha$ has been explored. The FAPF with integrator topology is designed to generate a sinusoidal waveform of 5Hz for which $g_{m1}$ and $g_{m2}$ are chosen as $3.96\mu A/V$ and $15.85\mu A/V$ respectively, while $C_\alpha$ and $C_\beta$ are taken as $2\mu\mho/s^\alpha$ and $1\mu\mho/s^\beta$ respectively with $\alpha = \beta = 0.5$. The frequency of the generated waveform is 4.9Hz against the designed value of 5Hz. The transient response and frequency spectrum are shown in Fig. 8

## 3.2 Robustness

To verify the robustness of the proposed circuits, PVT and Monte-Carlo analysis are performed for both topologies, but due to space constraint the results for $\alpha = \beta = 0.5$ are shown.

**Table 4** Specifications for passive components used to construct approximated FOC in Fig. 4b

| $C_\alpha$ **for** $\alpha = 0.5$ | $R_p$<br>$C_p$ | $R_{s1}$<br>$C_{s1}$ | $R_{s2}$<br>$C_{s2}$ | $R_{s3}$<br>$C_{s3}$ | $R_{s4}$<br>$C_{s4}$ | $R_{s5}$<br>$C_{s5}$ | $R_{s6}$<br>$C_{s6}$ | $R_{s7}$<br>$C_{s7}$ |
|---|---|---|---|---|---|---|---|---|
| $1\mu\mho/s^\alpha$ | 259.5kΩ<br>157.84pF | 173.01kΩ<br>57.8nF | 69.5kΩ<br>23.12nF | 27.68kΩ<br>9.248nF | 11.07kΩ<br>3.69nF | 4.34kΩ<br>1.48nF | 1.77kΩ 592pF | 708.63Ω<br>236.8pF |
| $2\mu\mho/s^\alpha$ | 129.75kΩ<br>315.68pF | 86.51kΩ<br>115.6nF | 34.75kΩ<br>46.24nF | 13.84kΩ<br>18.496nF | 5.535kΩ<br>7.38nF | 2.17kΩ<br>2.96nF | 0.885kΩ<br>1.184nF | 354.315Ω<br>473.6pF |

**Table 5** Simulation settings for proposed FO oscillators

| Topology | Case | $g_{m1}$ $[\mu A/V]$ | $g_{m2}$ $[\mu A/V]$ | $g_{m3}$ $[\mu A/V]$ | $g_{m4}$ $[\mu A/V]$ | $C_\alpha$ | $C_\beta$ | Theoretical $f_0$ $[Hz]$ | Simulated $f_0$ $[Hz]$ |
|---|---|---|---|---|---|---|---|---|---|
| FAPF+ | $\alpha = \beta = 0.5$ | 25 | 100 | – | – | $2\mu\mho/s^\alpha$ | $1\mu\mho/s^\beta$ | 198.1 | 180.8 |
| Integrator | $\alpha = 1, \beta = 0.5$ | 160.51 | 10 | – | – | $1\mu\mho/s^\alpha$ | $1\mu\mho/s^\beta$ | 28.5 | 27.3 |
| FAPF+ | $\alpha = \beta = 0.5$ | 241.42 | 100 | 100 | 100 | $1\mu\mho/s^\alpha$ | $1\mu\mho/s^\beta$ | 1.55k | 1.41k |
| Differentiator | $\alpha = 0.5, \beta = 1$ | 21.99 | 100 | 100 | 100 | $1\mu\mho/s^\alpha$ | $1\mu\mho/s^\beta$ | 103.9 | 96.5 |

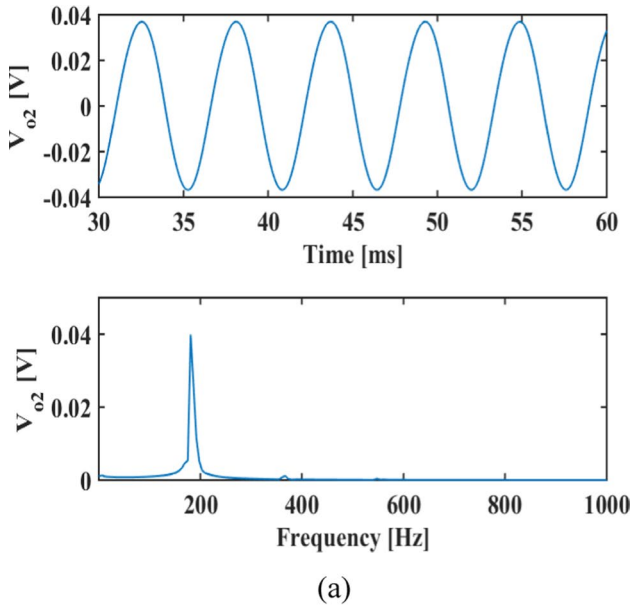



**Fig. 5** Proposed FO oscillator's transient response and frequency spectrum for FAPF with integrator topology **a** $\alpha$=$\beta$=0.5 **b** $\alpha$=1 & $\beta$=0.5

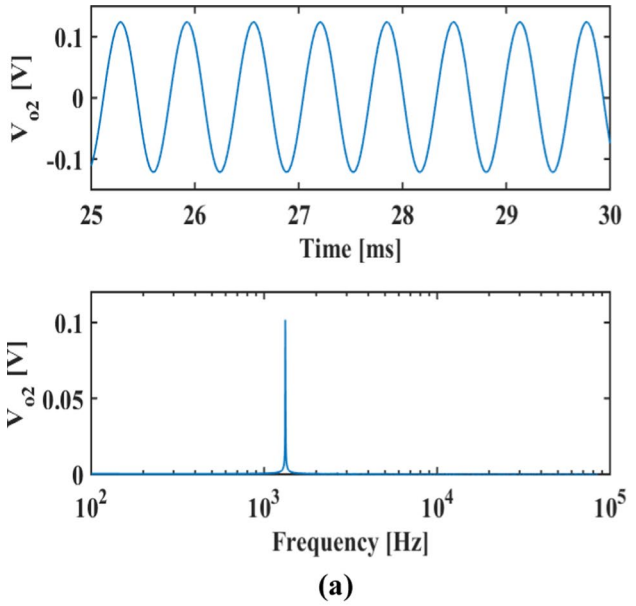



**Fig. 6** Proposed FO oscillator's transient response and frequency spectrum for FAPF with differentiator topology **a** $\alpha$=$\beta$=0.5 **b** $\alpha$=0.5 & $\beta$=1

#### 3.2.1 PVT analysis

In PVT analysis, PVT stands for the Process, Voltage, and Temperature. In order to test the robustness of the proposed FO oscillators, they are simulated at different corners of process, voltage and temperature.

Two letter identifiers are used to name a process corner, where the first letter refers to the NMOS corner and the second letter refers to the PMOS corner. Basically, three corners exist: typical, fast and slow. Fast and slow corners exhibit carrier mobilities that are higher and lower than normal, respectively. Therefore, there are five possible corners:

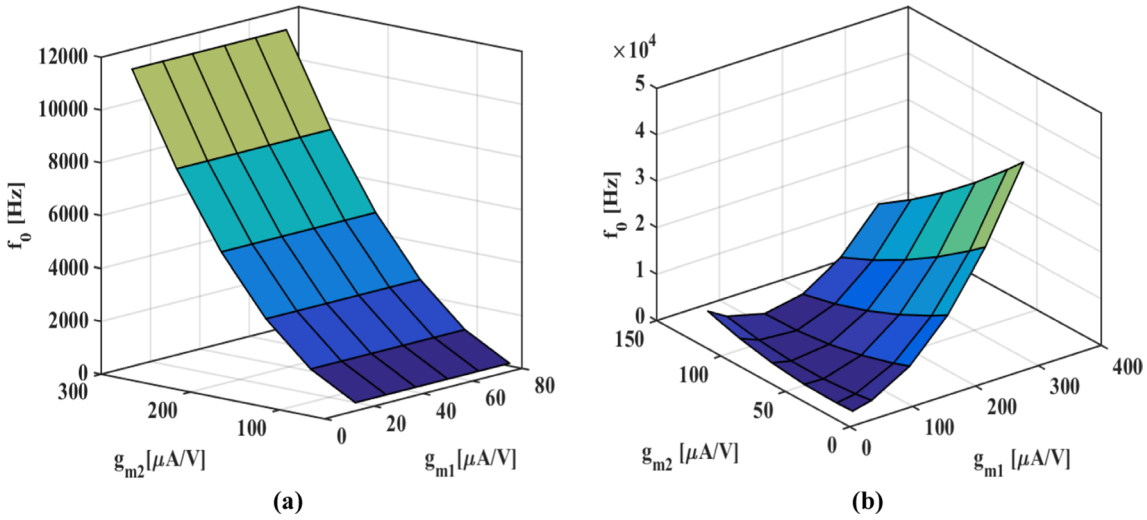



**Fig. 7** Electronic tuning of $f_0$ with $g_{m1}$ and $g_{m2}$ **a** FAPF with integrator topology **b** FAPF with differentiator topology

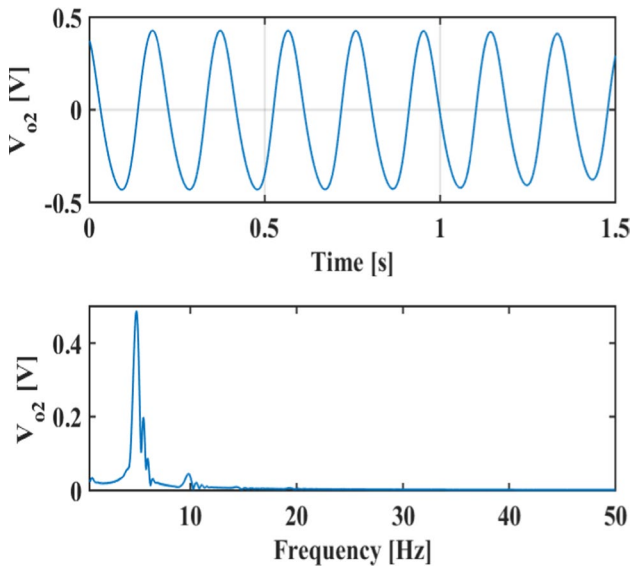



**Fig. 8** Proposed FO oscillator's transient response and frequency spectrum in VLF mode

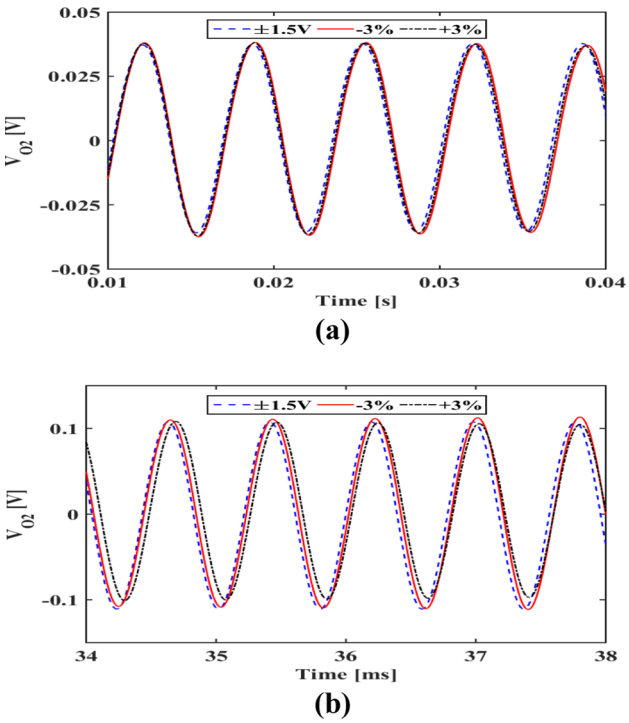



**Fig. 9** Transient response for supply voltage variations **a** FAPF and integrator topology **b** FAPF and differentiator topology

typical-typical (TT), fast-fast (FF), slow-slow (SS), fast-slow (FS), and slow-fast (SF).

The values of frequency of oscillation and percentage error from TT corner point for all the remaining four corner points are listed in Table 6 whereas Figs. 9 and 10 represent the transient response in presence of supply voltage variations (1.5V and 1.5V±3%) and temperature variations (0°, 27° and 70°) respectively for both the topologies of FO oscillator. It is clear from Table 6 that the error in frequency of oscillation lies within 5% for various corners and no abrupt change in the transient response in the presence of supply voltage and temperature variations is observed, which verify the robustness of the proposed circuits for PVT variations.

### 3.2.2 Monte-Carlo analysis

To draw further attention on sensitivity of the proposed FO oscillators with respect to random variations in FOC i.e. $C_\alpha$ & $C_\beta$, the passive components used for FOC realization

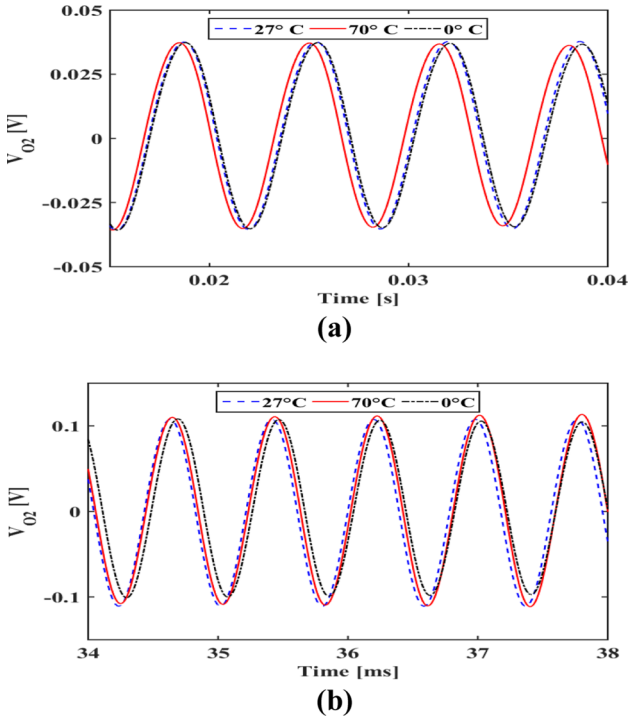



**Fig. 10** Transient response for temperature variations **a** FAPF and integrator topology **b** FAPF and differentiator topology

**Table 6** Oscillation frequency of fractional-order oscillator for all five process corners for $\alpha = \beta = 0.5$

| Process corner | Oscillation Frequency (Hz) | % error from $'tt'$ corner | Topology |
|---|---|---|---|
| TT | 180.93 | – | FAPF+ |
| FF | 188.93 | 4.50% | Integrator |
| FS | 183.75 | 1.63% | |
| SF | 183.13 | 1.29% | |
| SS | 172.88 | 4.38% | |
| TT | 1.41$k$ | – | FAPF+ |
| FF | 1.38$k$ | 2.29% | Differentiator |
| FS | 1.36$k$ | 3.82% | |
| SF | 1.36$k$ | 3.82% | |
| SS | 1.35$k$ | 4.58% | |

are varied in Gaussian manner with a tolerance of ±3% in resistances and capacitances and Monte-Carlo simulation is done on 250 random samples. The random variation in the value of $C_\alpha$ & $C_\beta$ affect the condition of oscillation. The histogram of the frequency of oscillation of $V_{o2}$ for all the samples is shown in Fig. 11 for $\alpha = \beta = 0.5$. It is observed from the above analysis that the spread in frequency of oscillation is within $(6 * \sigma)$ range which verifies the robustness of

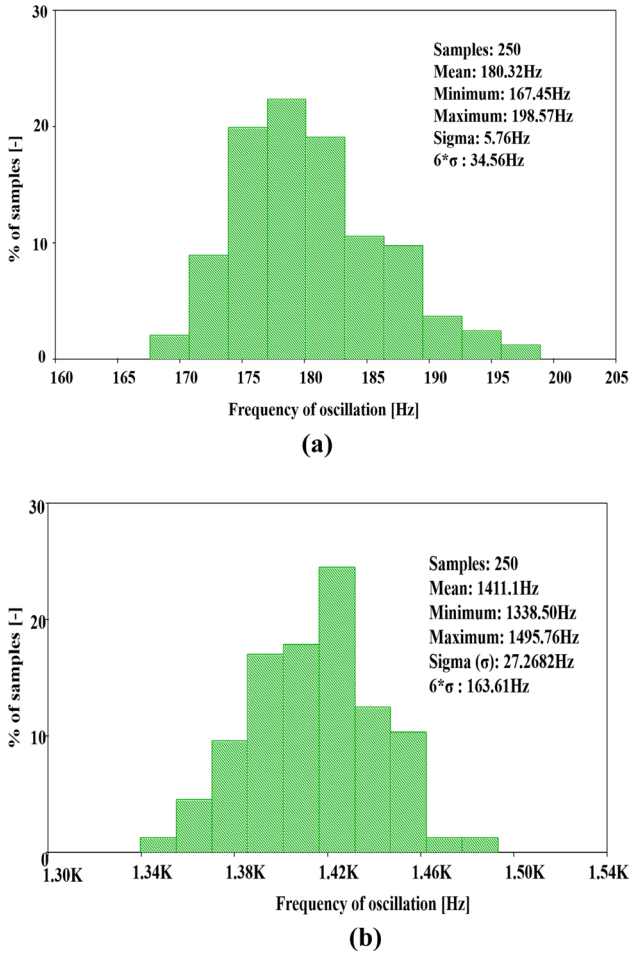



**Fig. 11** Monte Carlo Analysis for random variations in $C_\alpha$ **a** FAPF and integrator topology **b** FAPF and differentiator topology

the circuit in-spite the changes introduced in the condition of oscillation.

## 4 Stability and sensitivity analysis

### 4.1 Stability analysis

The stability of the proposed FO oscillators has been verified using root locus method for linear fractional-order system described in [38]. The integer-order s-plane is mapped onto fractional-order $s^\alpha$-plane and the $\pm\frac{\pi}{2}$ are mapped onto $\pm\frac{\alpha\pi}{2}$ while non-physical region from $\pm\pi$ to $\pm\alpha\pi$. The pole plots for characteristic equation of both the topologies of FO oscillator has been shown in Fig. 12 for $\alpha = \beta = 0.5$. As lowest common divisor $\lambda$ is taken as 2, the total number of poles are 2 for both the topologies. The boundaries for stable and unstable region are separated by $\pm\frac{\pi}{2\lambda}$ and shown as dotted dash lines. As it may be witnessed from Fig. 12 (a) and (b) that both the roots lie on the line $\pm\frac{\pi}{4}$, hence

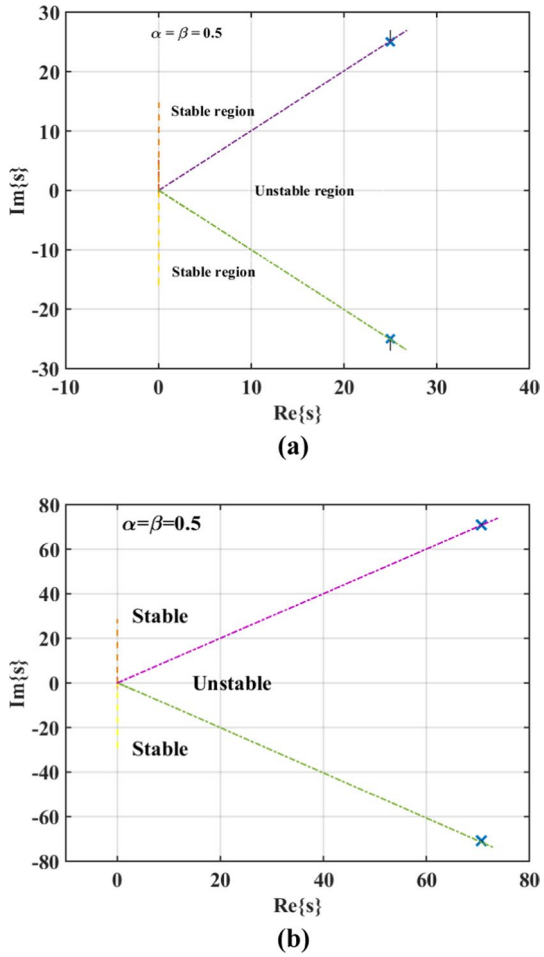



**Fig. 12** FO oscillator's pole plot in W-plane for $\alpha = \beta = 0.5$ **a** FAPF with integrator based topology **b** FAPF with differentiator based topology

proposed structures oscillate and none of the roots are in unstable region so the proposed FO oscillators do not show unstable behaviour.

### 4.2 Sensitivity analysis

The sensitivity of parameter (P) with respect to change in the value of parameter (Q) is defined as:

$S_Q^P = (\frac{Q}{P})(\frac{\delta P}{\delta Q})$

The sensitivity analysis is one of the most important analysis to study the performance of a circuit. Thus, the sensitivity analysis for both of the proposed topologies with respect to various circuit components has been carried out. The sensitivity equations are given in Tab.7 and it is clear that the proposed FO oscillators are more sensitive towards lower values of the fractional-order $\alpha$. As the value of fractional-order $\alpha$ approaches unity the sensitivity reduces.

**Table 7** Sensitivity with respect to various parameters

| $\lvert S_Y^{\omega_0} \rvert$ | FAPF+Integrator | FAPF+Differentiator |
|---|---|---|
| $\lvert S_{g_{m1}}^{\omega_0} \rvert$ | 0 | $\frac{1}{\alpha}\left[1 - \frac{g_{m3}g_{m4}C_\alpha}{g_{m1}g_{m2}C_\beta}\right]$ |
| $\lvert S_{g_{m2}}^{\omega_0} \rvert$ | $\frac{1}{\alpha}$ | $\frac{1}{\alpha}\left[1 - \frac{g_{m1}g_{m2}C_\beta}{g_{m3}g_{m4}C_\alpha}\right]$ |
| $\lvert S_{g_{m3}}^{\omega_0} \rvert$ | – | $\lvert S_{g_{m2}}^{\omega_0} \rvert$ |
| $\lvert S_{g_{m4}}^{\omega_0} \rvert$ | – | $\lvert S_{g_{m2}}^{\omega_0} \rvert$ |
| $\lvert S_{C_\alpha}^{\omega_0} \rvert$ | $\frac{C_\beta}{\alpha(C_\alpha - C_\beta)}$ | $\lvert S_{g_{m1}}^{\omega_0} \rvert$ |
| $\lvert S_{C_\beta}^{\omega_0} \rvert$ | $\frac{C_\alpha}{\alpha(C_\alpha - C_\beta)}$ | $\lvert S_{g_{m2}}^{\omega_0} \rvert$ |

## 5 Conclusion

Two new OTA-based topologies for FO oscillators are presented in this paper. Both topologies use trans-admittance mode FAPF as one of the block and trans-impedance mode integrator/differentiator as the other block. Both topologies are verified through SPICE simulations using 180nm CMOS technology node. The frequency of oscillation of the proposed FO oscillators can be electronically tuned and the proposed circuits have also been verified to work in VLF mode. The stability is verified using pole plots with the help of *forlocus* function of MATLAB. The mathematical equations for sensitivity of the oscillation frequency with respect to various circuit parameters have been derived and found that the proposed circuits are lesser sensitive towards higher values of $\alpha$. The robustness of the proposed circuits has also been ensured via process-voltage-temperature (PVT) and Monte Carlo analyses and no abrupt changes have been found in the transient responses of the proposed circuits.

**Data availibility** The datasets generated during the current study are available from the corresponding author on reasonable request.

## References

1. Debnath, L. (2013). Recent applications of fractional calculus to science and engineering. *International Journal of Mathematics and Mathematical Sciences, 54,* 3413–42.
2. Elwakil, A. S. (2010). Fractional-order circuits and systems: An emerging interdisciplinary research area. *IEEE Circuits and Systems Magazine, 10*(4), 40–50.
3. Dorcak, L., Valsa, J., Gonzalez, E., Terpak, J., Petras, I., & Pivka, L. (2013). Analogue realization of fractional-order dynamical systems. *Entropy, 15*(10), 4199–214.
4. Podlubny, I., Petras, I., Vinagre, B. M., Leary, P. O., & Dorcak, L. (2012). Analogue realizations of fractional-order controllers. *Nonlinear Dynamics, 29*(4), 281–296.

5. Das, S., & Pan, I. (2012). Fractional order signal processing-introductory concepts and applications. In: Springer Briefs in Applied Sciences and Technology.
6. Sheng, H., Chen, Y. Q., & Qiu, T. S. (2011). *Fractional processes and fractional-order signal processing: Techniques and applications*. New York: Springer.
7. Varshney, G., Pandey, N., & Pandey, R. (2021). Electronically tunable multifunction transadmittance-mode fractional-order filter. *Arabian Journal for Science and Engineering, 46,* 1067–78. https://doi.org/10.1007/s13369-020-04841-8.
8. Geddes, L. A., & Baker, L. E. (1989). *Principles of applied biomedical instrumentation* (3rd ed.). New York: John Wiley and Sons Ltd.
9. Faria, A. C., Veiga, J., Lopes, A. J., & Melo, P. L. (2016). Forced oscillation, integer and fractional-order modeling in asthma. *Computer methods and programs in biomedicine, 128,* 12–26.
10. Tlelo-Cuautle, E., Pano-Azucena, A.D., Guillén Fernández, O., & Silva-Juárez, A. (2020) Analog/digital implementation of fractional order chaotic circuits and applications, Springer.
11. Muñiz-Montero, C., Sánchez-Gaspariano, L.A., Sánchez-López, C., González-Díaz, V.R., & Tlelo Cuautle, E. (2017). On the electronic realizations of fractional-order phase-lead-lag compensators with OpAmps and FPAAs. In: Fractional order control and synchronization of chaotic systems, 688, 131-64. ISBN: 978-3-319-50248-9.
12. Radwan, A. G., Elwakil, A. S., & Soliman, A. M. (2008). Fractional-order sinusoidal oscillator: Design procedure and practical examples. *IEEE Transactions on Circuits and Systems I, 55*(7), 2051–63.
13. Said, L. A., Radwan, A. G., Madian, A. H., & Soliman, A. M. (2015). Fractional order oscillators based on operational transresistance amplifiers. *AEU-International Journal of Electronics and Communications, 69*(7), 988–1003.
14. Fouda, M., Soltan, A., Radwn, A. G., & Soliman, A. M. (2016). Fractional-order multi-phase oscillators design and analysis suitable for higher-order PSK applications. *Analog Integrated Circuits and Signal Processing, 87*(2), 301–12.
15. Said, L. A., Radwan, A. G., Madian, A. H., & Soliman, A. M. (2017). Three fractional-order-capacitorsbased oscillators with controllable phase and frequency. *Journal of Circuits, Systems and Computers, 26*(10), 1750160–78. https://doi.org/10.1142/S0218126617501602.
16. Said, L. A., Radwan, A. G., Madian, A. H., & Soliman, A. M. (2016). Two-port two impedances fractional order oscillators. *Microelectronics Journal, 55*(9), 40–52.
17. Kartci, A., Herencsar, N., Koton, J., Psychalinos, C.(2017) Compact MOS-RC voltage-mode fractional-order oscillator design. In: Procceeding of European Conference on Circuit Theory and Design (ECCTD), Catania, Italy, pp. 1-4
18. Kartci, A., Herencsar, N., Brancik, L., & Salama, K.N. (2018). CMOS-RC colpitts oscillator design using floating fractional-order inductance simulator. In: Procceding of the 2018 61st IEEE International Midwest Symposium on Circuits and Systems (MWSCAS). Windsor, Canada, 905-908.
19. Kartci, A., Agambayev, A., Hassan, A.H., Bagci, H., & Salama, K.N. (2018). Experimental verification of a fractional-order wien oscillator built using solid-state capacitors. In: Proc. Of the 2018 61st IEEE International Midwest Symposium on Circuits and Systems (MWSCAS). Windsor, Canada, 544-545.
20. Mishra, S. K., Gupta, M., & Upadhyay, D. K. (2019). Design and implementation of DDCC-based fractional-order oscillator. *International Journal of Electronics, 106*(4), 581–98.
21. Kartci, A., Herencsar, N., Dvorak, J., & Vrba, K. (2019). VDIBA-based fractional-order oscillator design. In: Proc. of the 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), Budapest, Hungary, 744-747.
22. Mishra, S. K., Gupta, M., & Upadhyay, D. K. (2019). Compact design of four-phase fractional-order oscillator with independent phase and frequency control. *Indian Journal of Physics, 93,* 891–901.
23. Mishra, S. K., Gupta, M., & Upadhyay, D. K. (2019). Compact design of four-phase fractional-order oscillator with independent phase and frequency control. *Indian Journal of Physics, 93,* 891–901.
24. Elwy, O., Said, L. A., Madian, A. H., & Radwan, A. (2019). All possible topologies of the fractional-order Wien oscillator family using different approximation techniques. *Circuits, Systems, and Signal Processing, 38,* 3931–51.
25. Ahmed, G.M., Said, L.A., Madian, A.H., & Radwan, A.G. (2019). Fractional-order oscillators based on double op-amp. In: Proc. of 2019 Fourth International Conference on Advances in Computational Tools for Engineering Applications (ACTEA), Beirut, Lebanon, 2019, 1-4.
26. Tsirimokou, G., Psychalinos, C., Elwakil, A., & Maundy, B. Fractional-order multiphase sinusoidal oscillator design using current-mirrors. In: Proc. of 2018 41st International Conference on Telecommunications and Signal Processing (TSP), Athens, 2018, pp. 1-4, https://doi.org/10.1109/TSP.2018.8441399.
27. Pradhan, A., Subhadhra, K.S., Atique, N., Sharma, R.K., Gupta, S.S. (2018). MMCC-based current mode fractional-order voltage-controlled oscillators. In: Proc. of 2018 2nd International Conference on Inventive Systems and Control (ICISC), Coimbatore, 763-768, https://doi.org/10.1109/ICISC.2018.8398901.
28. Subhadhra, K. S., Sharma, R. K., & Gupta, S. S. (2020). Realisation of some current-mode fractional-order VCOs/SRCOs using multiplication mode current conveyors. *Analog Integrated Circuits and Signal Processing, 103,* 31–55. https://doi.org/10.1007/s10470-020-01590-4.
29. Sotner, R., Jerabek, J., Domansky, O., Herencsar, N., Kartci, A., & Dvorak, J. (2018). Practical design of fractional-order oscillator employing simple resonator and negative resistor. In: Proceeding of 2018 10th Int. Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Moscow, Russia, 1-4.
30. Sotner, R., Jerabek, J., Polak, L., Langhammer, L., Stolarova, H., Petrzela, J., et al. (2021). On the performance of electronically tunable fractional-order oscillator using grounded resonator concept. *AEU-International Journal of Electronics and Communications, 129,* 153540.
31. Swamy, M.N.S., Raut, R. & Tang, Z. (2004). Generation of new OTA-Coscillator structures using network transposition. Proceeding of 47th IEEE Midwest Symp Circuits and Systems, I. 73-76.
32. Montree, K., Fabian, K., Tomasz, K., & Costas, P. (2021). Multiple-input universal filter and quadrature oscillator using multiple-input operational transconductance amplifiers. *IEEE Access, 9,* 56253–56263.
33. Senani, R., Bhaskar, D. R., Singh, V. K., & Sharma, R. K. (2016). *Sinusoidal oscillators and waveform generators using modern electronic circuit building blocks*. Berlin: Springer.
34. Wen, D., Zhou, Y., & Li, X. (2015). A critical review: Coupling and synchronization analysis methods of EEG signal with mild cognitive impairment. *Frontiers in Aging Neuroscience*. https://doi.org/10.3389/fnagi.2015.00054.
35. Fouda, M. E., Soltan, A., Radwan, A. G., & Soliman, A. M. (2016). Fractional- order multi-phase oscillators design and analysis suitable for higher-order PSK applications. *Analog Integrated Circuits and Signal Processing, 87*(2), 301–312.

36. Tsukutani, T., Sumi, Y., & Fukui, Y. (2006). Electronically controlled current-mode oscillators using MO-OTAs and grounded capacitors. *Frequenz, 60*(11–12), 220–23.
37. Valsa, J., Dvorak, P., & Friedl, M. (2011). *Network model of the CPE. Radioengineering, 20,* 619–26.
38. Radwan, A. G., Soliman, A. M., Elwakil, A. S., & Sedeek, A. (2009). On the stability of linear systems with fractional-order elements. *Chaos, Solitons and Fractals, 40*(5), 2317–2328.

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

**Garima Varshney** Garima Varshney received B.Tech. degree in Electronics and Communication Engineering (ECE) from Aligarh Muslim University and M.Tech. degree in VLSI design and Embedded systems from Delhi Technological University, Delhi. At present, she is a research scholar in the department of Electronics & Communication Engineering, DTU. Her research interest includes fractional-order analog signal processing and signal generating circuits. She has published 4 journal papers, 4 conference papers and 1 book chapter.

**Neeta Pandey** Neeta Pandey is Professor in ECE department, Delhi Tech nological University. She received her M.E. in Microelectronics from BITS Pilani and Ph.D. from Guru Gobind Singh Indraprastha University, Delhi. She has served in Central Electronics Engineering Research Institute, Pilani, Indian Institute of Technology, Delhi, Priyadarshini College of Computer Science, Noida and Bharati Vidyapeeth's College of Engineering, Delhi in various capacities. She is having approximately 27 years of teaching and research experience in Electronics and Communication Engineering. She has published papers in international, national journals of repute and conferences. Her research interests are in analog and digital VLSI design. She is life member of ISTE and senior member of IEEE,IEEE WIE for 13 years.

**Rajeshwari Pandey** Rajeshwari Pandey received her B.Tech. (Electronics and Telecommunication) from J. K. Institute of Applied Physics, University of Allahabad and her M.E. (Electronics and Control) from BITS, Pilani, Rajasthan, India and Ph.D. from Faculty of Technology, Delhi University, India. She has served BITS Pilani, AERF, Noida and Priyadarshini College of Computer Science, Noida in various capacities. Currently, she is Professor in Department of Electronics and Communication Engineering, Delhi Technological University, Delhi. Her research interests include analog integrated circuits and microelectronics. She is life member of ISTE, IETE and IEEE and IEEE WIE.

Optical and Quantum Electronics (2022) 54:512
https://doi.org/10.1007/s11082-022-03890-7

# Design of novel MIMO UOWC link using gamma–gamma fading channel for IoUTs

**Dushyant Singh Chauhan**[1] · **Gurjit Kaur**[1] · **Dinesh Kumar**[1]

Received: 26 January 2022 / Accepted: 3 June 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

**Abstract**
Internet of underwater things (IoUT) describes the network of physical objects interconnected smartly in the underwater medium for monitoring/carrying out various underwater operations. IoUT allows a network of autonomous underwater vehicles, remotely operated underwater vehicles and underwater sensors to communicate with each other and collect, sensing, transmitting information to control centers located above the sea surface at high speed. This information can be used for various tasks like detection of tsunamis, animal health monitoring, shipwrecks discovery, crash surveying, etc. Optical beam undergoes fluctuation in its intensity due to fading effect in underwater optical wireless communication (UOWC). Fading is a phenomenon in which there is a variation in the refractive index due to salinity and temperature variations. It also refers as underwater optical turbulence. In this paper, we propose a model that best suits this degradation in the optical intensity. We analytically study the performance of spatial diversity in UOWC systems under optical heterodyne detection and intensity-modulated with direct detection schemes. We have also investigated various modulation techniques and obtained closed-form expressions for average bit error rate. We derive the closed-form expressions for probability density function and cumulative density function in terms of the Meijer-G function, which help in computing the system performance in terms of outage probability, and ergodic capacity.

**Keywords** Internet of underwater things (IoUT) · Scintillation index (SI) · Equal gain combiner (EGC) · Autonomous underwater vehicles (AUVs) · Multiple-input multiple-output (MIMO)

## 1 Introduction

Underwater communications have gained a lot of interest nowadays due to their numerous applications like detection of tsunamis, animal health monitoring, shipwrecks discovery, crash surveying, etc. (Kaushal and Kaddoum 2016). These applications require

✉ Gurjit Kaur
gurjeet_kaur@rediffmail.com

Dinesh Kumar
dineshkumar@dtu.ac.in

[1] Department of ECE, Delhi Technological University, Delhi 102045, India

huge amount of data generated from various underwater sensors (Hanjiang et al. 2022) installed under the sea. Microsoft has installed the underwater datacentres in 2018 and after two years they found that due to the cold water of the sea the energy consumption reduces up to 30% caused by device cooling (Zhang et al. 2020). Traditional underwater technologies like radio frequency and acoustic communication are unable to deal with this huge amount of generated data due to low data transmission rate (i.e., kbps) and high latency (i.e., sec) resulting in the long transmission delays (Sozer et al. 2000). So, optical communication became the state of the art which provides high data rate (i.e., Gbps) and low transmission latency. It has been observed that attenuation may become weak when operated in the blue-green of visible spectrum (i.e., 450 nm to 550 nm) (Ergul et al. 2015). But there are several challenges with the optical beam in sea water like absorption, scattering and fading of signals which reduces the transmission range and deteriorates the system performance (Oubei et al. 2015). Fading is a phenomenon in which there is a variation in the refractive index (RI) due to salinity and temperature variations. This UOT results in the distortion of phase and intensity of laser diodes while propagating inside the sea, which significantly degraded the performance of the system. Various techniques and models have been presented to reduce these effects. Model based on Beer lambers law is simple for UOWC channel but its accuracy is less (Hanson et al. 2010). Monte-Carlo model is easy to solve through numerical integration but it contains some statistical errors moreover it is slow (Li et al. 2012). Analytical RTE model is accurate but difficult to solve analytically (Cochenour et al. 2008). Some statistical distribution functions are available, which explains the behaviour of UOT, and these function helps in designing the UOWC system more reliable and robust. FSO atmospheric turbulence model using log-normal distribution is used to explain the laser fluctuations in the ocean (Gerçekcioğlu 2014) utilizes. Since refractive index fluctuations in sea water and irradiance fluctuations in free-space are bit different phenomenon. So, log-normal distribution function was not able to explain the behaviour of irradiance fluctuation of laser in underwater during moderate and strong turbulence. So, further there is a need to investigate some other statistical models that best suited for UOT. It has been observed that gamma–gamma (GG) distribution well explains the fading effect during moderate and strong UOT (Fu and Du 2018). Optical amplification technique with multiple receivers is used to investigate the performance of UOWC systems (Boucouvalas et al. 2016). Spatial diversity MIMO scheme along with OOK modulation technique is used to study the performance of UOWC system (Jamali et al. 2018). Multi-PPM technique along with the spatial diversity has been utilize to investigate the performance of the system (Peppas et al. 2017). Various modulation techniques are available for optical communication. On–off keying (OOK) modulation is implemented in UOWC system, this is a binary level modulation technique (Wang et al. 2017). 2.7 Gbps data rate has been achieved by using NRZ-OOK modulation technique (Liu et al. 2017). But energy efficiency of RZ-OOK modulation is more in comparison with its counterpart NRZ-OOK at a cost of high channel bandwidth. A different modulation technique known as orbital angular momentum (OAM) has been demonstrated in which discrete OAM modes are used to modulate the user data in spatial domain (Peng et al. 2018).

Our goal is to address the challenges that occur in the underwater surveillance system due to turbulence-induced UOWC fading channel and to compute the performance metrics. Precisely, we aim to minimize the average BER and outage probability while maximizing the capacity of the channel. Towards achieving these goals:

- We propose an UOWC model to describe the UOT in the surveillance system.
- We apply the spatial diversity MIMO scheme to increase total transmitted power by the number of transmitters which improves the range of the UOWC system.
- Since OOK modulation has been used in many research works, we have analyzed the UOWC system in IM/DD scheme using quadrature amplitude modulation (QAM) and phase-shift keying (PSK).
- Analytically we have obtained mathematical expressions for OP, ABER, and EC that best describe the performance of the UOWC system (Fig. 1).

The paper is organized as follows. We present the UOWC system and channel modeling by using Rytov variance and applied spatial diversity MIMO using the EGC technique in Sect. 2. Performance metrics in terms of OP, ABER, and OP have been analytically evaluated using the MeigerG function. In Sect. 3, the numerical results along with the comparison of the exponential generalized-gamma (EGG) model with our proposed model are presented. Finally, Sect. 4 concludes the paper.

## 2 System and channel model

Modeling of channels plays a vital role in determining the performance of any communication system. Several studies depicted that incident light is affected due to three major phenomena scattering, absorption (Matta et al. 2022), and UOT in the underwater channel. When an optical signal propagates through an underwater medium, then particles present in the medium cause the light wave to absorb and scatter. The classical Kolmogorov model

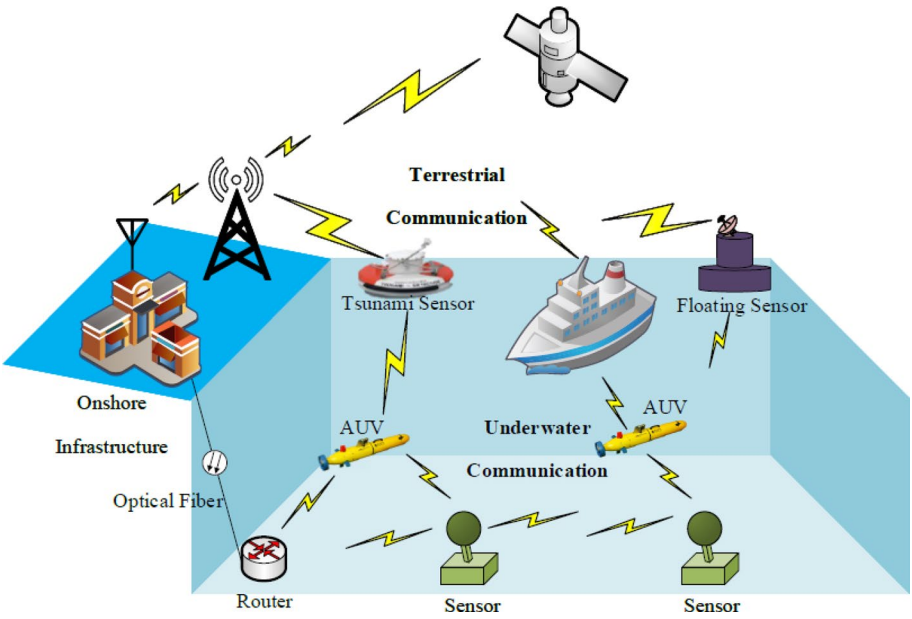



**Fig. 1** Arrangement of underwater optical wireless communication network consisting of various sensors and AUVs

(Ata and Baykal 2014) is used to calculate the power spectrum of the refractive index (RI) in the underwater medium as Eq. (1) with various parameters for its calculation are shown in Table 1.

$$\phi_n^{ocean}(k) = \frac{1}{2.577} \times 10^{-8} \frac{k^{-11/3}}{\varepsilon^{1/3}} \chi_T \times \left( e^{-A_T \delta} - \frac{2}{\tau} e^{-A_{TS} \delta} + \frac{1}{\tau^2} e^{-A_S \delta} \right) \times \left[ 1 + 2.35(k\eta)^{2/3} \right] \tag{1}$$

with $\delta = 8.284(k\xi)^{4/3} + 12.978(k\xi)^2$, $A_T = 1.863 \times 10^{-2}, K = 2\pi/\lambda$, $A_{TS} = 1.9 \times 10^{-4}$, $A_S = 9.41 \times 10^{-3}$.

The Rytov variance model is used to calculate SI values for various levels of turbulence (Ata and Baykal 2014) as

$$\sigma_R^2(z, \lambda) = 78.87 z K^2 \int_0^1 \int_0^\infty k \phi_n^{ocean}(k) \left[ 1 - \cos\left( \frac{zk^2}{K} \xi \right) \right] dk d\xi \tag{2}$$

where $\sigma_R^2$ is known as Rytov's variance A log-normal distribution model has been used for weak UOT is used to determine the channel capacity of UOWC systems (Boucouvalas et al. 2016). Strong underwater turbulence can be studied by using gamma–gamma (GG) distribution, and BER has been evaluated for OOK modulation. If I being the optical intensity and is a random variable then the probability density function (PDF) of GG distribution can be expressed as Wang et al. (2020). GG statistical model has been used under weak to strong turbulence conditions due to temperature and salinity variations, under different spatial diversity techniques. Practically our application has to perform in medium and strong fading regions so, GG model has been used.

$$f_I(I) = 2(\alpha\beta)^{(\alpha+\beta)/2} K_{\alpha-\beta}\left( 2\sqrt{\alpha\beta I} \right) \frac{I^{(\alpha+\beta)/2-1}}{\Gamma(\alpha)\Gamma(\beta)}, \tag{3}$$

where $K$ is the modified Bessel function, $\alpha$ and $\beta$ are parameters associated with the Rytov variance as Jamali et al. (2018)

$$\alpha = \frac{1}{\exp\left[ \frac{0.49\sigma_R^2}{\left(1+1.11\sigma_R^{12/5}\right)^{7/6}} \right] - 1} \tag{4}$$

**Table 1** Various factors used to calculate the power spectrum of RI

| Symbol | Parameter | Values |
|---|---|---|
| $k$ | Magnitude of spatial frequency | 15 MHz at $\lambda = 0.417$ μm |
| $\eta$ | Kolmogorov scale | $10^{-3}$ m |
| $\chi_T$ | Dissipation rate of temperature | $10^{-2} - 10^{-10}$ K$^2$/s to |
| $\tau$ | Temperature and salinity driven RI | [−5 and 0] |
| $\varepsilon$ | Rate of dissipation of the turbulent kinetic energy | $10^{-1}$ m$^2$/s$^3$ to $10^{-10}$ m$^2$/s$^3$ |

$$\beta = \frac{1}{\exp\left[\frac{0.51\sigma_R^2}{\left(1+0.69\sigma_R^{12/5}\right)^{5/6}}\right] - 1}. \tag{5}$$

Through the relationship $\sigma_I^2 = 1/\alpha + 1/\beta + 1/\alpha\beta$ the intensity of the turbulence like weak, moderate, and strong can be decided. To improve the system performance in terms of capacity, we proposed a model using GG turbulence-induced fading. Figure 2 shows the proposed UOWC system with $M$ lasers and $N$ photo-diodes employed at the transmitter and receiver end. Various optical signals collected by the photodiodes and are combined using EGC scheme (Bhatnagar and Ghassemlooy 2016). The advantage of EGC is that it is easy to implement whereas no adaptive amplifiers are required at the receiver side (Jamali et al. 2016).

The information is transmitted through the turbulence channel with the help of a laser in the case of single input multiple output (SIMO) transmission. Due to fading the intensity gets degraded, also due to scattering the multiple photo-diodes collect the information bits which get amplified and collected at the receiver using the EGC technique. In EGC technique the amplifiers at the receiving end have equal gains as $A_1 = A_2 = A_N$. Since intensity variation is collected due to various photodiodes as a result there exist a joint distribution for the signal intensities received for MIMO scheme. In Luong and Pham (2014) the distribution of intensity has been evaluated by using the generalized power series method and the received signal is given as

$$R = \frac{x\psi}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} I_{mn} + Y, \tag{6}$$

where $R$ is the received signal, $\psi$ is the power conversion factor, $Y$ is the additive white gaussian noise with zero mean and variance $N_o/2$. Now using Eq. (6) in Eq. (3) along with $\alpha_1 = MN\alpha$ and $\beta_1 = MN\beta$ the intensity $I$ can be viewed as

$$f_I(I) = \frac{2\left(\frac{\alpha_1\beta_1}{MN}\right)^{(\alpha_1+\beta_1)/2}}{\Gamma(\alpha_1)\Gamma(\beta_1)} I^{(\alpha_1+\beta_1)/2-1} K_{\alpha_1-\beta_1}\left(2\sqrt{\frac{\alpha_1\beta_1}{MN}I}\right) \tag{7}$$

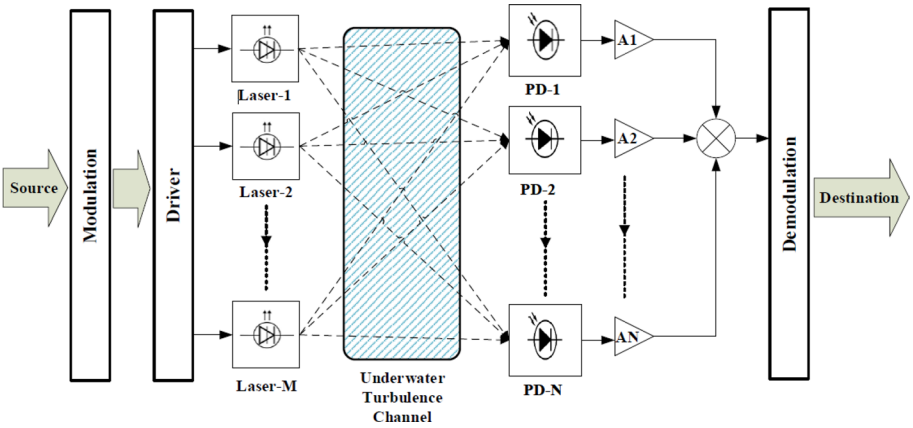



**Fig. 2** MIMO scheme for UOWC system using EGC

Now, we utilize the GG-PDF for the modeling of the fading channel in the UOWC system. The expression for PDF derived as Eq. (7) can help in determining the various performance metrics, moreover, it simplifies the analytical calculation also. We have proposed two detection mechanisms known as OHD and IM/DD with a variety of modulation schemes. In our analysis, the fading characteristics that have dominated the effect of absorption/scattering are negligible and their effects like inter symbol interference are also not significant. First, we derive the general form of GG-PDF and CDF for the UOWC system under different detection schemes. So that it can be utilized to obtain the performance parameters.

It has been shown that IM/DD has low complexity and is cost-effective (Chaaban and Alouini 2015). While OHD has a high capacity as compared to the previous technique. Also, at a low SNR the gap between these two detection techniques is constant in terms of capacity, however it increases at high SNR. On substituting in Eq. (7) and utilizing Eqs. (07.34.03.0605.01) and (07.34.16.0001.01) from Wolfram (2014), the PDF of the UOWC system when operated in OHD can be given as

$$f_\gamma(\gamma) = \frac{1}{\Gamma(\alpha_1)\Gamma(\beta_1)\gamma} G_{0,2}^{2,0}\left[\frac{\alpha_1\beta_1}{MN}\frac{\gamma}{\mu_1}\middle| \begin{matrix} - \\ \alpha_1, \beta_1 \end{matrix}\right], \tag{8}$$

$G_{p,q}^{m,n}[.]$ is the Meiger-G function defined in Wolfram (2014).

In IM/DD mechanism, average electrical SNR $\mu_2$ is obtained using $\gamma = \frac{(\eta I)^2}{MNN_0}$, $\mu_2 = \frac{(\eta E[I])^2}{MNN_0}$ $\overline{\gamma} = \frac{\mu_2 E[I^2]}{E[I]^2}$ as

$$\mu_2 = \frac{\overline{\gamma}\alpha\beta}{1+\alpha+\beta+\alpha\beta}. \tag{19}$$

On substituting in Eq. (7) and utilizing Eqs. (07.34.03.0605.01) and (07.34.16.0001.01) from Wolfram (2014), the PDF of the UOWC system can be expressed as

$$f_\gamma(\gamma) = \frac{1}{2.\Gamma(\alpha_1)\Gamma(\beta_1)\gamma} \times G_{0,2}^{2,0}\left[\frac{\alpha_1\beta_1}{MN}\left(\frac{\gamma}{\mu_2}\right)^{1/2}\middle| \begin{matrix} - \\ \alpha_1, \beta_1 \end{matrix}\right]. \tag{10}$$

Using Eqs. (8) and (10), by utilizing Eqs. (8.4.3/1) and (8.2.2/15) from Wang et al. (2020), we obtained a unified expression for PDF which expressed both OHD and IM/DD schemes as

$$f_\gamma(\gamma) = \frac{1}{r\Gamma(\alpha_1)\Gamma(\beta_1)\gamma} G_{0,2}^{2,0}\left[\frac{\alpha_1\beta_1}{MN}\left(\frac{\gamma}{\mu_r}\right)^{1/r}\middle| \begin{matrix} - \\ \alpha_1, \beta_1 \end{matrix}\right]. \tag{11}$$

The performance of any communication system can be best analyzed using its performance parameters like capacity, BER, and outage probability. In Zedini et al. (2019), obtaining an analytical expression for performance metrics is intractable, and closed-form expressions are not achievable.

EC is an important feature that illustrates the performance of UOWC system. Because instantaneous SNR is a random variable, EC is also a random value, and its average value for UWOC system can be given as Lin et al. (2022).

$$\overline{C} = \int_0^\infty B \log_2(1+\gamma) f_\gamma(\gamma) d\gamma. \tag{12}$$

On substituting Eq. (11) in Eq. (12), we derived a closed-form expression for EC as

$$\begin{aligned}\overline{C} = {} & \frac{B}{4\pi \ln 2} \frac{(\alpha_1\beta_1)^{(\alpha_1+\beta_1)/2}}{\Gamma(\alpha_1)\Gamma(\beta_1)} \left(\frac{1}{\mu_r}\right)^{(\alpha_1+\beta_1)/4} \\ & \times \mathrm{G}_{2,6}^{6,1}\left[\frac{\alpha_1^2\beta_1^2}{16\mu_r}\middle| \begin{matrix} -\frac{\alpha_1+\beta_1}{4}, -\frac{\alpha_1+\beta_1}{4}+1 \\ \frac{\alpha_1-\beta_1}{4}, \frac{\alpha_1-\beta_1+2}{4}, \frac{\beta_1-\alpha_1}{4}, \frac{\beta_1-\alpha_1+2}{4}, -\frac{\alpha_1+\beta_1}{4}, -\frac{\alpha_1+\beta_1}{4} \end{matrix}\right]\end{aligned} \tag{13}$$

It has been seen from Eq. (13) that as SNR increases, channel capacity also increases. In Zedini et al. (2019), a generalized expression has been computed which evaluates the average BER for various modulation formats under both IM/DD and OHD techniques. The general expression of ABER can be expressed as

$$P_e = \frac{\delta}{2\Gamma(p)} \sum_{k=1}^{n} \int_0^\infty \Gamma(p, q_k\gamma) f_\gamma(\gamma) d\gamma, \tag{14}$$

where $\Gamma(p, q_k\gamma)$ is incomplete gamma function and $\delta, p, q_k, n$ are the parameter that describes the specific modulation scheme. Table 2 shows different values of the parameter associated with various modulation schemes and detection mechanism i.e.IM/DD or OHD.

Since most frequently used modulation technique practical UOWC systems is OOK so we investigate the ABER for the IM/DD technique (Bhatnagar and Ghassemlooy 2016). We also examine the ABER for M-PSK and M-QAM, which are commonly used in coherent systems. By substituting Eq. (11) into Eq. (14), utilizing Eq. (07.34.03.0613.01) and Eq. (07.34.21.0013.01) in (Wolfram 2014), an expression for ABER has been calculated for OOK, BPSK, M-PSK, and M-QAM, and it is given by

$$P_e = \frac{\delta}{2\Gamma(p)} \left(\frac{r^{\alpha_1+\beta_1}}{(2\pi)^{r-1}}\right) G_{2,2r+1}^{2r,r}\left[\left(\frac{\alpha_1\beta_1}{MNr^2}\right)^r \frac{1}{\mu_r q_k}\middle| \begin{matrix} 1, (1-p) \\ \Delta(r,\alpha_1), \Delta(r,\beta_1), 0 \end{matrix}\right]. \tag{15}$$

For underwater fading channels OP is a useful parameter that estimates the UOWC system performance. It is defined as the probability of the instantaneous SNR falling below a predefined threshold value and can be calculated as Zedini et al. (2020)

$$P_{out} = P_r(\gamma \le \gamma_{th}) = F_\gamma(\gamma_{th}). \tag{16}$$

**Table 2** Parameters associated with different modulation schemes

| Modulation | $\delta$ | $p$ | $q_k$ | $n$ | Detection Type |
|---|---|---|---|---|---|
| On–off keying (OOK) | 1 | 1/2 | 1/4 | 1 | IM/DD |
| Binary phase shift keying (BPSK) | 1 | 1/2 | 1 | 1 | OHD |
| M-ary phase shift keying (M-PSK) | $\frac{2}{\max(\log_2 M, 2)}$ | 1/2 | $\sin^2\left(\frac{(2k-1)\pi}{M}\right)$ | $\max\left(\frac{M}{4}, 1\right)$ | OHD |
| M-ary quadrature amplitude modulation (M-QAM) | $\frac{4}{\log_2 M}\left(1-\frac{1}{\sqrt{M}}\right)$ | 1/2 | $\frac{3(2k-1)^2}{2(M-1)}$ | $\frac{\sqrt{M}}{2}$ | OHD |

**Table 3** Estimated parameters of the EGG model, along with the parameters of the proposed system for gradient temperature UOWC system

| Temp. Grad. °C $cm^{-1}$ | SI $\sigma_I^2$ | $\sigma_r^2$ | $\alpha$ | $\beta$ |
|---|---|---|---|---|
| 0.22 | 3.1952 | 0.23903 | 0.9815 | 0.9276 |
| 0.22 | 1.9328 | 0.1795 | 1.4388 | 1.3661 |
| 0.1 | 0.4769 | 0.065 | 4.7488 | 4.5442 |
| 0.05 | 0.4201 | 0.0585 | 5.3286 | 5.1012 |
| 0.15 | 0.1915 | 0.0292 | 11.1556 | 10.699 |
| 0.05 | 0.1486 | 0.0231 | 14.2302 | 13.653 |

**Table 4** Estimated parameters of the EGG model, along with the parameters of the proposed system for uniform temperature UOWC system

| Water type | SI $\sigma_I^2$ | $\sigma_r^2$ | $\alpha$ | $\beta$ |
|---|---|---|---|---|
| Salty water | 0.1006 | 0.01596 | 20.8024 | 19.9673 |
| | 0.3111 | 0.04515 | 7.0445 | 6.7495 |
| | 1.1273 | 0.1258 | 2.2346 | 2.1298 |
| Fresh water | 0.1088 | 0.0172 | 19.2796 | 18.5042 |
| | 0.315 | 0.0457 | 6.954 | 6.6626 |
| | 1.0409 | 0.1189 | 2.3898 | 2.2788 |

The cumulative density function (CDF) can be obtained by evaluating $F_\gamma(\gamma) = \int_0^\gamma f_\gamma(\gamma)d\gamma$. On substituting Eq. (10), and utilizing Eq. (07.34.21.0084.01) from Wolfram (2014). The CDF can be obtained as

$$F_I(I) = \frac{(\alpha_1\beta_1)^{(\alpha_1+\beta_1)/2}}{\Gamma(\alpha_1)\Gamma(\beta_1)MN} I^{(\alpha_1+\beta_1)/2} G_{1,3}^{2,1}\left[\alpha_1\beta_1 I \middle| \begin{matrix} 1-\frac{\alpha_1+\beta_1}{2} \\ \frac{\alpha_1-\beta_1}{2}, \frac{\beta_1-\alpha_1}{2}, -\frac{\alpha_1+\beta_1}{2} \end{matrix}\right]. \tag{17}$$

Substituting Eq. (17) into Eq. (16), the expression for OP can be obtained as

$$P_{out} = \left(\frac{\gamma_{th}}{\overline{\gamma}}\right)^{(\alpha_1+\beta_1)/4} \times \frac{(\alpha_1\beta_1)^{(\alpha_1+\beta_1)/2}}{\Gamma(\alpha_1)\Gamma(\beta_1)} \times G_{1,3}^{2,1}\left[\alpha_1\beta_1\sqrt{\frac{\gamma_{th}}{\overline{\gamma}}} \middle| \begin{matrix} 1-\frac{\alpha_1+\beta_1}{2} \\ \frac{\alpha_1-\beta_1}{2}, \frac{\beta_1-\alpha_1}{2}, -\frac{\alpha_1+\beta_1}{2} \end{matrix}\right]. \tag{18}$$

In Zedini et al. (2019), the EGG model is used to predict the performance of the UOWC system both analytically and experimentally. Table 3, shows the effect of the temperature gradient on the strength of turbulence. As temperature gradient increases the SI values increases which signifies strong attenuation.

The effect of water under uniform temperature is also analyzed experimentally (Zedini et al. 2019). Table 4, shows the estimated parameters of the EGG and our proposed model in both type of water.

In the next section, the numerical results along with the comparison of the EGG model (Zedini et al. 2019) with our proposed model are presented.

## 3 Numerical results

We present a detailed discussion on the parameters which are responsible for affecting the UOWC system performance in this section. Analytical results obtained in the previous section are verified using plots and are used to demonstrate the EC, ABER, and OP of the UOWC link using our proposed GG turbulence fading channel. Moreover, various transmitter–receiver diversity MIMO schemes based on the EGC technique have also been incorporated in the proposed model to improve system performance.

From Table 3 it can be verified that higher the level of temperature gradient, higher is the value of SI leading to stronger turbulence resulting in the deterioration of system performance. In Fig. 3, OP is depicted under strong ($\sigma_I^2 = 3.5918$) and moderate ($\sigma_I^2 = 1.9358$) turbulence conditions. But on introducing transmitter–receiver diversity, outage probability decreases, which signifies the effect of fading strength reduces. Also, for low SNR values (below 15 dB) (Hessien et al. 2018), the MIMO scheme has a very insignificant effect on the UOWC system performance. At high SNR values (above 15 dB), outage probability decreases significantly using MIMO schemes, thereby increasing the UOWC system performance.

With the help of Table 4, OP is plotted and a comparison of the EGG turbulence model with our proposed GG model has been presented in Fig. 4 for fresh and salty water under uniform temperature conditions. OP as a function of normalized average SNR is presented for the moderate turbulence ($\sigma_I^2 = 1.12735$) for thermally uniform salty water. Moreover, it can be seen from Fig. 4 that salinity which creates rapid fluctuations in intensity affects the UOWC system performance but to a smaller extent. In this case, where the effect of turbulence decreases, OP keeps on decreasing. The proposed model outperforms the EGG model. Also, as we introduce transmitter–receiver diversity, performance has been improved drastically.

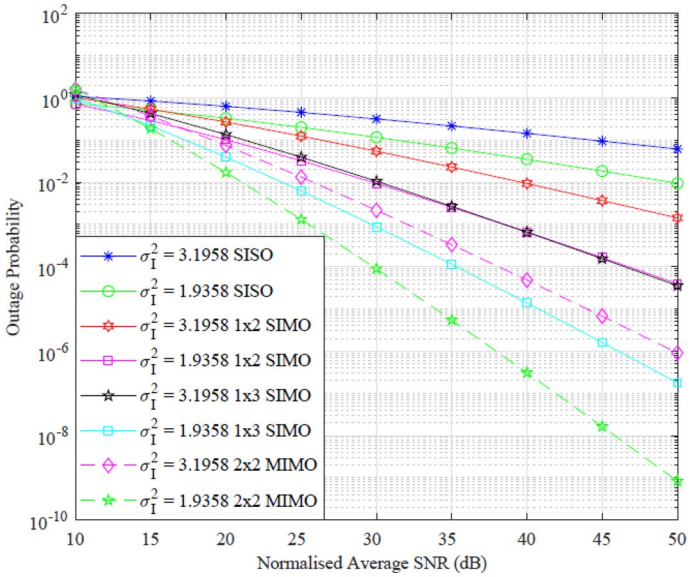



**Fig. 3** Outage probability versus normalized average SNR for moderate and strong turbulence using various MIMO schemes operated under optical heterodyne detection

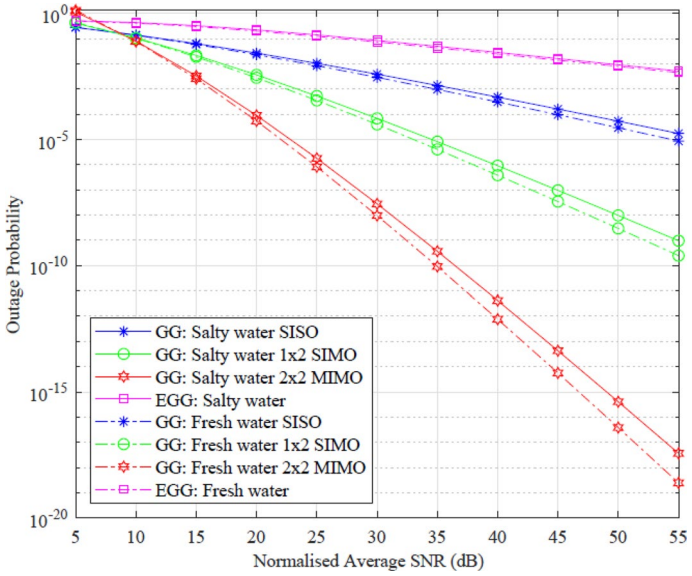



**Fig. 4** Outage probability under salty and fresh water using MIMO schemes for thermally uniform UOWC channel

Because direct detection only reaches the shot-noise limit for relatively high signal powers, whereas OHD can reach the standard quantum limit for much lower signal powers, so OHD offers advantage in terms of sensitivity. As a result, this detection method is particularly well suited to very weak signals. ABER as a function of average SNR for both IM/DD and OHD techniques is shown in Fig. 5. Under strong fading ($\sigma_I^2 = 3.1952$), it has been shown that OHD outperforms IM/DD technique. At $P_e = 7.3 \times 10^{-6}$ utilizing 2×2 MIMO almost 10 dB SNR has been achieved in comparison with 1×2 SIMO. However, below 15 dB 1×3 SIMO scheme performs better than 2×2 MIMO scheme. Also, it has been observed that the UOWC system based on the SISO scheme using OHD performs nearly the same as 1×2 SIMO IM/DD technique.

The temperature gradient 0.22 °C $cm^{-1}$ has been selected from Table 3, resulting in strong turbulence of SI ($\sigma_I^2 = 3.1952$). A comparison of the EGG turbulence model with our proposed GG model is shown in Fig. 6. Analytical results obtained in Eq. (15) have been verified for the UOWC system operated under a strong turbulence region with OHD using various modulation techniques. BPSK modulation outperforms 16-QAM,16-PSK, and 64-QAM modulation. By utilizing 1×2 SIMO scheme, system performance improves over the conventional SISO scheme. At $P_e = 12.9 \times 10^{-4}$, utilizing 1×2 SIMO almost 15 dB has been achieved. It has been shown that the proposed SISO system performs similarly to the EGG model, but at high SNR, the EGG model performs better. The proposed system using 1×2 SIMO scheme achieved better system performance than the EGG model.

Because power conservation is an important aspect in UOWC, this scheme does not require a carrier signal at the receiver during demodulation, resulting in reduction in complex receivers as compared to QAM. Moreover, the BW requirements are less as compared to BPSK and QAM schemes. Therefore, this scheme is preferred for wireless communication in underwater medium. Figure 7 compares the ABER performance of

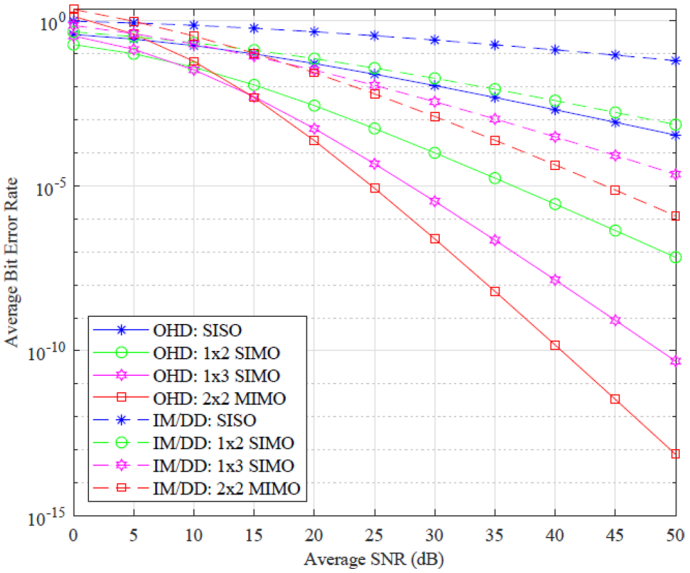



**Fig. 5** Average BER versus average SNR operated under either IM/DD or OHD with various MIMO schemes under strong fading ($\sigma_I^2 = 3.1952$) condition

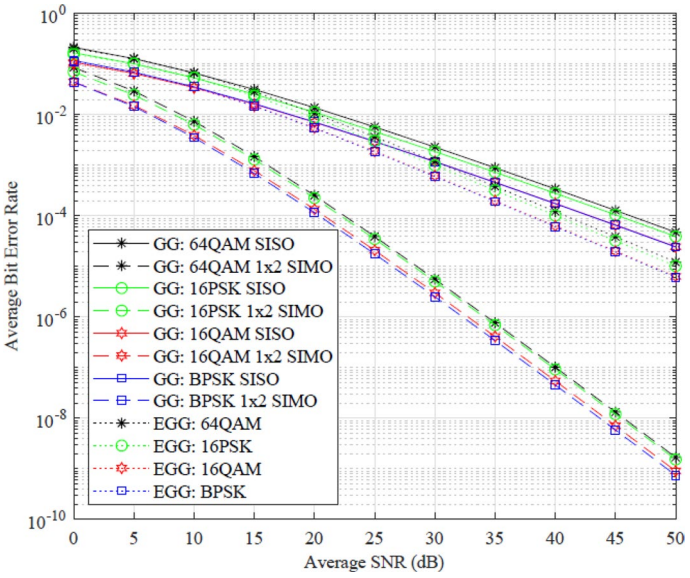



**Fig. 6** Comparison of the EGG and proposed GG model in terms of average BER for various modulation schemes at a temperature gradient 0.22 °C cm$^{-1}$ operated under IM/DD technique

various modulation schemes under strong underwater turbulence ($\sigma_I^2 = 3.1952$). Temperature gradient 0.22 °C cm$^{-1}$ has been selected from Table 3, resulting in strong turbulence of SI ($\sigma_I^2 = 3.1952$). Figure 8 shows EC as a function of average SNR. As average SNR

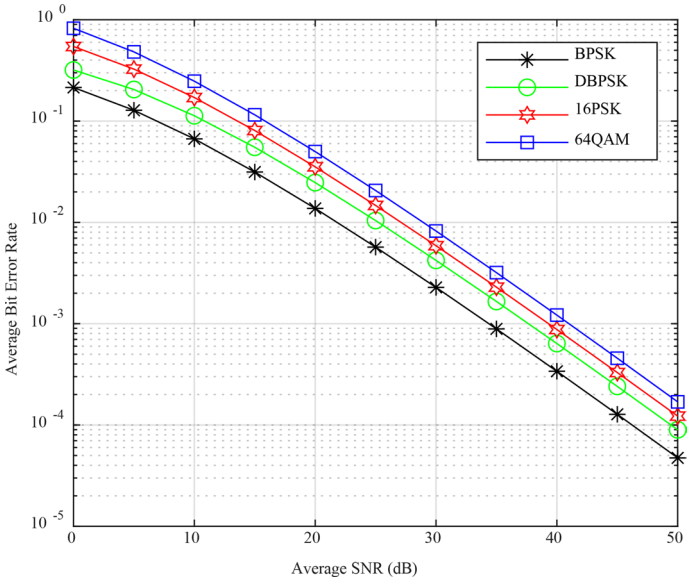



**Fig. 7** Comparison of various modulation schemes for the proposed GG model in terms of average BER under strong turbulence ($\sigma_I^2 = 3.5918$) operated under IM/DD technique

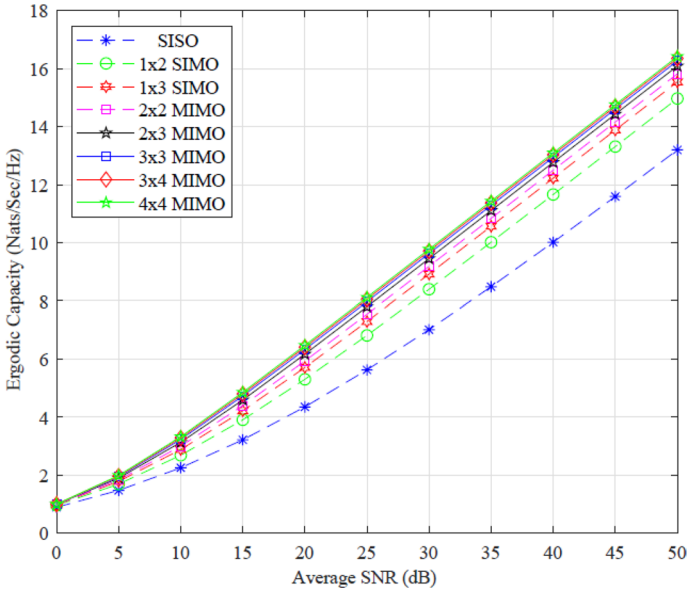



**Fig. 8** Ergodic capacity versus average SNR for temperature gradient of 0.22 °C cm$^{-1}$ using various MIMO schemes

increases, EC increases which is consistent with the previous work. At low SNR values, it is clear that there is a very small change in the Ergodic channel capacity on introducing the MIMO scheme. But at high SNR above 25 dB, the performance drastically improves. As

we see from Fig. 8, on introducing receiver diversity SIMO, the performance of the UOWC system significantly improves in comparison with SISO scheme.

For various temperature gradients as shown in Table 3, Fig. 9 shows the variation of EC as a function of average SNR. It has been shown that capacity decreases by increasing the turbulence-induced fading in the UOWC system. For strong fading ($\sigma_I^2 = 3.1952$), if $1 \times 3$ and above MIMO scheme is employed, then the EC improves and the UOWC system performs as it is working in the weak turbulence regime. Moreover, it can be seen that EC improves more in case of weak and moderate turbulence regime as compared to strong turbulence on employing receiver's diversity. This signifies that if increasing temperature gradient above 0.22 °C $cm^{-1}$, system performance will not improve much on further employing higher MIMO schemes.

Figure 10 shows the performance comparison of EGG with our proposed GG model in terms of EC for various levels of turbulence-induced fading. Figure validates the significant effect of the temperature gradient on the UOWC system performance. However, it can be seen that when the temperature gradient decreases, ABER decreases resulting in the performance improvement of the UOWC system. Also, from Fig. 10 it can be seen that proposed model outperforms the EGG model present in Zedini et al. (2019) for turbulence-induced fading.

## 4 Conclusion

In this paper, we present a UOWC model for the turbulence-induced fading channel using spatial diversity MIMO in which the intensity fluctuations of an optical beam due to degradation effects like fading are characterized by the GG model. Moreover, MIMO schemes further improve the UOWC system performance and we derived the closed-form

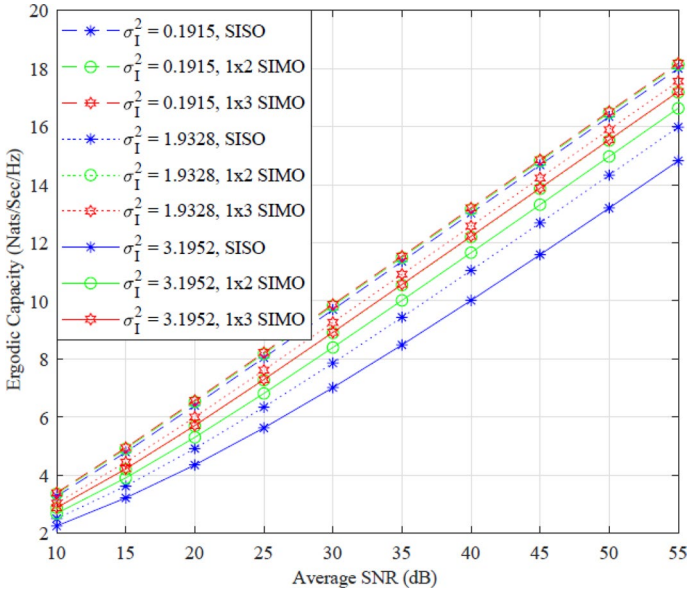



**Fig. 9** Ergodic capacity versus average SNR under various turbulence induced fading conditions

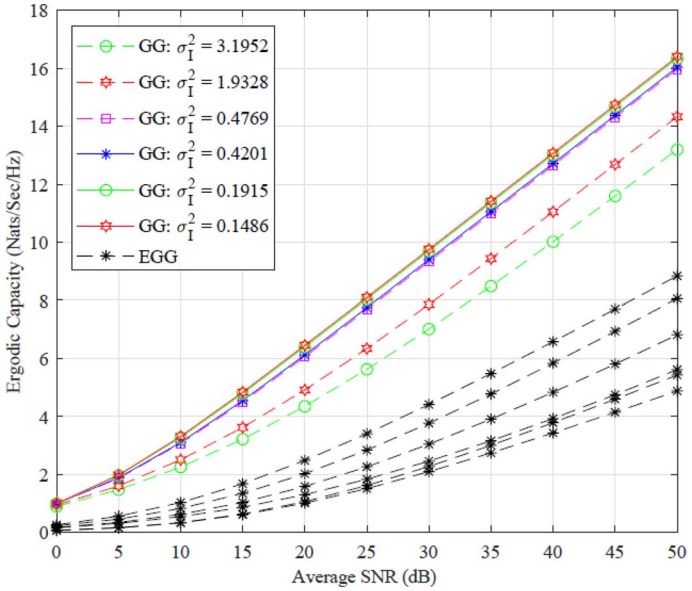



**Fig. 10** A comparison of EGG with proposed GG model in terms of Ergodic capacity for IM/DD technique under different levels temperature gradient

expressions for system performance using the Meijer-G function. Also, we analytically obtained expressions for the ABER for various modulation techniques operated under either IM/DD or OHD. However, the OP and the EC under diverse turbulence conditions for both saltwater and freshwater are also evaluated. So, our proposed model along with the results presented in this paper can help in evaluate the UOWC system in various turbulence conditions.

**Acknowledgements** The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

## References

Ata, Y., Baykal, Y.: Scintillations of optical plane and spherical waves in underwater turbulence. JOSA A **31**(7), 1552–1556 (2014)

Bhatnagar, M.R., Ghassemlooy, Z.: Performance analysis of gamma–gamma fading FSO MIMO links with pointing errors. J. Lightwave Technol. **34**(9), 2158–2169 (2016)

Boucouvalas, A.C., Peppas, K.P., Yiannopoulos, K., Ghassemlooy, Z.: Underwater optical wireless communications with optical amplification and spatial diversity. IEEE Photon. Technol. Lett. **28**(22), 2613–2616 (2016)

Chaaban, A., Alouini, M.S.: Optical intensity modulation direct detection versus heterodyne detection: a high-SNR capacity comparison. In: 2015 5th International Conference on Communications and Networking (COMNET), pp. 1–5. IEEE (2015) https://doi.org/10.1109/COMNET.2015.7566632

Cochenour, B.M., Mullen, L.J., Laux, A.E.: Characterization of the beam-spread function for underwater wireless optical communications links. IEEE J. Ocean. Eng. **33**(4), 513–521 (2008)

Ergul, O., Dinc, E., Akan, O.B.: Communicate to illuminate: state-of-the-art and research challenges for visible light communications. Phys. Commun. **17**, 72–85 (2015)

Fu, Y., Du, Y.: Performance of heterodyne differential phase-shift-keying underwater wireless optical communication systems in gamma–gamma-distributed turbulence. Appl. Opt. **57**(9), 2057–2063 (2018)

Gerçekcioğlu, H.: Bit error rate of focused Gaussian beams in weak oceanic turbulence. JOSA A **31**(9), 1963–1968 (2014)
Hanjiang, L., Jinglong, W., Fanfeng, B., Rukhsana, R., Kaishun, W., Zhongwen, G.: Recent progress of air/water cross-boundary communications for underwater sensor networks: a review. IEEE Sens. J. **22**(9), 8360–8382 (2022)
Hanson, F., Lasher, M.: Effects of underwater turbulence on laser beam propagation and coupling into single-mode optical fiber. Appl. Opt. **49**(16), 3224–3230 (2010)
Hessien, S., Tokgöz, S.C., Anous, N., Boyacı, A., Abdallah, M., Qaraqe, K.A.: Experimental evaluation of OFDM-based underwater visible light communication system. IEEE Photon. J. **10**(5), 1–13 (2018)
Jamali, M.V., Nabavi, P., Salehi, J.A.: MIMO underwater visible light communications: comprehensive channel study, performance analysis, and multiple-symbol detection. IEEE Trans. Veh. Technol. **67**(9), 8223–8237 (2018)
Jamali, M.V., Salehi, J.A., Akhoundi, F.: Performance studies of underwater wireless optical communication systems with spatial diversity: MIMO scheme. IEEE Trans. Commun. **65**(3), 1176–1192 (2016)
Kaushal, H., Kaddoum, G.: Underwater optical wireless communication. IEEE Access **4**, 1518–1547 (2016)
Li, J., Ma, Y., Zhou, Q., Zhou, B., Wang, H.: Monte Carlo study on pulse response of underwater optical channel. Opt. Eng. **51**(6), 066001 (2012)
Lin, Z., Xu, G., Zhang, Q., Song, Z.: Average symbol error probability and channel capacity of the underwater wireless optical communication systems over oceanic turbulence with pointing error impairments. Opt. Express **30**(9), 15327–15343 (2022)
Liu, X., Yi, S., Zhou, X., Fang, Z., Qiu, Z.J., Hu, L., Tian, P.: 34.5 m underwater optical wireless communication with 2.70 Gbps data rate based on a green laser diode with NRZ-OOK modulation. Opt. Express **25**(22), 27937–27947 (2017)
Luong, D.A., Pham, A.T.: Average capacity of MIMO free-space optical gamma-gamma fading channel. In 2014 IEEE International Conference on Communications (ICC), pp. 3354–3358. IEEE (2014)
Matta, G., Pandey, P., Agrawal, M., Bahl, R.: Capacity analysis of underwater visible light communication systems over a lossy channel in the presence of noises. JOSA A **39**(5), 948–958 (2022)
Oubei, H.M., Duran, J.R., Janjua, B., Wang, H.Y., Tsai, C.T., Chi, Y.C., Ooi, B.S.: 4.8 Gbit/s 16-QAM-OFDM transmission based on compact 450-nm laser for underwater wireless optical communication. Opt. Express **23**(18), 23302–23309 (2015)
Peng, J., Zhang, L., Zhang, K., Ma, J.: Channel capacity of OAM based FSO communication systems with partially coherent Bessel–Gaussian beams in anisotropic turbulence. Opt. Commun. **418**, 32–36 (2018)
Peppas, K.P., Boucouvalas, A.C., Ghassemloy, Z.: Performance of underwater optical wireless communication with multi-pulse pulse-position modulation receivers and spatial diversity. IET Optoelectron. **11**(5), 180–185 (2017)
Sozer, E.M., Stojanovic, M., Proakis, J.G.: Underwater acoustic networks. IEEE J. Ocean. Eng. **25**(1), 72–83 (2000)
Wang, W., Wang, P., Cao, T., Tian, H., Zhang, Y., Guo, L.: Performance investigation of underwater wireless optical communication system using M-ary OAMSK modulation over oceanic turbulence. IEEE Photon. J. **9**(5), 1–15 (2017)
Wang, W., Wang, P., Guo, L., Pang, W., Chen, W., Li, A., Han, M.: Performance investigation of OAMSK modulated wireless optical system over turbulent ocean using convolutional neural networks. J. Lightwave Technol. **38**(7), 1753–1765 (2020)
Wolfram. The Wolfram Functions Site, 2014. [Online]. http://functions.wolfram.com
Zedini, E., Kammoun, A., Soury, H., Hamdi, M., Alouini, M.S.: Performance analysis of dual-hop underwater wireless optical communication systems over mixture exponential-generalized gamma turbulence channels. IEEE Trans. Commun. **68**(9), 5718–5731 (2020)
Zedini, E., Oubei, H.M., Kammoun, A., Hamdi, M., Ooi, B.S., Alouini, M.S.: Unified statistical channel model for turbulence-induced fading in underwater wireless optical communication systems. IEEE Trans. Commun. **67**(4), 2893–2907 (2019)
Zhang, L., Wang, Z., Wei, Z., Chen, C., Wei, G., Fu, H.Y., Dong, Y.: Towards a 20 Gbps multi-user bubble turbulent NOMA UOWC system with green and blue polarization multiplexing. Opt. Express **28**(21), 31796–31807 (2020)

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

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/361689973

# Determinants of Spiritual Tourism Consumption: A Hierarchical Approach

**Article** · July 2022

DOI: 10.21427/t97s-s131

CITATIONS
0

READS
27

**3 authors:**

Suneel Kumar
University of Delhi
**28** PUBLICATIONS **122** CITATIONS
SEE PROFILE

Shivangi Gupta
Delhi Technological University
**3** PUBLICATIONS **1** CITATION
SEE PROFILE

Shekhar Asthana
University of Delhi
**17** PUBLICATIONS **105** CITATIONS
SEE PROFILE

All content following this page was uploaded by Suneel Kumar on 14 July 2022.

The user has requested enhancement of the downloaded file.

International Journal of Religious Tourism and Pilgrimage

Volume 10 | Issue 2 Article 9

2022

# Determinants of Spiritual Tourism Consumption: A Hierarchical Approach

Suneel Kumar
*Shaheed Bhagat Singh College, University of Delhi*, drsuneel.sbsc@gmail.com

Shivangi Gupta
*University School of Management and Entrepreneurship, Delhi Technological University*, shivangigupta_phdusme2K19@dtu.ac.in

Shekhar
*Faculty of ManagementStudies, University of Delhi*, shekhar0481.sa@gmail.com

Follow this and additional works at: https://arrow.tudublin.ie/ijrtp

Part of the Tourism and Travel Commons

Recommended Citation
Kumar, Suneel; Gupta, Shivangi; and Shekhar (2022) "Determinants of Spiritual Tourism Consumption: A Hierarchical Approach," *International Journal of Religious Tourism and Pilgrimage*: Vol. 10: Iss. 2, Article 9.
doi:https://doi.org/10.21427/t97s-s131
Available at: https://arrow.tudublin.ie/ijrtp/vol10/iss2/9

Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

© International Journal of Religious Tourism and Pilgrimage
ISSN : 2009-7379
Available at: http://arrow.tudublin.ie/ijrtp/ Volume 10(ii) 2022

# Determinants of Spiritual Tourism Consumption: A Hierarchical Approach

**Suneel Kumar**
Shaheed Bhagat Singh College, University of Delhi
drsuneel.sbsc@gmail.com
**Shivangi Gupta**
University School of Management and Entrepreneurship, Delhi Technological University
shivangigupta_phdusme2K19@dtu.ac.in
**Shekhar**
Faculty of Management Studies, University of Delhi
shekhar0481.sa@gmail.com

This paper aims to analyse the relationship between factors influencing the travel intentions of the spiritual tourist. The research design comprises three segments, i.e., identifying factors from the literature, conducting interviews with the academic experts and managers, and analysing the responses recorded. The ***Interpretive Structural Modelling*** (ISM) technique is used to determine the interlinkage between the factors and develop a hierarchical relationship. This paper identified twelve factors that influence the travel intention of spiritual tourists. The results indicate that the pro-tourism attitude of the management of religious monuments, presence of relaxation and recreational activities, historical and cultural heritage of a destination, infrastructure development, and accessibility of place significantly drive tourist motivation to undertake spiritual tourism. Also, destination cost, reference group influence, marketing of the destination, destination image, and stress and spirituality level of tourist influence intention to undertake spiritual tourism, but get influenced by tourist motivation drivers. The study's value lies in clarifying the relationship between factors influencing the travel intentions of spiritual tourists, an area where limited research has been done.

**Key Words**: spiritual tourism; motivation; hospitality management; Interpretive structural modelling

## Introduction

Spirituality is defined as an aspect of human existence wherein one tries to find out the meaning and purpose of life. It aims to give direction to one's journey to seek answers to questions that guide one's living (Baker, 2003). It is proven to have a significant positive impact on a person's health and well-being, thereby improving the quality of life (Adegbola, 2006; Counted, Possamai & Meade, 2018; Jafari *et al.*, 2010; Koenig, 2012; Vitorino *et al.*, 2018). Tourism, where the primary purpose is leisure, and mental peace, has imbibed the principle of spirituality in its product, spiritual tourism. Spiritual Tourism, therefore, can be defined as tourism undertaken to know oneself, seeking answers to the purpose of existence, for mental and emotional well-being, and connecting with the supreme divine (Bone, 2013; Buzinde, 2020; Ivona & Privitera, 2019; Sharpley, 2016; Singh, 2009). Several studies have been conducted on various aspects of spiritual tourism, highlighting it as a sub-section but different product from religious tourism, its relative importance for the economy and visitors, and the theoretical framework for its development. Studies have pointed out the difference between religious and spiritual tourism. A religious tourist's travel intention is to act on their religious beliefs, whereas a spiritual tourist travels to explore life's meaning (Sirirat, 2019).

Researchers globally have hovered around the factors that can help develop spiritual tourism. Each of the studies is unique in identifying different factors influencing the travel intention of spiritual tourists. Insight into these factors helps in strategic policy-making by explaining the 'What' and 'How' of spiritual tourism. However, a study on the interrelationship between these factors is missing and demands the researcher's attention. In many developing and underdeveloped nations, good investment proposals suffer from crowding out due to

scarcity of funds. Thus, there is a significant chance that spiritual tourism, although a potentially lucrative niche tourism segment, might fail to gain policymakers' attention compared to other mass tourism segments.

Identifying and working on every possible factor will become ineffective as it would heavily consume scarce resources. Therefore, it is of utmost importance to strategically develop the industry by focusing on factors with the greatest potential. Thus, this study attempts to bridge the literature gap and help policy makers formulate spiritual tourism strategies by highlighting the most important factors influencing spiritual tourist travel intentions. This will help allocate resources strategically and act as a base for future research in this direction.

The study objectives can be summarised in the form of the following research questions.

***RQ1****:* What factors influence the motivation of spiritual tourism consumption?

***RQ2****:* How are these motivators interrelated?

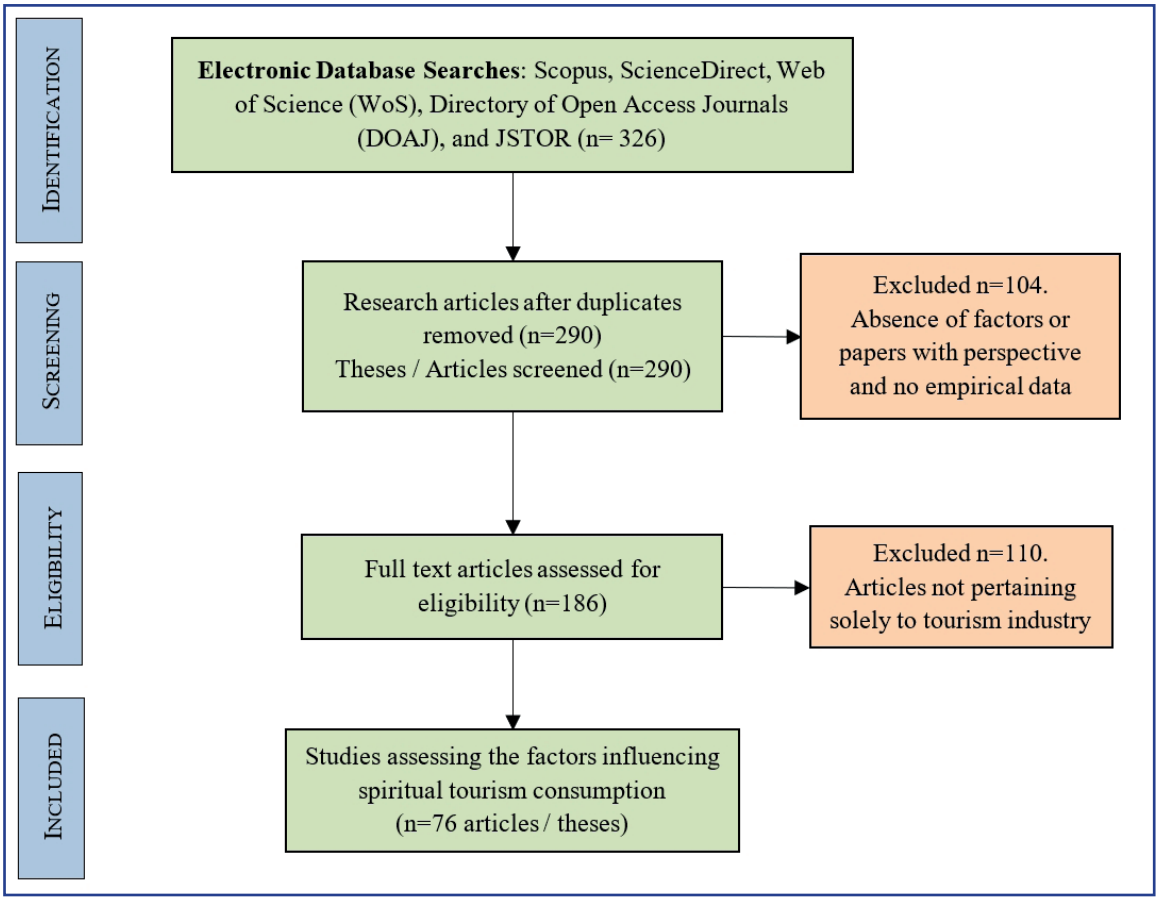
## Review of Literature

The authors conducted an extensive literature review to identify the factors influencing the spiritual tourist intention to undertake spiritual tourism (Figure 1). 326 articles on different databases (Scopus, Web of Science, ScienceDirect, DOAJ, and JSTOR), were identified and screened to identify factors. At each stage, articles / theses were removed if they had no empirical data or were perspective articles. Articles were also rejected from the process if they did not focus on the tourism industry. A final number of 76 articles / theses were explored to identify the different factors influencing spiritual tourists' travel intention.

### *Factors impacting travel intention of spiritual tourists*

Spiritual tourism has been a hot topic of academic interest. The role of marketing this sector will be to attract spiritual tourists by building a compelling brand and image for spiritual tourism (Beerli & Martín, 2004; Geary, 2008; Phukan, Rahman & Devdutt, 2012). Proper

**Figure 1 Review of Literature - Process Followed in the Identification of Factors**




Source- Author's own

marketing will also ensure tourists are aware of spiritual tourism, assess how much they are interested in it, how well they desire spiritual tourism, and how much they are willing to undertake it (Haq & Jackson, 2009). Spiritual events, activities, and journeys act as catalysts in spiritual tourism development by increasing spiritual tourism acceptance amongst businesses and academics (Haq & Wong, 2010). Heritage buildings have proven to be a spiritual transmission sources for tourists (Willson & McIntosh, 2013). Spirituality is felt more in places with a history of rich heritage and culture (Kumar, 2018; Piewdang, Mekkamol & Untachai, 2013). People are influenced by cultural values and try to seek answers to their problems and attain enlightenment. This often links to a mechanism which blends with the community's rich cultural heritage (Power, 2015).

The cost incurred in travelling is one of the critical factors impacting tourist behaviour (Nicolau & Más, 2006; Seyidov & Adomaitienė, 2017). Destination costs will impact tourist decisions in relation to factors such as activities performed, length of stay and expenditure incurred during the trip. This will motivate or demotivate a spiritual tourist to undertake a journey. Public potency, the religious importance of a place, religious administration, art and culture, the attraction of potential opportunities, and community cooperation are among the primary attributes for developing spiritual tourism (Kuralbayev *et al.*, 2017). Individual factors such as wellness, adventure or recreation, and religious motivators such as religious observance, ritualised practice, and reaffirmation of identity are the primary influencers for spiritual tourism (Cheer, Belhassen & Kujawa, 2017). Spirituality plays a role in altering the consumption pattern of a consumer (Skousgaard, 2006). A person with a low spiritual level might not feel the need to undertake spiritual tourism in the same way as a spiritually awakened person.

Spirituality has often been linked to positive well-being (Lifshitz, Nimrod & Bachner, 2019). Relaxation and recreational activities at destinations will tend to have positive impact in reducing the stress level of tourists (Dowson, 2016; Moal-Ulvoas, 2016). Religions are a prominent means of attaining spirituality (Fry, 2000; Noronha, 2015). Many people visit religious monuments to seek answers to their questions and change their lifestyle. If the administration is reluctant to allow people to stay there for long or has certain other restrictions, spiritual tourism will not prosper. Destination accessibility and tourist travel motivation are interrelated (Kumar, Shekhar & Guleria, 2019). An accessible destination will attract a tourist that wishes to enjoy their time traveling and seeking solitude away from mental turmoil. The absence of infrastructure, however, might discourage spiritual tourists from visiting a destination (Hartini & Dewi, 2019).

People tend to follow their reference group consciously or subconsciously (Fernandes & Londhe, 2015). A person might have a spiritual interest on their own, or because reference groups influence tourist behaviour. Families with elders who have high spiritual intellect tend to undertake such trips with their families - including small children who, at their age, might not be spiritually awakened. Stress is a modern-day concept and even young children are impacted by stress (Kleber, 2019; Lodha, 2018). Spirituality is one of the ways to get relief from this stress. People with high-stress levels tend to seek solace from their problems by undertaking spiritual tourism. Intention to undertake spiritual experience can be a crucial motivator for a tourist. This creates demand. A product has to be demanded by the tourist to become successful (Cooper, 1994) and in the absence of demand, there will be no market and hence no revenue.

## Research Methods

### *Conceptual Framework*

The extant literature suggests that spiritual tourism consumption is a function of several factors. However, there exists interrelatedness between these factors. This interrelatedness gives rise to hierarchical relationships between these factors. For instance, relaxation and recreational activities will increase destination cost and directly influence the marketing mix of the destination. It would also impact the intention of a tourist to undertake spiritual tourism. However, to develop these activities, there must be adequate infrastructure that must be established, mindful of the cultural and historical heritage of the destination. Similar interrelationships exist between other factors. The development of strategies requires time and money that in short supply. Thus, it is not advisable to target all of the factors associated with the Spiritual Tourism system. Thus, there is a need for a flexible hierarchical system within which, policy makers can adjust a number of factors and evaluate how these have impacted on the system's influences.

### *Methodology*

Interpretive Structural Modelling (ISM) finds its origin in John Warfield's work in the early 1970s. Since then, researchers have widely used it in strategic management (Alzebdeh, Bashir & Al Siyabi, 2015; Guo, Li & Stevens, 2012; Tripathy, Sahu & Ray, 2013). Due to its simplicity and effectiveness, ISM's use in the tourism industry increases (Kumar & Shekhar, 2020). ISM is an interpretive modelling technique based on an experts' interpretation and gives final output in the form of a model. Developing a hierarchical model using ISM involves a series of steps (Gupta, Gupta & Shekhar, 2021). The first step involves identifying factors by an extensive review of the literature. If the need arises, experts are consulted for additional factors. The factors are scored on the nature of their relationships, and insignificant factors are removed. This gives us the final list of factors to be included in the study. Once the factors have been identified and coded, the next step is to prepare a self-structural interpretation matrix (SSIM). This matrix is also known as the V-A-X-O matrix, where the relationship between variables is coded in the following manner:

- If a factor leads or drives another factor; the relationship is denoted as **V**;
- If a factor is led or driven by another factor; the relationship is denoted as **A**;
- If the two factors are interdependent; the relationship is denoted as **X**;
- If the two factors are unrelated; the relationship is denoted as **O**;

The SSIM is then converted into a binary form, with V and X relationships converted into one, and the rest coded as zero. The resultant matrix obtained from this is known as the Initial Reachability Matrix. From the initial reachability matrix, the next step is to incorporate transitive links into the matrix. The driving power and dependence level help to calculate the hierarchical position of the factors. The driving power and dependence level calculation helps classify factors into autonomous factors, linkage factors, driving factors, and dependent factors. It is also known as MICMAC (Matrix of Cross-Impact Multiplications Applied to Classification) analysis. Categorising the factors helps in identifying the controllable and uncontrollable factors in a system. In the end, the connections are established; firstly the process includes all the links, and then, transitive links are removed in the final di-graph.

### *Survey Instruments and Measurement*

The validity of the identified factors is examined through conducting structured interviews with experts. In this research, the authors used a bilingual questionnaire (in Hindi and English) to gauge the authenticity of the factors and the nature of interrelationship among them. There were additional questions on respondents' demographics. The experts were asked if the various factors act as motivators for spiritual tourists and were asked to rate each on a five-point Likert scale, where 1 means not a motivator and 5 means highly significant motivator. Respondents were then asked to fill a Self-Structuring Interaction Matrix (SSIM). The author obtained the respondents' consent before participating in the research and ensured that their responses were kept confidential and used solely for academic purposes. Two scholars proofread the survey that resulted in a few grammatical edits.

### *Data Collection*

The data were collected in three months from January 2020 to March 2020. The authors collected data solely from Indian respondents. The authors contacted the experts by phone and briefed them about collecting data. These experts are people that have expertise in the field of religious and spiritual tourism or are working in organizations in the spiritual tourism industry. Initially, thirty experts were identified through convenience and purposive sampling techniques. Snowball sampling technique was used to identify subsequent experts. For ease of recording and analysis, the questionnaire was mailed to a hundred and forty respondents, of which one hundred and twenty questionnaires were returned. Twenty of the returned surveys were rejected as they were incomplete, and thus, analysis was done on the one hundred usable responses.

## Results

### *Sample Characteristics*

The sample was dominated by academic experts (65 of 100 respondents), meaning only 35 percent of the respondents were industry experts. The sample was slightly dominated by males (58 percent). Of the total sample, 41 were between 35 and 40 years of age, 25 were between 30 and 35 years, 19 percent indicated that they were between 25 and 30, and 15 were above 40 years of age.

**Table 1: Average Scores of Factors Determined Through Expert Interaction**

| Factor Code | Factor | Average Score |
|---|---|---|
| 01 | Destination cost | 3.58 |
| 02 | Spirituality level of tourist | 3.88 |
| 03 | Marketing Mix of Destination | 3.64 |
| 04 | Historical and Cultural Heritage | 4.24 |
| 05 | Relaxation and Recreational facilities | 3.98 |
| 06 | Management of religious monuments | 4.48 |
| 07 | Accessibility of Destination | 4.12 |
| 08 | Infrastructure | 4.16 |
| 09 | Reference Group Influence on tourist | 2.69 |
| 10 | Destination Image | 2.85 |
| 11 | Stress level of Tourists | 4.01 |
| 12 | Intention to undertake spiritual tourism | 4.84 |

Source: Author's calculation

About their highest level of education, most of the respondents (71) had a PhD, while 19 had a professional degree, and 7 had an MPhil. However, only 3 had a graduation degree. Concerning their spiritual tourism consumption, 45 had been on a spiritual trip within six months, 35 within one year, and the remaining 20 had been on a spiritual trip some time during their life.

### *Examining the interrelationship between the factors*

Experts' scores on each factor were averaged to calculate the score for the factor out of five (Table 1). Intention to undertake spiritual tourism, management of religious monuments, and historical and cultural heritage of the destination were the factors that scored the highest with 4.84, 4.48, and 4.24 respectively. Reference group influence on tourists scored the lowest with a value of 2.69. All of the factors managed to score above 2.5 (a cut-off score pre-determined by the authors in the study), so all factors were included in the study.

The SSIM (Table 2) shows the relationships identified by the experts between the factors.

It shows that ***destination cost*** influences the marketing mix of the destination and the destination image.

The ***spirituality level of the tourist*** influences their desire to undertake spiritual tourism.

The ***marketing mix*** of the destination influences the destination image.

The destination's ***historical and cultural heritage*** influences the destination cost, marketing mix of destination, relaxation and recreational activities, infrastructure development, and destination image.

The presence of ***relaxation and recreational activities*** influences the destination cost, destination marketing, the stress level of the tourists, and destination image.

The ***attitude of the management*** of a religious monument influences the historical and cultural heritage of the destination, development of relaxation and recreational activities, accessibility of the destination, infrastructure development, and destination image.

The ***accessibility*** of the destination influences the destination cost, marketing mix of the destination,

**Table 2: Self-Structural Interaction Matrix (SSIM)**

| | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **01** | - | O | V | A | A | O | A | A | O | V | O | O |
| **02** | O | - | O | O | O | O | O | O | A | O | A | V |
| **03** | A | O | - | A | A | O | A | A | O | V | O | O |
| **04** | V | O | V | - | V | A | O | X | O | V | O | O |
| **05** | V | O | V | A | - | A | A | A | O | V | V | O |
| **06** | O | O | O | V | V | - | V | V | O | V | O | O |
| **07** | V | O | V | O | V | A | - | A | O | V | O | O |
| **08** | V | O | V | X | V | A | V | - | O | V | O | O |
| **09** | O | V | O | O | O | O | O | O | - | O | O | V |
| **10** | A | O | A | A | A | A | A | A | O | - | O | V |
| **11** | O | O | O | O | A | O | O | O | O | O | - | V |
| **12** | O | A | O | O | O | O | O | O | A | A | A | - |

Source: Author's compilation

**Table 3: Initial Reachability Matrix**

| | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **01** | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| **02** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **03** | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| **04** | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| **05** | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| **06** | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 |
| **07** | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| **08** | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |
| **09** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| **10** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| **11** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| **12** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

Source: Author's compilation

development of relaxation and recreational activities, and destination image.

The ***development of infrastructure*** influences the destination cost, marketing mix of the destination, historical and cultural heritage, development of relaxation and recreational activities, accessibility of the destination, and destination image.

The ***reference group*** pressure influences tourists' intention to undertake spiritual tourism and spirituality level.

The ***destination image*** influences tourists' intention to undertake spiritual tourism.

The ***stress level*** of the tourist influences the decision to undertake spiritual tourism.

The SSIM is converted to its binary form to produce an Initial Reachability Matrix (Table 3).

Transitivity helps to in identify new indirect relationships (Table 4).

The ***marketing mix*** of the destination, by transitivity, drives a tourist's intention to undertake spiritual tourism.

The destination's ***historical and cultural heritage*** also influences the accessibility of the destination and stress level of the tourist, and intention to undertake spiritual tourism.

The presence of ***relaxation and recreational*** activities now shows an influence on the intention to undertake spiritual tourism.

**Table 4: Final Reachability Matrix**

| | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | Driving power |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **01** | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | **3** |
| **02** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | **2** |
| **03** | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1* | **3** |
| **04** | 1 | 0 | 1 | 1 | 1 | 0 | 1* | 1 | 0 | 1 | 1* | 1* | **9** |
| **05** | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1* | **6** |
| **06** | 1* | 0 | 1* | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1* | 1* | **10** |
| **07** | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1* | 1* | **7** |
| **08** | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1* | 1* | **9** |
| **09** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | **3** |
| **10** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | **2** |
| **11** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | **2** |
| **12** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | **1** |
| **Dependence Level** | **6** | **2** | **7** | **3** | **5** | **1** | **4** | **3** | **1** | **8** | **6** | **11** | |

Source: Author's compilation

**Table 5 Partition Level of Factors**

| Code | Factor Name | Partition Level |
|---|---|---|
| 01 | Destination cost | Level 5 |
| 02 | Spirituality level of tourist | Level 7 |
| 03 | Marketing Mix of Destination | Level 6 |
| 04 | Historical and Cultural Heritage | Level 2 |
| 05 | Relaxation and Recreational facilities | Level 4 |
| 06 | Management of religious monuments | Level 8 |
| 07 | Accessibility of Destination | Level 3 |
| 08 | Infrastructure | Level 2 |
| 09 | Reference Group Influence on tourist | Level 6 |
| 10 | Destination Image | Level 7 |
| 11 | Stress level of Tourists | Level 7 |
| 12 | Intention to undertake spiritual tourism | Level 1 |

Source: Author's calculation

The ***attitude of management*** at a religious monuments now additionally drives destination cost, marketing mix of the destination, the stress level of the tourists, and intention to undertake spiritual tourism.

The ***accessibility*** of the destination and infrastructure development now influences the tourists' stress level and intention to undertake spiritual tourism.

The 'driving power' and the 'dependence level' is calculated for each of the factors.

With the help of the driving power and dependence level calculated in the previous step, the partition level of each factor is thus determined (Table 5). For instance, the 'driving power' and 'dependence level' of the management of religious monuments are six and one. This means that the management of religious monuments influences six factors and gets influenced by one factor.

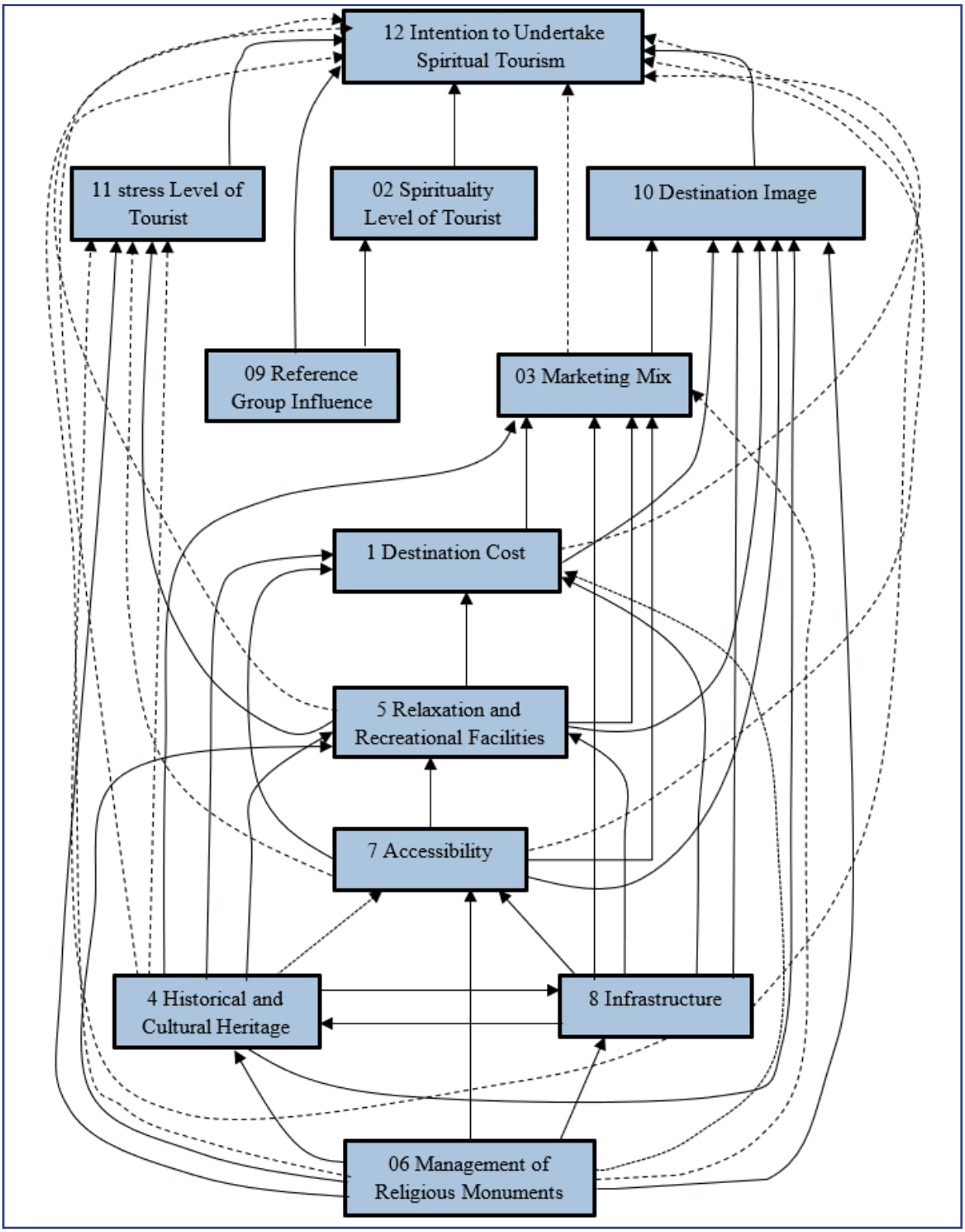
Partition level means the positioning of each of the factors in the hierarchical model (Table 5). The driving power (reachability set) and dependence level (antecedent set) form an intersection set. If the reachability set is equal to the intersection set at a level for a factor, that level becomes the partition level. This cycle is repeated until the partition level of each factor has been determined (see appendix). The results show that the management of religious monuments is placed at level one. At level two are infrastructure and historical and cultural heritage. The accessibility of the destination occupies the third level. At the fourth level is the presence of relaxation and recreational facilities. Destination cost occupies the fifth level. At the sixth level are reference group influence and marketing mix of the destination. Stress level, spirituality level of the tourists, and destination image occupy level seven. At level eight is the intention to undertake spiritual tourism.

The initial digraph (Figure 2) is prepared and represents all the links that exist between the factors. The placement of factors in the digraph follows increasing order, i.e., factors with level 1 are placed at the top and followed by factors at level 2, 3, up to level 8. Factors identified at the same level are placed at the same level in the digraph, i.e., horizontally adjacent.

The final ISM digraph (Figure 3) shows eight levels in the hierarchical structure of factors. Factors that are at a lower level drive the factors at the upper level of the hierarchy. At the bottom of the hierarchy is the management of religious monuments. It means that it channels the tourist motivation or intent to undertake spiritual tourism. It directly influences infrastructure development and historical and cultural heritage that are interdependent. Infrastructure development directly influences the accessibility of destinations which influences the presence of relaxation and recreational activities. These activities will impact the destination costs and the stress level of tourists, which influences the intent to undertake spiritual tourism. The destination cost influences the marketing mix directly and alters the destination image. The reference group influences the tourists' spirituality level and destination image, impacting the intent to undertake spiritual tourism.

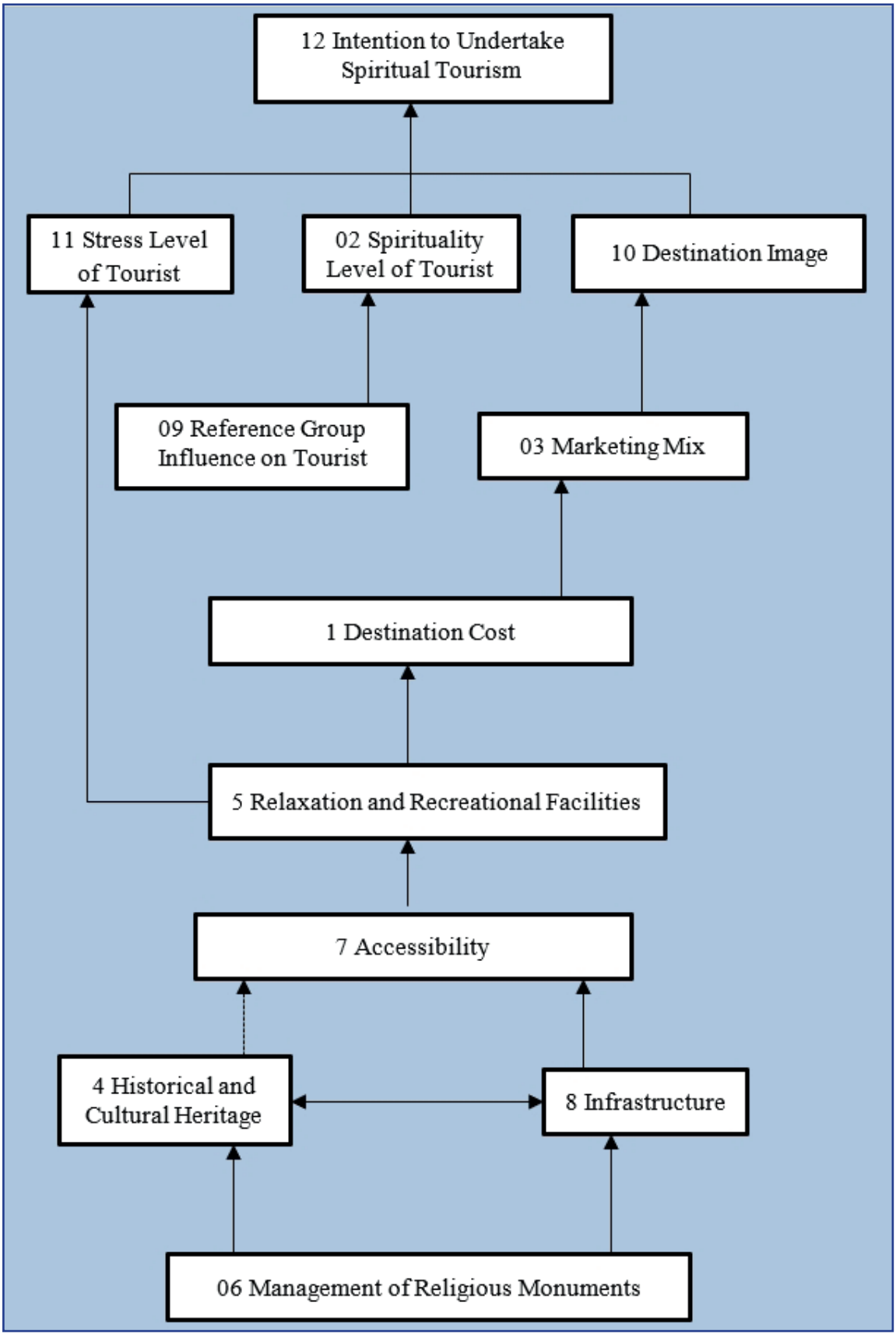
MICMAC analysis (Figure 4) shows that reference group influence on a tourist and spirituality level of tourists are autonomous factors. The management of

**Figure 2: Initial Digraph Showing Hierarchical Relationship Among Factors Influencing Spiritual Tourism**



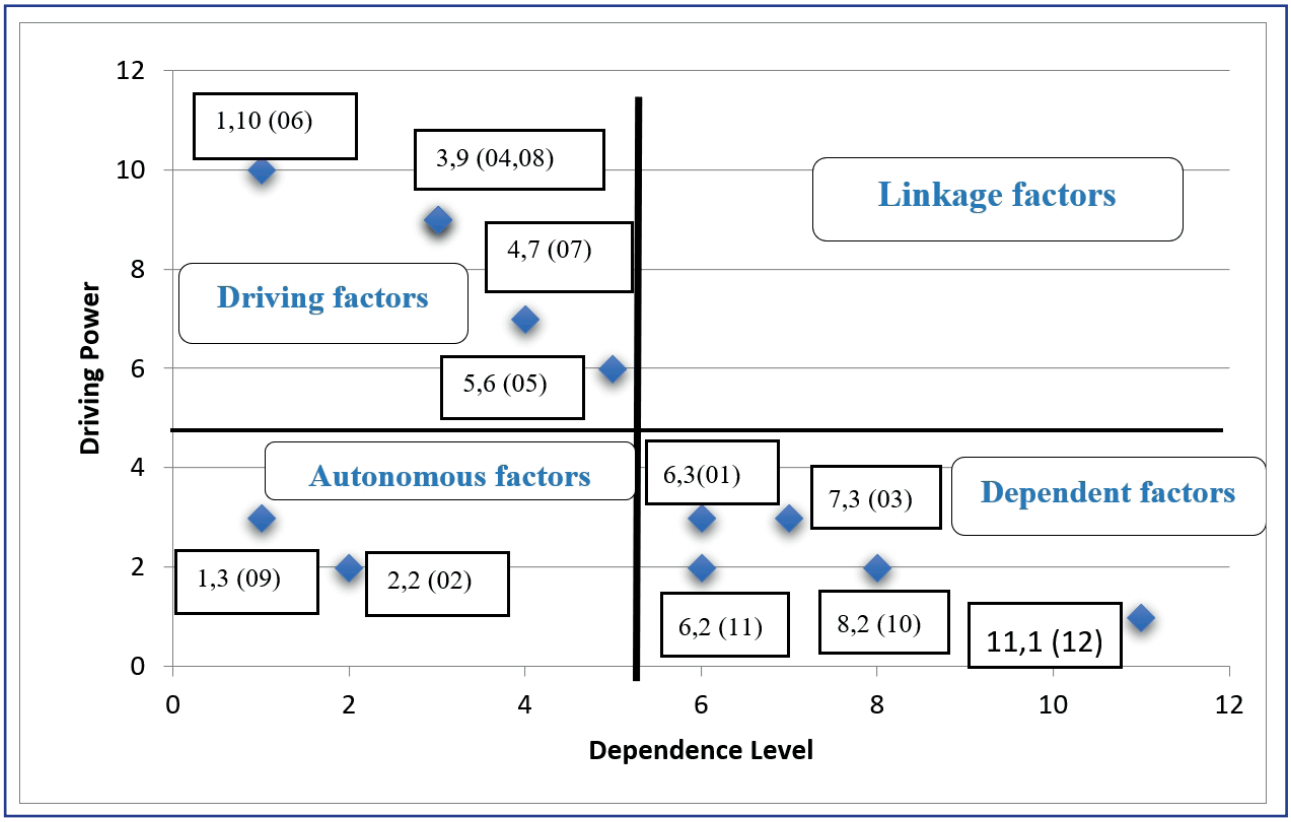

Source- Author's own

**Figure 3: Final Digraph Showing Hierarchical Relationship Among Factors Influencing Spiritual Tourism**




Source- Author's own

**Figure 4: Driving and Dependence Matrix**




Source- Author's own

religious monuments, historical and cultural heritage, infrastructure, relaxation and recreational facilities, and destination accessibility are driving factors. The destination cost, marketing mix, destination image, the stress level of tourists, and tourists' intention to undertake spiritual tourism are dependent factors. The analysis clarifies that all the factors except tourists' stress level and reference group influence on tourists are uncontrollable. Also, the results obtained are consistent with the digraph.

## Discussion and Implications

This study suggests several ways a spiritual tourist can be motivated for spiritual tourism consumption. The attitude of the management of religious monuments plays a pivotal role in attracting a spiritual tourist. They act as the central authority that permits or forbids tourism activity on their premises (Shinde, 2018). Tussles with the management of religious monuments will harm the development of spiritual tourism. At first, the management of the religious monuments will influence the development of infrastructure at a place. Infrastructure development is necessary for promoting tourism (Jovanović & Ilić, 2016; Seetanah, Teeroovengadum & Nunkoo, 2020). And then, the management will have control over the cultural and historical heritage of a place. However, infrastructure development should be aligned with cultural and heritage values as many destinations are visited more for their heritage qualities than religious qualities (Chrysanthi, Panoraia & Lagos, 2015; Egresi, Bayram & Kara, 2012).

Moreover, heritage can act as a transmitter of traditional customs and values (Alizadeh, 2011; Aulet & Vidal, 2018; Nolan & Nolan, 1992). Also, too much modern infrastructure tends to harm cultural values (Kumar & Shekhar, 2020). Infrastructure development will positively influence the accessibility of destinations; thereby, improving tourist satisfaction (Khuong & Luan, 2015; Mistry, 2018). Further, while increasing accessibility, care should be taken that historical and cultural heritage is not destroyed due to over-tourism. Relaxation and recreational facilities have proven to be stress-busters (Chen, Petrick & Shahvali, 2016; Ohe *et al.*, 2017). The development of these activities at tourist destinations will reduce incoming tourists' stress, and they might develop a liking for the place (Bagri & Kala, 2015). This liking might translate into intention to revisit the destination (Seetanah, Teeroovengadum & Nunkoo,

2020). Improved accessibility and infrastructure will further enhance the spiritual tourist motivation.

The development of relaxation activities will increase destination costs for tourists and the local community. There should be an assessment of social carrying capacity and a check on destination costs while developing these activities (Nicolau & Más, 2006; Seyidov & Adomaitienė, 2017; Zhuang, Yao & Li, 2019). An increase in destination cost will impact the destination's marketing mix, and that will alter the destination image and intention of a tourist to undertake spiritual tourism. The marketing mix determines the success of a product or a service (Alipour & Darabi, 2011; Bahman, Kamran & Mostafa, 2013). Lucrative offers, product bundling, and strategic development of spiritual tourism will prove beneficial for spiritual tourism (Zouni & Digkas, 2019). Destination image should be tailored for spiritual tourism. It should be created, keeping in mind the cultural and historical heritage, destination cost, marketing mix, and infrastructure at the destination (Kocyigit, 2016). However, once an image has been created, it can lead to stereotyping (Tu & Su, 2014; Tung, King & Tse, 2020), and there can be concerns that other typologies of tourism that are in contrast with spiritual tourism might not be able to prosper in such destinations.

Increasing stress levels are widespread in young people in urban areas (Lodha, 2018; Parikh *et al.*, 2019). So, businesses should be targeted as a market to spread awareness about spiritual tourism (Fedorova, 2016). The spirituality level of tourists is dependent on the influence of their reference group (Shekhar, Gupta & Valeri, 2021). So, elders in a family should be the target audience for spiritual tourism, influencing others to visit the destinations. Furthermore, measures should be undertaken to increase people's spirituality levels (Carr, 2008).

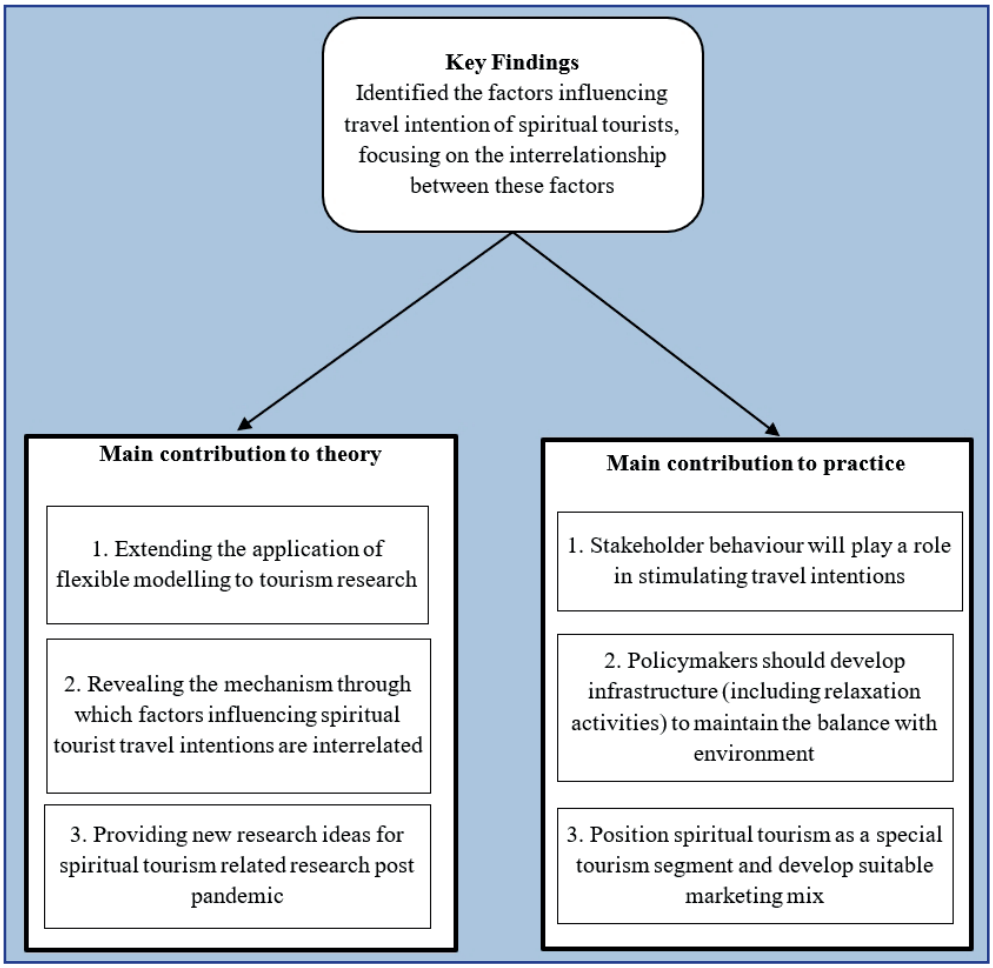
## Implications of Study

Thus study highlights gaps in spiritual tourism research and contributes to theory and practice (Figure 5). Several challenges and opportunities for spiritual tourism development are identified in this model. The spiritual level of the tourists and reference group influence are the autonomous factors. The pandemic outbreak has directly influenced people's spirituality level (Goodman, 2020; Ribeiro *et al.*, 2020). Thus, the possibility for spiritual tourism post-pandemic might increase amongst tourists. The intention to undertake spiritual tourism might be higher in older adults. As this age-group people are decision-makers or influence decisions, and reference group pressure is a motivator for spiritual tourism, this age group should be targeted by effective niche marketing. Further, destination costs should remain competitive to attract tourists. Building a destination image for spiritual tourism might be a challenge for policymakers. A primary challenge for policymakers will be to develop spiritual tourism as a 'mass tourism' or 'niche tourism' segment. The administrator's focus should be on preserving the heritage and monuments of places by ensuring development is undertaken to the highest levels of sustainable development.

Developing standardised infrastructure, marketing mix, and relaxation and recreational activities will depend on the nature of consumer targeted. Local community participation will be a must as locals are the flagbearers of a destination's cultural heritage. Small business entrepreneurs must be used to make spiritual tourism inclusive. Therefore, there should be a core group or working committee comprising entrepreneurs from all concerned sectors and the local community to improve the industry. Policymakers must develop trust with the management of institutions and balance tourist inflow and religious sanctity. Unwarranted incidents, (for instance, recent tussles with religious monuments happening across India), should be avoided. To provide authentic information about heritage and local customs, traditions, and values should be imparted to incoming tourists. Management must help tourists concentrate on enjoyment rather than diverting their attention towards struggling to find information.

This study suggests different ways to stimulate spiritual tourism development. Coupling this analysis with continuity-change matrix, High-Medium-Low (HML), and Vital-Desirable-Burden (VDB) analyses would help select a suitable channel to boost the industry's development. Spiritual tourism destinations are spread globally. Spiritually abundant nations must carry out exploratory studies to explore their destinations which could serve as tourist destinations, existing infrastructure at those destinations, and possibilities for tourism development in those sites and regions. To create necessary infrastructure at destinations, carrying capacity analysis must not be overlooked. Studies should be conducted to determine the physical, social, and economic carrying capacity of such destinations. Spiritual tourism must be

Figure 5: Key Findings and Contributions of Study




Source- Author's own

positioned in various communication channels to remind the audience of the product availability and its associated benefits. Researchers should empirically establish the role of spiritual tourism in reducing the stress levels of tourists.

## Conclusion and Study Limitations

The Sustainable Development Goals (SDGs) were designed by the United Nations in 2015. SDG 3, 8, and 11, propose that by 2030 regions must , devise policies and activities that promote good health and well-being, generate decent work and economic growth, resulting in sustainable communities and cities. Spiritual tourism development can significantly contribute to achieving sustainable development these goals. Spiritual tourism development also supports SDG 16, which focuses on achieving global peace and justice.

These desired goals can be achieved by adopting a strategic development approach for spiritual tourism. Strategic assessment will answer the necessary 'What' and 'How,' to frame effective policies. Thus, this paper attempts to promote strategic policy and decision-making in the development of spiritual tourism. It is a unique project where analysis has been undertaken to identify 'what motivates the spiritual tourist?' and addresses 'how the various variables are interrelated?' The study identified twelve variables influencing spiritual tourist motivation and modelled them through the interpretation of a hundred experts to show the interrelationships.

The analysis confirms the presence of several driving motivators that influence the other motivators for a spiritual tourist. By focusing on policy development around the drivers identified, the desired impact on the dependent factors can be achieved. Incorporating flexibility will remove the need to reformulate strategies from scratch in a dynamic environment. It will help in avoiding effort duplication and help in the optimum allocation of resources. The paper has listed managerial

and academic implications for future research in spiritual tourism.

The study, however, is not free from limitations. The study is based on respondents from India. Incorporating respondents from various countries might result in additional spiritual tourism motivators. On increasing the factors identified, the categorisation of the variables might change. The number of experts reached was limited. Increasing the sample size of experts (particularly including those with industry expertise) may lead to variations in results (Watson, 1978). The variations in results are an inherent drawback of ISM which is an interpretive methodology (Kumar & Shekhar, 2020). ISM also fails to explain the cause of the relationship between variables. The impact of COVID-19 on spiritual tourist motivators has not been addressed in the study.

Despite its limitations, this study will contribute to the literature on spiritual tourism and attract the interest of academics and policymakers.

## References


Adegbola, M.A. (2006) Spirituality and Quality of Life in Chronic Illness. *The Journal of Theory Construction & Testing* 10(2): 42–46.

Alipour, M. and Darabi, E, (2011) The role of service marketing mix and its impact on marketing audit In engineering and technical service corporations. *Global Journal of Management and Business Research* 11(6): 69–77.

Alizadeh, K. (2011) The Role of Historic-Religious Monuments in the Future Economy of Mashhad City. *Canadian Association of Geographers Conference*. Calgary, Alberta: University of Calgary: 1–11.

Alzebdeh, K., Bashir, H.A. and Al Siyabi, S.K. (2015) Applying interpretive structural modeling to cost overruns in construction projects in the sultanate of Oman. *Journal of Engineering Research* 12(1): 53–68.

Aulet, S. and Vidal, D. (2018) Tourism and religion: sacred spaces as transmitters of heritage values. *Church, Communication and Culture* 3(3): 237–259.

Bagri, S. and Kala, D. (2015) Tourists' satisfaction at Trijuginarayan: An emerging spiritual and adventure tourist destination in Garhwal Himalaya India. *Turizam* 19(4): 165–182.

Bahman, S.P., Kamran, N. and Mostafa, E. (2013) The effect of marketing mix in attracting customers: Case study of Saderat Bank in Kermanshah Province. *African Journal of Business Management* 7(34): 3272–3280.

Baker, D.C. (2003) Studies of the inner life: The impact of spirituality on quality of life. *Quality of Life Research* 12: 51–57.

Beerli, A. and Martín, J.D. (2004) Factors influencing destination image. *Annals of Tourism Research* 31(3): 657–681.

Bone, K, (2013) Spiritual Retreat Tourism in New Zealand. *Tourism Recreation Research* 38(3): 295–309.

Buzinde, C.N. (2020) Theoretical linkages between well-being and tourism: The case of self-determination theory and spiritual tourism. *Annals of Tourism Research* 83: 102920.

Carr, D. (2008) Music, Spirituality, and Education. *The Journal of Aesthetic Education* 42(1): 16–29.

Cheer, J.M., Belhassen, Y. and Kujawa, J. (2017) The search for spirituality in tourism: Toward a conceptual framework for spiritual tourism. *Tourism Management Perspectives* 24: 252–256.

Chen, C.C., Petrick, J.F. and Shahvali, M. (2016) Tourism Experiences as a Stress Reliever. *Journal of Travel Research* 55(2): 150–160.

Chrysanthi, B., Panoraia, P. and Lagos, D. (2015) Religious Tourism in Greece and regional development: The case of Samos Island. *55th Congress of the European Regional Science Association: 'World Renaissance: Changing roles for people and places*. Lisbon, Portugal: European Regional Science Association (ERSA): 1-25.

Cooper, R.G. (1994) New Products: The Factors that Drive Success. *International Marketing Review* 11(1): 60–76.

Counted, V., Possamai, A. and Meade, T. (2018) Relational spirituality and quality of life 2007 to 2017: An integrative research review. *Health and Quality of Life Outcomes* 16: 1-18.

Dowson, R. (2016) Event as spiritual pilgrimage: A case study of the 'Cherish' Christian Women's Conference. *International Journal of Religious Tourism and Pilgrimage* 4(2): 12–28.

Egresi, I., Bayram, B. and Kara, F. (2012) Tourism at Religious Sites: A Case from Mardin, Turkey. *Geographica Timisiensis* 21(1): 5–15.

Fedorova, A. (2016) *New Product Development in Spiritual Tourism Case Study: Omalaya Travel Ltd*. Savonia University of Applied Sciences: 1-62.

Fernandes, S. and Londhe, B.R. (2015) Influence of Social Reference Group on Buying Behavior, A Comparative Study of Working and Non Working Women in Bangalore-A Pilot Study Analysis. *Indian Journal of Science and Technology* 8(S6): 95–124.

Fry, P.S. (2000) Religious involvement, spirituality and personal meaning for life: Existential predictors of psychological wellbeing in community-residing and institutional care elders. *Aging and Mental Health* 4(4): 375–387.

Geary, D. (2008) Destination enlightenment: Branding Buddhism and spiritual tourism in Bodhgaya, Bihar. *Anthropology Today* 24(3): 11–14.

Goodman, B. (2020) *Faith in a time of crisis*, *American Psychological Association*. Available at: https://www.apa.org/topics/covid-19/faith-crisis.

Guo, Z., Li, Y. and Stevens, K.J. (2012) Analyzing Students' Technology Use Motivations: An Interpretive Structural Modeling Approach. *Communications of the Association for Information Systems* 30(14): 199–224.

Gupta, A., Gupta, S. and Shekhar (2021) Determining Interrelationship Between Factors Impacting Foreign Direct Investment in Tourism: An ISM-based Approach. *Asia-Pacific Journal of Management Research and Innovation* 17(1–2): 17–30.

Haq, F. and Jackson, J. (2009) Spiritual journey to Hajj: Australian and Pakistani experience and expectations. *Journal of Management, Spirituality and Religion* 6(2): 141–156.

Haq, F. and Wong, H.Y. (2010) Is spiritual tourism a new strategy for marketing Islam?. *Journal of Islamic Marketing* 1(2): 136–148.

Hartini, N.M. and Dewi, N.M.D.M. (2019) Factors are considered by spiritual tourist traveling in Tirta Empul Tampaksiring, Gianyar. *International Journal of Multidisciplinary Educational Research* 8(7): 355–362.

Ivona, A. and Privitera, D. (2019) Places and religious bands: Explorations in spiritual tourism. *International Journal of Religious Tourism and Pilgrimage* 7(4): 54–63.

Jafari, E., Najafi, M., Sohrabi, F., Dehshiri, G.R., Soleymani, E. and Heshmati, R. (2010) Life satisfaction, spirituality well-being and hope in cancer patients. *Procedia Social and Behavioral Sciences*, 5: 1362–1366.

Jovanović, S. and Ilić, I. (2016) Infrastructure As Important Determinant Of Tourism Development In The Countries Of Southeast Europe. *EcoForum* 5(1): 1–34.

Khuong, M.N. and Luan, P.D. (2015) Factors Affecting Tourists' Satisfaction towards Nam Cat Tien National Park, Vietnam — A Mediation Analysis of Perceived Value. *International Journal of Innovation, Management and Technology* 6(4): 238–243.

Kleber, R.J. (2019) Trauma and Public Mental Health: A Focused Review. *Frontiers in Psychiatry* 10: 451.

Kocyigit, M. (2016) The Role of Religious Tourism in Creating Destination Image: The Case of Konya Museum. *International Journal of Religious Tourism and Pilgrimage* 4(7): 21–30.

Koenig, H.G. (2012) Religion, Spirituality, and Health: The Research and Clinical Implications. *ISRN Psychiatry* 2012: 1–33.

Kumar, S. and Shekhar (2020) Digitalization: A strategic approach for development of tourism industry in India. *Paradigm* 24(1): 93–108.

Kumar, S., Shekhar and Guleria, N. (2019) Understanding dynamics of niche tourism consumption through Interpretive Structure Modeling. *Saaransh RKG Journal of Management* 11(1): 40–48.

Kumar, Y.P, (2018) Factors Influencing Development of Spiritual Tourism in Puducherry. *International Journal of Science and Research*, 8(9): 1087–1089.

Kuralbayev, A., Abishov, N., Abishova, A. and Urazbayeva, G. (2017) The Measurement of the Spiritual Tourism in Regions of South Kazakhstan. *European Research Studies Journal* 0(3A): 115–133.

Lifshitz, R., Nimrod, G. and Bachner, Y.G. (2019) Spirituality and wellbeing in later life: a multidimensional approach. *Aging and Mental Health* 23(8): 984–991.

Lodha, P. (2018) Internet addiction, depression, anxiety and stress among Indian youth. *Indian Journal of Mental Health* 5(4): 427–442.

Mistry, R.B. (2018) Influence of Destination Attributes Importance to Tourists' Satisfaction-Gujarat Tourism. *Amity Journal of Management Research* 3(1): 14–26.

Moal -Ulvoas, G. (2016) Spirituality and the Travel Motivations of Older Adults. *International Journal of Religious Tourism Pilgrimage* 4(4): 79-85.

Nicolau, J.L. and Más, F.J. (2006) The influence of distance and prices on the choice of tourist destinations: The moderating role of motivations. *Tourism Management* 27(5): 982–996.

Nolan, M.L. and Nolan, S. (1992) Religious sites as tourism attractions in Europe. *Annals of Tourism Research* 19(1): 68–78.

Noronha, K.J. (2015) Impact of Religion and Spirituality on Older Adulthood. *Journal of Religion, Spirituality and Aging* 27(1): 16–33.

Ohe, Y., Ikei, H., Song, C. and Miyazaki, Y. (2017) Evaluating the relaxation effects of emerging forest-therapy tourism: A multidisciplinary approach. *Tourism Management* 62: 322–334.

Parikh, R., Sapru, M., Krishna, M., Cuijpers, P., Patel, V. and Michelson, D. (2019) 'It is like a mind attack': Stress and coping among urban school-going adolescents in India. *BMC Psychology* 7(1): 1-9.

Phukan, H., Rahman, Z. and Devdutt, P. (2012) Emergence of spiritual tourism in India. *International Journals of Marketing and Technology* 2(4): 300–317.

Piewdang, S., Mekkamol, P. and Untachai, S. (2013) Measuring Spiritual Tourism Management in Community : A Case Study of Sri Chom Phu Ongtu Temple, Thabo district, Nongkhai province,Thailand. *Procedia-Social and Behavioral Sciences* 88: 96–107.

Power, R. (2015) Walking the Spiritual Ways – West of Ireland experience of modern pilgrimage. *International Journal of Religious Tourism and Pilgrimage*, 3(1): 46–54.

Ribeiro, M.R.C., Damiano, R.F., Marujo, R., Nasri, F., Lucchetti, G. (2020) The role of spirituality in the COVID-19 pandemic: a spiritual hotline project. *Journal of Public Health* 42(4): 855-856,

Seetanah, B., Teeroovengadum, V. and Nunkoo, R. (2020) Destination Satisfaction and Revisit Intention of Tourists: Does the Quality of Airport Services Matter?. *Journal of Hospitality & Tourism Research* 44(1): 134–148.

Seyidov, J. and Adomaitienė, R. (2017) Factors influencing local tourists' decision-making on choosing a destination: a Case of Azerbaijan. *Ekonomika* 95(03): 112–127.

Sharpley, R. (2016) Tourism and spirituality: An evolving relationship. *International journal of Tourism & Spirituality* 1(1): 8–24.

Shekhar, A.G. and Valeri, M. (2021) Mapping research on family business in tourism and hospitality: a bibliometric analysis. *Journal of Family Business Management* 00: 1-20.

Shinde, K.A. (2018) Governance and Management of Religious Tourism in India. *International Journal of Religious Tourism and Pilgrimage* 6(1): 58–71.

Singh, S. (2009) Spirituality and tourism An anthropologist's view. *Tourism Recreation Research* 34(2): 143–155.

Sirirat, P. (2019) Spiritual tourism as a tool for sustainability: A case study of Nakhon Phanom province, Thailand. *International Journal of Religious Tourism and Pilgrimage* 7(3): 97–111.

Skousgaard, H. (2006) A Taxonomy of Spiritual Motivations For Consumption. In Pechmann C and Price L (eds) *NA - Advances in Consumer Research*. Duluth: MN : Association for Consumer Research: 294–296.

Tripathy, S., Sahu, S. and Ray, P.K. (2013) Interpretive structural modelling for critical success factors of R&D performance in Indian manufacturing firms. *Journal of Modelling in Management* 8(2): 212–240.

Tu, H.C. and Su, S.J. (2014) How Tourist Destinations Could Be Stereotyped and Misplaced by Structural Forces? - A Case of Penghu Archipelago, Taiwan. In Mydin M A O and Marzuki A (eds) *SHS Web of Conferences*. EDP Sciences: 01019.

Tung, V.W.S King, B.E.M. and Tse, S. (2020) The Tourist Stereotype Model: Positive and Negative Dimensions. *Journal of Travel Research* 59(1): 37–51.

Vitorino, L.M., Lucchetti, G., Leão, F.C., Vallada,H., Peres, M.F.P. (2018) The association between spirituality and religiousness and mental health. *Scientific Reports* 8(1): 1–9.

Watson, R.H. (1978) Interpretive structural modeling-A useful tool for technology assessment?. *Technological Forecasting and Social Change* 11(2): 165–185.

Willson, G.B. and McIntosh, A.J. (2013) The spiritual dimension of heritage buildings. *ASEAN Journal on Hospitality and Tourism* 12(2): 107-118.

Zhuang, X., Yao, Y. and Li, J. (Justin) (2019) Sociocultural Impacts of Tourism on Residents of World Cultural Heritage Sites in China. *Sustainability* 11(3): 840.

Zouni, G. and Digkas, D. (2019) Marketing suggestions for multi-religious tourism development: The case of Thessaloniki. *Journal of Tourism, Heritage & Services Marketing* 5(2): 36–42.

# Appendix

| Iteration | Factor No | Reachability Set | Antecedent Set | Intersection Set | |
|---|---|---|---|---|---|
| Iteration 1 | 1 | 1,3,10 | 1,4,5,6,7,8 | 1 | I |
| | 2 | 2,12 | 2,9 | 2 | |
| | 3 | 3,10,12 | 1,3,4,5,6,7,8 | 3 | |
| | 4 | 1,3,4,5,7,8,10,11,12 | 4,6,8 | 4,8 | |
| | 5 | 1,3,5,10,11,12 | 4,5,6,7,8 | 5 | |
| | 6 | 1,3,4,5,6,7,8,10,11,12 | 6 | 6 | |
| | 7 | 1,3,5,7,10,11,12 | 4,6,7,8 | 7 | |
| | 8 | 1,3,4,5,7,8,10,11,12 | 4,6,8 | 4,8 | |
| | 9 | 2,9,12 | 9 | 9 | |
| | 10 | 10,12 | 1,3,4,5,6,7,8,10 | 10 | |
| | 11 | 11,12 | 4,5,6,7,8,11 | 11 | |
| | 12 | 12 | 12 | 12 | |
| Iteration 2 | 1 | 1,3,10 | 1,4,5,6,7,8 | 1 | II |
| | 2 | 2 | 2,9 | 2 | |
| | 3 | 3,10 | 1,3,4,5,6,7,8 | 3 | |
| | 4 | 1,3,4,5,7,8,10,11 | 4,6,8 | 4,8 | |
| | 5 | 1,3,5,10,11 | 4,5,6,7,8 | 5 | |
| | 6 | 1,3,4,5,6,7,8,10,11 | 6 | 6 | |
| | 7 | 1,3,5,7,10,11 | 4,6,7,8 | 7 | |
| | 8 | 1,3,4,5,7,8,10,11 | 4,6,8 | 4,8 | |
| | 9 | 2,9 | 9 | 9 | |
| | 10 | 10 | 1,3,4,5,6,7,8,10 | 10 | |
| | 11 | 11 | 4,5,6,7,8,11 | 11 | |
| Iteration 3 | 1 | 1,3 | 1,4,5,6,7,8 | 1 | III |
| | 3 | 3 | 1,3,4,5,6,7,8 | 3 | |
| | 4 | 1,3,4,5,7,8 | 4,6,8 | 4,8 | |
| | 5 | 1,3,5 | 4,5,6,7,8 | 5 | |
| | 6 | 1,3,4,5,6,7,8 | 6 | 6 | |
| | 7 | 1,3,5,7 | 4,6,7,8 | 7 | |
| | 8 | 1,3,4,5,7,8 | 4,6,8 | 4,8 | |
| | 9 | 9 | 9 | 9 | |
| Iteration 4 | 1 | 1 | 1,4,5,6,7,8 | 1 | IV |
| | 4 | 1,4,5,7,8 | 4,6,8 | 4,8 | |
| | 5 | 1,5 | 4,5,6,7,8 | 5 | |
| | 6 | 1,4,5,6,7,8 | 6 | 6 | |
| | 7 | 1,5,7 | 4,6,7,8 | 7 | |
| | 8 | 1,4,5,7,8 | 4,6,8 | 4,8 | |
| Iteration 5 | 4 | 4,5,7,8 | 4,6,8 | 4,8 | V |
| | 5 | 5 | 4,5,6,7,8 | 5 | |
| | 6 | 4,5,6,7,8 | 6 | 6 | |
| | 7 | 5,7 | 4,6,7,8 | 7 | |
| | 8 | 4,5,7,8 | 4,6,8 | 4,8 | |
| Iteration 6 | 4 | 4,7,8 | 4,6,8 | 4,8 | VI |
| | 6 | 4,6,7,8 | 6 | 6 | |
| | 7 | 7 | 4,6,7,8 | 7 | |
| | 8 | 4,7,8 | 4,6,8 | 4,8 | |
| Iteration 7 | 4 | 4,8 | 4,6,8 | 4,8 | VII |
| | 6 | 4,6,8 | 6 | 6 | |
| | 8 | 4,8 | 4,6,8 | 4,8 | |
| Iteration 8 | 6 | 6 | 6 | 6 | VIII |



View publication stats

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.

# Determination of ciprofloxacin using the Ziziphus mauritania derived fluorescent carbon dots based optical sensor

**Navneet Chaudhry**
Delhi Technological University
**Damini Verma**
Jawaharlal Nehru University
**Amit K Yadav**
Jawaharlal Nehru University
**Jai Gopal Sharma**
Delhi Technological University
**Pratima Solanki** ( pratimarsolanki@gmail.com )
Jawaharlal Nehru University

**Research Article**

**Keywords:** Fluorescent, Ciprofloxacin, Carbon quantum dots, Ziziphus mauritiana, Green synthesis

**Posted Date:** July 6th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1775535/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

# Abstract

The rapid determination and sensitive monitoring of antibiotics residues are critical challenges and highly essential for the protection of human and environment. Hence, in the present work, facile green synthesis of fluorescent carbon quantum dots (CDs) has been done using one step hydrothermal treatment of fruit exact of *Ziziphus mauritiana* for the first time, acting as natural source of carbon for optical determination of ciprofloxacin (CIP). These prepared CDs was used to selectively detect CIP, exhibiting high quantum yield of 30% at excitation wavelength of 340 nm. After the optimal parameters, the sensor gave the linear detection range from 20 to 100 µM along with limit of detection of 9.20 µM. The current work depicts the potential to effectively and rapidly determine other forthcoming environmental pollutants.


# 1. Introduction

Fluoroquinolones (FQs) comes under the class of antibiotics that are extensively used as human and veterinary medicines for fighting against many Gram positive (+) and gram negative (-) bacteria (Lavaee et al. 2017). Ciprofloxacin (CIP) is the most frequently used antibiotics among FQs that is oftenly utilize for treating animal and human diseases. (Picó and Andreu 2007) They are highly engaged in agricultural, medical and other areas owing to its high-cost efficiency and antimicrobial properties (Aldred et al. 2013) (Ionescu et al. 2007). But antibiotics are not completely digested by the body, a portion of their active component is discharged into the environment *via* wastewater (Borba et al. 2018)(Chaudhary et al. 2021). The leftover CIP might be transferred to the human body through the food chain and accumulate there in the system. In order to protect human health, numerous organisations or governments have established maximum residue limits. If a huge proportion of these residues stay in the body, it can cause a variety of negative responses as well as diseases, including pneumonia, respiratory tract infection, and cancer (Yu et al. 2017)(Zhang et al. 2014). As a result, it is critical that CIP should be detected using a cost effective, easy, environmentally friendly, and sensitive technique.

Many approaches have been investigated for CIP determination such as electrochemical techniques (Hu et al. 2018), colorimetry (Lavaee et al. 2017), capillary electrophoresis (Montes et al. 2014), high-performance liquid chromatography (Jin et al. 2016), and so on. However, these approaches need complex procedures, lengthy, are costly, and extensive sample processing. On contrast, the fluorescence sensor or optical sensor approach offers the benefits of non-destructive, simple, and cost effective (Li et al. 2019). Owing to its rapid, lack of pre-processing, and high sensitivity detection, fluorescence based optical sensors have been extensively applied to detect many analytes that includes biomolecules (Chai et al. 2021); proteins (Xue et al. 2021); and metal ions (Wang et al. 2021).

Carbon dots (CDs) are suitable alternatives for constructing fluorescent probes ascribing to its exceptional features, such as superb optical qualities (Wang et al. 2017)(Ramezani et al. 2018), water solubility, high chemical stability, and great biocompatibility (Gao et al. 2018)(Liu et al. 2019). There are widely employed in various fields such as medication administration, metal ion sensing, biological

imaging (Atchudan et al. 2019), biosensors (Anh et al. 2017), as well as in other domains (Zong et al. 2014). However, preparation of CDs typically entails the employment of chemical agents such as powerful oxidizers or acids, which can pose a risk to human health and pollute the environment. The hydrothermal approach was often employed by researchers since it is an environmentally friendly, facile, one-step procedure with a cheap synthesis cost than other methods (Hua et al. 2017). The hydrothermal technique of carbonization at low temperature with self-generated pressure is a green approach since no pollutant is emitted into the environment. To synthesize fluorescent CDs, several natural products such as tejpata leaves (Chaudhary et al. 2022), orange juice (Sahu et al. 2012), amino acids (Pei et al. 2015) etc. were employed as precursors. Therefore, it becomes critical to look into environmentally friendly, green and new precursors for the cost-effective preparation of fluorescent CDs. The use of green plant or ecologically friendly resources in the hunt for innovative precursors for CDs is becoming increasingly appealing. However, CDs quantum yield obtained from green resources is usually low (Feng and Zhang 2019)(Dager et al. 2019). In the hunt for achieving high quantum yield, we explored the CDs synthesis from natural precursor i.e., fruit extract of *Ziziphus mauritiana* as a source of carbon. *Ziziphus mauritiana* popularly recognized as Indian jujube,[1] or Indian plum is a fruit tree species that belongs to Rhamnaceae family. For centuries, it has been used as a herbal treatment. Its fruits that are rich in antioxidant property and valuable medical leaves makes this tree more attractive for providing economical products (Rahman et al., 2021). Hence, the goal of this study is to implement and create a reliable technique for detecting CIP utilising *Ziziphus mauritania* derived fluorescent CDs as a fluorescent agent.

In the present research, a fluorescent CDs derived from *Ziziphus Mauritania*' fruit extract were used for the optical detection of CIP synthesized *via* green methodology using one step hydrothermal treatment. The entire characterization of CDs was done and examined using X-ray diffraction (XRD), UV-Visible (UV-Vis), Fourier transform infrared (FTIR) and photoluminescence (PL) spectroscopy and scanning electron microscope (SEM). The fabricated CIP optical sensor was developed using the technique of PL and preparing CDs from chemical-free green source for the first time. The effect of CIP concentrations on the CDs PL intensity as well as the selectivity of CIP against various interferents were investigated. These CDs exhibits linearity from 20 to 100 µM for CIP determination, give blue emission along with high water solubility.

## 2. Experiments

# 2.1 Chemicals and Instrumentation

The fruit of *Ziziphus mauritiana* (Ber) were obtained from the nearby market of New Delhi, CIP, and other antibiotics were procured from SRL. All chemical reagents were utilized as such without pre-treatment. The crystal structure of CDs was done using X-ray Diffractometer (Miniflex, Rigaku). The investigation of the various functional groups is done through Fourier transformation infrared spectroscopy (FTIR), Perkin Elmer. The fluorescence as well as absorbace of CDs were performed employing the fluorescence spectroscopy, (Aligent Technologies, Cary Eclipse) and UV-Visible spectroscopy (T90 + UV/VIS

Spectrometer), respectively. The elemental composition and morphology of CDs were performed using the energy-dispersive X-ray (EDS) spectroscopy and scanning electron microscope (SEM), JEOL, respectively.

## 2.2 Synthesis of Carbon Dots

The preparation of fluorescent CDs was done using a facile green route through one step hydrothermal treatment. For this, 40 g of *Ziziphus mauritiana*' fruit was weighed and cut it into small pieces. The small pieces were then grinded and converted into paste that was diluted by dissolving it in 50 mL of distilled water to make the fruit extract. Next 20 mL of the fruit exact of *Ziziphus mauritiana* was poured in a Teflon-equipped stainless steel autoclave reactor and kept in muffle furnace for hydrothermal reaction to take place for 5 h at 190°C. Later after cooling the autoclave, centrifugation was done for 20 min at 8000 rpm for obtaining the dark coloured supernatant liquid (Kim and Kim 2018). Moreover, the supernatant was dialyzed *via* the dialysis membrane in 1 L of distilled water for 12 h with 3.5 KDa porosity in which distilled water was replaced in every 30 min. At last, yellowish-brown liquid was collected carrying the fluorescent CDs using a micropipette and placed in fridge 4°C for future experiment. Scheme 1 illustrate the preparation of CDs from *Ziziphus mauritiana* by low temperature, one step hydrothermal reaction followed by the detection of CIP *via* fluorescence enhancement of CDs.

## 3. Results And Discussion

The quantum yield of CDs derived from fruit extract of *Ziziphus mauritiana* was estimated at excitation wavelength of 340 nm in distilled water employing quinine sulfate as a reference. It was found that CDs yielded 30% of quantum yield from hydrothermal reaction which is much higher as expected from the others natural precursors of CDs. The possible reason for high yield could be ascribed to the high pressure as well as temperature conditions formed during the hydrothermal process providing well-organized carbonization of fruit extract.

The crystal structure of CDs was investigated via XRD. Figure 1 (a) depicts the XRD pattern showing as broad peak at 25.4**°**, which specify the interlayer spacing of (002) plane characteristic of graphitic diffraction pattern (Chaudhary et al. 2020).

The functional group was performed using FTIR present in CDs [Fig. 1 (b)]. The FTIR of CDs shows characteristic peak at 1507 and 3289 $cm^{-1}$ that are ascribed to N-H and N-H/O-H stretching, respectively. This depicts the existence of amino functional groups present on the surface of CDs (Liang et al. 2013). Also, other peaks such as 1020, 1697, 1736, and 2932 $cm^{-1}$ are accredited to C-O, C-C C-O, C-H stretching vibrations, respectively (Mehta et al. 2014). The strong peaks at 1388 and 1595 $cm^{-1}$ signify the symmetric and asymmetric stretching vibrations, respectively corresponding to the carboxylate anions (Mandal et al. 2012).

The dependence of PL intensity of CDs on excitation wavelength is depicted in Fig. 2 (a). The PL emission spectra of CDs at different excitation wavelengths ranging from 320 to 400 nm as shown in

Fig. 2 (b) (Liang et al. 2013). From this figure, an intense emission peak can be seen at 450 nm with 340 nm excitation wavelength. It is ascribed that at 340 nm wavelength, highest number of CDs are excited. Moreover, the surface characteristic of CDs could also be accountable for excitation-dependent fluorescence performance

Figure 2 (c) display the CDs UV-Visible absorption spectra. In this figure, two peaks are visible. One intense absorption peak at 284 nm and other minor shoulder peak at 250 nm that corresponds to $n$-π∗ transitions of the C-O and $\pi$-$\pi$∗ transitions of the C-C bond, respectively (Hu et al. 2014)(Zhang et al. 2012). The inset depicts the CDs in the aqueous suspension in daylight and exhibiting bright blue emission under UV light.

The pH response was performed in order to see the change in fluorescence intensity of synthesized CDs as displayed in Fig. 2 (d). For this experiment, various pH ranging from 2 to 12 was taken to observe change in florescence. The data depicts that maximum fluorescence was observed at pH 7.0 of CDs derived from fruit extract of *Ziziphus mauritiana*.

The SEM-EDS was performed to examine its morphology and find the composition of elements present in it, respectively [Fig. 3 (a,b)]. The EDS showed nine elements for example carbon (40.7wt.%), calcium (0.7wt.%), oxygen (41.8wt.%), nitrogen (1.2wt.%), potassium (6.9 wt.%), chlorine (2.8 wt.%), sulfur (0.07wt.%), silicon (4.4 wt.%), and sodium (1.0wt.%) as depicted in Fig. 3 (a). The existence of nitrogen (N) peak in the EDS depicts the nitrogen-doped CDs in which N atoms are provided by the amino acids present in the fruit extract of *Ziziphus mauritania*. Also, the oxygen (O) element could be ascribed to alkoxy, and hydroxy groups present on the CDs surface.

Figure 3 (a) EDS of CDs synthesized from fruit extract of *Ziziphus mauritiana*

The SEM image of CDs shows that they are spherical in shape. Moreover, the confirmation of EDS results was done using elemental mapping study. Figure 3 (b) shows the composition of nine elements in CDs, designated by different colors (C, O, Ca, K, S, Si, Cl, Na, and N).

# 3.1 Detection of Ciprofloxacin using Carbon Dots

The interaction of different CIP concentrations on CDs, PL intensity data is depicted in Fig. 4 (a). For this experiment, CIP concentrations from 1 to 100 μM were added in solution which lead to enhancement in fluorescence intensity as can be seen from the PL spectra at 448 nm excitation wavelength. This could be ascribed to the production of a complex formed between the amine group of CDs and carboxylate group of CIP. Also, a charge transfer from CDs to CIP can enhance the PL intensity (Kim and Kim 2018). Moreover, the increase in PL intensity of CDs could also be accredited to the hydrogen bonding interactions, as the CDs comprise of hydroxyl and carboxyl groups as confirmed from FTIR analysis that provides aqueous solubility to bind with CIP *via* hydrogen bonding.

Figure 4 (b) represents a relation between the PL intensity of CDs with concentrations of CIP. A good linearity can be seen between CIP concentrations and PL intensity of CDs ranging from 20–100 μM having regression coefficient ($R^2$) of 0.982 and LOD of 9.20 μM. The LOD was estimated using the standard formula of $3\sigma/S$ (where, $\sigma$ depicts the linear curve slope and $S$ is the standard error of intercept). This confirms that CIP could be determined from the enhancement in fluorescence of CDs.

## 3.2 Selectivity analysis of Carbon Dots against Ciprofloxacin

The CDs selectivity was performed utilizing different potential interfering chemicals for CIP determination [Fig. 4 (c)]. For this study, various antibiotics such as gentamycin, ampicillin, amoxicillin, tetracycline, moxifloxacin, levofloxacin and norfloxacin were taken, added one by one and its PL intensity was noted. Among all, only CIP exhibited a very intense fluorescence enhancement that displayed that CDs derived from *Ziziphus mauritiana* fruit extract have high selectivity towards it, enabling the determination of CIP against other interfering antibiotics.

# 4. Conclusions

In conclusion, a chemical-free synthesis procedure using *Ziziphus mauritiana* fruit extract as a carbon-rich source was used to develop an efficient and simple CDs in the aqueous medium. The characterizations of the green, fluorescent CDs were well confirmed and supported by XRD, FTIR, UV-Vis and PL for fluorescence sensor applications. The intense blue emission of CDs displayed good aqueous medium stability. The green synthesised CDs demonstrated remarkable selectivity for CIP determination among the other analytes examined by enhancement in fluorescence intensity having quantum yield of 30%. As a result, the fluorescent based optical sensor exhibited wide linearity from 20 to 100 μM along with LOD of 9.20 μM. The synthesised CDs possesses the benefits of selectivity and therefore can bind to the CIP *via* the functional group present on CDs surface. The developed synthesis procedure of CDs derived from the fruit extract of *Ziziphus mauritiana* is highly economically feasible, greener, convenient and could be utilized for detecting other antibiotics in future.

# Declarations


### Acknowledgments

Navneet Chaudhary acknowledge Indian Council of Medical Research (ICMR), New Delhi, India for financial help. Amit K Yadav is highly grateful to the Ministry of Education, Govt. of India for the Prime Minister Research Fellowship for economic aid.


### Author Contributions

**Navneet Chaudhary, Damini Verma:** conceptualization, formal analysis, methodology, software, visualization, data curation, writing— original draft, writing—review and editing. **Damini Verma:**

methodology, data curation, writing—review and editing. **Amit K Yadav:** writing— original draft, writing—review and editing; **Navneet Chaudhary:** investigation. **Jai Gopal Sharma:** supervision, writing— review and editing, resources. **Pratima R Solanki:** project administration, conceptualization, methodology, supervision, writing— review and editing, funding acquisition and Editing

**Ethical approval:** Not applicable

**Consent to participate:** Not applicable

**Consent for publication:** Not applicable.

**Competing interests**: The authors declare no competing interests.

**Data availability:** The datasets utilized or/and investigated during the present study are available from the corresponding author on reasonable request.

# References


Aldred KJ, McPherson SA, Turnbough Jr CL, et al (2013) Topoisomerase IV-quinolone interactions are mediated through a water-metal ion bridge: mechanistic basis of quinolone resistance. Nucleic Acids Res 41:4628–4639

Anh NTN, Chowdhury AD, Doong R (2017) Highly sensitive and selective detection of mercury ions using N, S-codoped graphene quantum dots and its paper strip based sensing application in wastewater. Sensors Actuators B Chem 252:1169–1178

Atchudan R, Edison TNJI, Perumal S, et al (2019) Green synthesized multiple fluorescent nitrogen-doped carbon quantum dots as an efficient label-free optical nanoprobe for in vivo live-cell imaging. J Photochem Photobiol A Chem 372:99–107

Borba FH, Schmitz A, Pellenz L, et al (2018) Genotoxicity and by-products assessment in degradation and mineralization of Ciprofloxacin by UV/$H_2O_2$ process. J Environ Chem Eng 6:6979–6988

Chai S, Zan G, Dong K, et al (2021) Approaching superfoldable thickness-limit carbon nanofiber membranes transformed from water-soluble PVA. Nano Lett 21:8831–8838

Chaudhary N, Gupta PK, Eremin S, Solanki PR (2020) One-step green approach to synthesize highly fluorescent carbon quantum dots from banana juice for selective detection of copper ions. J Environ Chem Eng 8:103720

Chaudhary N, Verma D, Sharma JG, Solanki PR (2022) A novel bioinspired carbon quantum dots based optical sensor for ciprofloxacin detection. Mater Lett 308:131090

Chaudhary N, Yadav AK, Sharma JG, Solanki PR (2021) Designing and characterization of a highly sensitive and selective biosensing platform for ciprofloxacin detection utilizing lanthanum oxide nanoparticles. J Environ Chem Eng 9:106771

Dager A, Uchida T, Maekawa T, Tachibana M (2019) Synthesis and characterization of mono-disperse carbon quantum dots from fennel seeds: photoluminescence analysis using machine learning. Sci Rep 9:1–12

Feng X, Zhang Y (2019) A simple and green synthesis of carbon quantum dots from coke for white light-emitting devices. Rsc Adv 9:33789–33793

Gao W, Song H, Wang X, et al (2018) Carbon dots with red emission for sensing of $Pt^{2+}$, $Au^{3+}$, and $Pd^{2+}$ and their bioapplications in vitro and in vivo. ACS Appl Mater Interfaces 10:1147–1154

Hu C, Yu C, Li M, et al (2014) Chemically tailoring coal to fluorescent carbon dots with tuned size and their capacity for Cu (II) detection. Small 10:4926–4933

Hu X, Goud KY, Kumar VS, et al (2018) Disposable electrochemical aptasensor based on carbon nanotubes-$V_2O_5$-chitosan nanocomposite for detection of ciprofloxacin. Sensors Actuators B Chem 268:278–286

Hua J, Yang J, Zhu Y, et al (2017) Highly fluorescent carbon quantum dots as nanoprobes for sensitive and selective determination of mercury (II) in surface waters. Spectrochim Acta Part A Mol Biomol Spectrosc 187:149–155

Ionescu RE, Jaffrezic-Renault N, Bouffier L, et al (2007) Impedimetric immunosensor for the specific label free detection of ciprofloxacin antibiotic. Biosens Bioelectron 23:549–555

Jin T, Wu H, Gao N, et al (2016) Extraction of quinolones from milk samples using bentonite/magnetite nanoparticles before determination by high-performance liquid chromatography with fluorimetric detection. J Sep Sci 39:545–551

Kim K, Kim J (2018) Synthesis of carbon quantum dots from jujubes for detection of iron (III) ions. J Nanosci Nanotechnol 18:1320–1322

Lavaee P, Danesh NM, Ramezani M, et al (2017) Colorimetric aptamer based assay for the determination of fluoroquinolones by triggering the reduction-catalyzing activity of gold nanoparticles. Microchim Acta 184:2039–2045

Li Z, Cui Z, Tang Y, et al (2019) Fluorometric determination of ciprofloxacin using molecularly imprinted polymer and polystyrene microparticles doped with europium (III)(DBM) 3phen. Microchim Acta 186:1–10

Liang Q, Ma W, Shi Y, et al (2013) Easy synthesis of highly fluorescent carbon quantum dots from gelatin and their luminescent properties and applications. Carbon N Y 60:421–428

Liu ML, Chen B Bin, Li CM, Huang CZ (2019) Carbon dots: synthesis, formation mechanism, fluorescence origin and sensing applications. Green Chem 21:449–471

Mandal A, Dandapat A, De G (2012) Magic sized ZnS quantum dots as a highly sensitive and selective fluorescence sensor probe for $Ag^{+}$ ions. Analyst 137:765–772

Mehta VN, Jha S, Kailasa SK (2014) One-pot green synthesis of carbon dots by using Saccharum officinarum juice for fluorescent imaging of bacteria (Escherichia coli) and yeast (Saccharomyces cerevisiae) cells. Mater Sci Eng C 38:20–27

Montes RHO, Marra MC, Rodrigues MM, et al (2014) Fast determination of ciprofloxacin by batch injection analysis with amperometric detection and capillary electrophoresis with capacitively coupled contactless conductivity detection. Electroanalysis 26:432–438

Pei S, Zhang J, Gao M, et al (2015) A facile hydrothermal approach towards photoluminescent carbon dots from amino acids. J Colloid Interface Sci 439:129–133

Picó Y, Andreu V (2007) Fluoroquinolones in soil—risks and challenges. Anal Bioanal Chem 387:1287–1299

Ramezani Z, Qorbanpour M, Rahbar N (2018) Green synthesis of carbon quantum dots using quince fruit (Cydonia oblonga) powder as carbon precursor: application in cell imaging and $As^{3+}$ determination. Colloids Surfaces A Physicochem Eng Asp 549:58–66

Sahu S, Behera B, Maiti TK, Mohapatra S (2012) Simple one-step synthesis of highly luminescent carbon dots from orange juice: application as excellent bio-imaging agents. Chem Commun 48:8835–8837

Wang S, Chen H, Xie H, et al (2021) A novel thioctic acid-carbon dots fluorescence sensor for the detection of $Hg^{2+}$ and thiophanate methyl via S-Hg affinity. Food Chem 346:128923

Wang S, Wang Y, Yang K, et al (2017) Synthesis of carbon dots originated from hydroxypropylmethyl cellulose for sensing ciprofloxacin. Anal Sci 33:1129–1134

Xue J, Yi J, Zhou H (2021) Label-Free Fluorescence Molecular Beacon Probes Based on G-Triplex DNA and Thioflavin T for Protein Detection. Molecules 26:2962

Yu W, Zhan S, Shen Z, et al (2017) Efficient removal mechanism for antibiotic resistance genes from aquatic environments by graphene oxide nanosheet. Chem Eng J 313:836–846

Zhang L, Xing Y, He N, et al (2012) Preparation of graphene quantum dots for bioimaging application. J Nanosci Nanotechnol 12:2924–2928

Zhang R, Wang Y, Yu L-P (2014) Specific and ultrasensitive ciprofloxacin detection by responsive photonic crystal sensor. J Hazard Mater 280:46–54

Zong J, Yang X, Trinchi A, et al (2014) Carbon dots as fluorescent probes for “off–on” detection of $Cu^{2+}$ and l-cysteine in aqueous solution. Biosens Bioelectron 51:330–335

# Figures

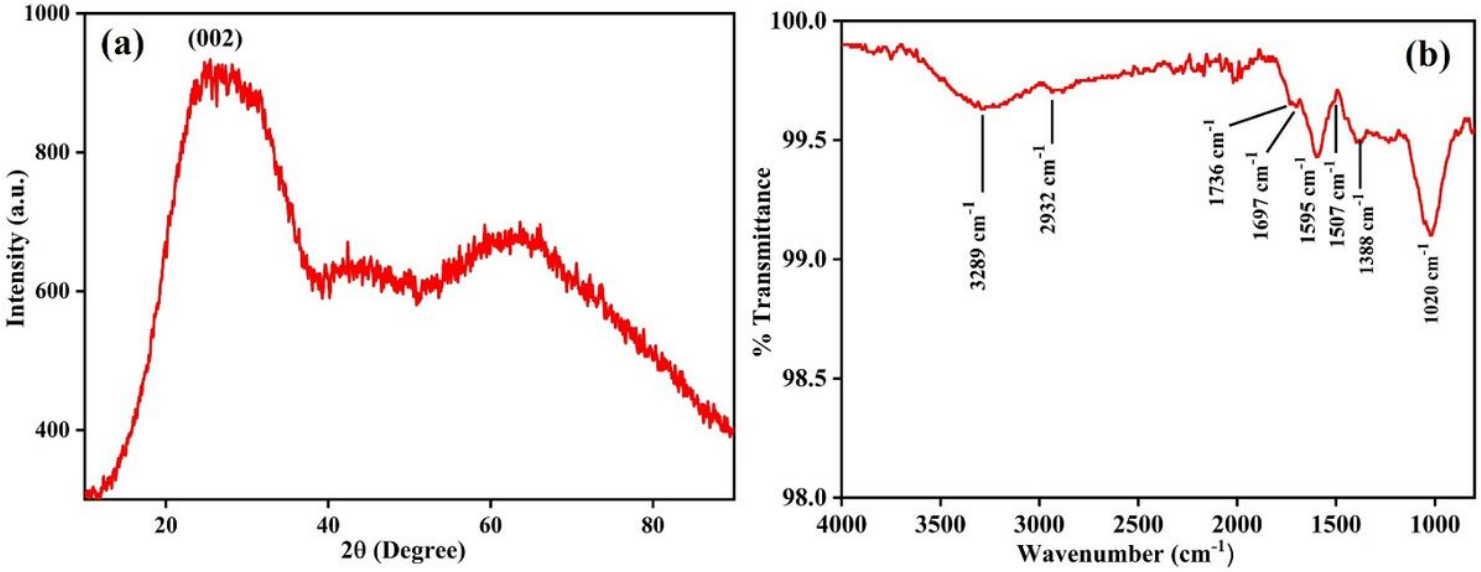
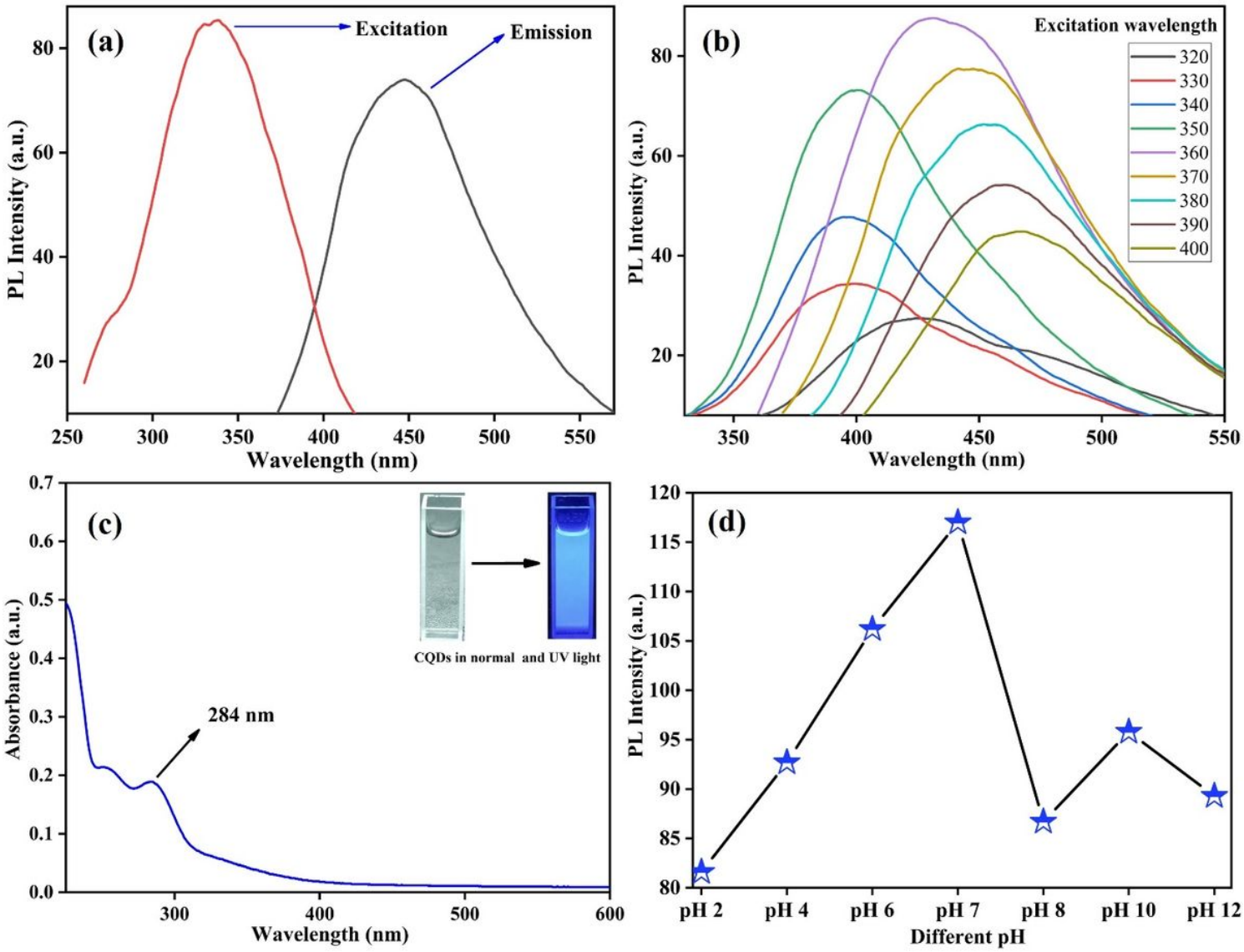



**Figure 1**

(a) XRD of fluorescent CDs, (b) FT-IR of CDs displaying the various functional groups




## Figure 2

(a) PL spectra showing emission and excitation wavelength of CDs, (b) PL spectra of CDs derived from fruit extract of *Ziziphus mauritiana* with 320-400 nm excitation wavelengths, (c) UV-Visible spectra of CDs, inset shows the fluorescent CDs in daylight and UV light; (d) pH response with change in fluorescence of synthesized CDs

## Figure 3

(a) EDS of CDs synthesized from fruit extract of *Ziziphus mauritiana*

**(b)** SEM image and EDS mapping of CDs

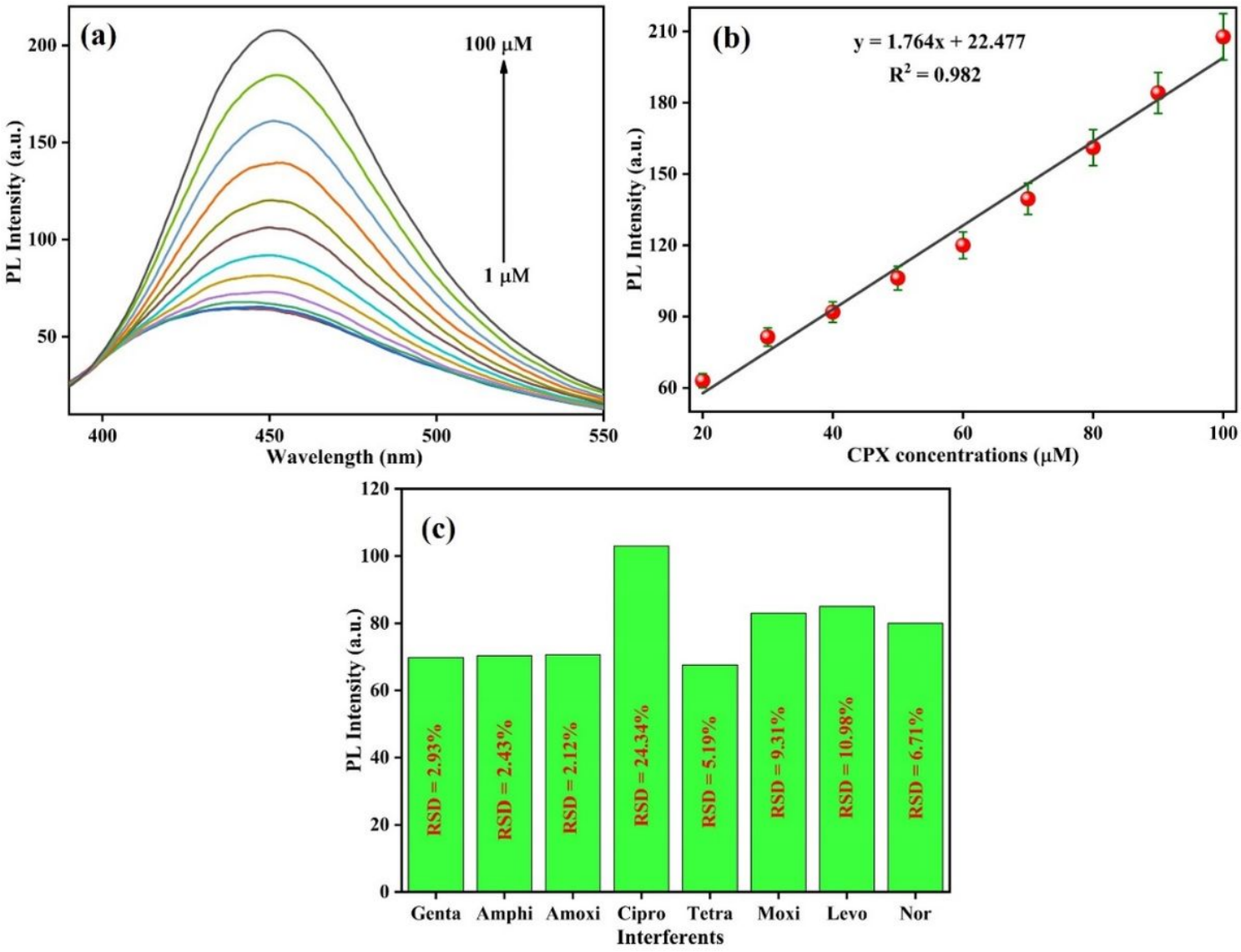



**Figure 4**

(a) The sensing response between the concentrations of CIP and PL intensity of CDs, (b) Calibration curve between concentrations of CIP and CDs PL intensity and; (c) Selectivity analysis of CDs against other antibiotics

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Scheme1.jpeg

*2022 3rd International Conference for Emerging Technology (INCET)*
*Belgaum, India. May 27-29, 2022*

# Dro-Mal Detector: A Novel Method of Android Malware Detection

Aman Chandok
*Department of Information Technology*
*Delhi Technological University*
Delhi, India
official.amanchandok@gmail.com

Aditya Verma
*Department of Information Technology*
*Delhi Technological University*
Delhi, India
polas.aditya@gmail.com

Rahul Gupta
*Department of Information Technology*
*Delhi Technological University*
Delhi, India
rahulgupta100689@gmail.com

2022 3rd International Conference for Emerging Technology (INCET) | 978-1-6654-9499-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/INCET54531.2022.9824877

***Abstract*—The increasing popularity of Android applications accompanies with it an ever-growing field of malware. Hence, the need of the hour is to design efficient malware detection techniques to ensure security. The API call graph of an Android Application Package (APK) is an efficient way to understand the behavior of an application and caller-callee relationships. However, traditional algorithms of graph similarity are not optimized enough to perform classification for large Android Application packages (APKs). Our study aims to use the function call graph of an API as a model to track malware behavior statically and dynamically. The call graph embeddings and permissions when converted to reduced dimensional feature vectors, are then trained with the help of Support Vector Classifier (SVC) whose hyperparameters are tuned with the help of the Grid Search algorithm. The same dataset is also trained with Artificial Neural Networks (ANNs) and compared by varying different hyperparameters. The results of both algorithms were compared and it was found that the Support Vector Classifier (SVC) performed better than the deep learning algorithm.**

***Keywords— APK, API, SVC, OS, ANN, LSTM, CNN, SGD, NLP***

## I. INTRODUCTION

Android is an open-source OS based on Linux essentially made for cellular phones and touchpads. Multiple program functions have come up with updates in technology in a much compact form by means of Android. Various features like a camera, calculator, calendar, games, modules, etc. have been integrated with Android devices. And now, the pressing use of these different applications due to their user-friendly nature has made Android immensely popular. Many gestures have been developed and awesome features like computing apps, memory management, cloud platforms, etc. have also come up. The Android applications are made up of a package in which various functions are interconnected to each other. The connection between the functions is termed as a caller-callee relationship. These functions and the relationships between them enable the smooth working of an Android application.

The Android market is becoming increasingly wealthy as Android smartphones have become so important for humans to communicate with each other around the globe. Nearly 344 million smartphones were manufactured in just the first quarter of 2017! According to the data published by International Data Corporation, Android has more than 85% share in the global mobile market, and this figure is nearly 5 times the share of iOS. The 2019 statistics show that Android maintains more than 2500 million active users per month. These applications are building up the industry of software and IT. They attract many customers but also malware developers to satisfy their malicious intentions. The self-signature facility on Android apps makes it hard to follow back to their designers. In 2016, about 20 million mobile malwares were identified with an abrupt increase of 100% relative to the previous year. 4 families and 60 new mobile threat clusters were found 2016. Malware designers are also focusing equally on remodeling existing malware families so as to alleviate their detection. As a result, Android malware needs to be classified well while dealing with a wide number of applications while using a smartphone. The detection system needs to be such that its parameters are set optimally to perform the most accurate prediction.

## II. RELATED WORK

There can be various approaches by which an APK can be classified as a malware or good ware. The analysis of malware for android applications can be classified into three major categories which include code-based, permission-based and behavioural-based models. The code-based analysis takes a large application dataset and groups malicious applications into similar categories [1]. It just uses code for analysis of an APK and it does not consider the obfuscated methods in that APK. The permission-models analyze the APK dynamically and statically and this analysis is used to draft out major permissions and system calls from application logs. The permissions gathered by the applications also do not cover the whole application at once [2] because in static analysis, the model just aggregates the static permissions which are not used during the execution of an application. The behavioral-based approach detects or monitors behavioral matches with the signatures defined by matching with regular expressions to find or detect the malign activity inside an application. However, this approach weakens with the increment in obfuscation of any application [3, 4].

The call tracking and system invocation data of an Android application are derived from its API call graph. This data is further represented inside a graph kernel to find the structural similarities and differences. The calls are tracked and features are extracted so that each node and edge can be identified and can be shortened to create a subgraph [5]. The call graphs can be represented with the Markov Chain system which is also known as the shortest path graph algorithm [6]. This algorithm is applied to find the similarities in various graphs. In [7], if the



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:26:50 UTC from IEEE Xplore. Restrictions apply.

node-set includes 160 instructions, the trace data is turned into a Markov process. This is depicted by a weighted-graph with unidirectional edges. The structural features and Android APK usually consist of 4 elements i.e., services, activities, content observers, and broadcast receivers. An edge is based on the lifecycle of a single node [8]. The malicious pattern and the threat modalities pattern are seen in unknown APKs to find out if a given application is malware or good ware. Call graphs are extracted by statically analyzing an executable file as mentioned in [9]. Data flow between the processes, files, sockets, and also the system registers are represented as a data flow diagram in [10]. Using the features of these data flows, malware detection is presented on behavioral databases. One of the methodologies also involves the opcode extraction from their instruction IDs followed by associating each opcode as in [11] with other potential opcodes through probabilistic edge weights. Call graphs of an APK are obtained to detect resemblance in an NP and a non-trivial problem [12] because the nodes of the graph are classified in accordance with Dalvik's fifteen instruction categories. A neural network can be used to convert the graph into embeddings as shown in [13] and then the final hybrid of these networks is trained to guarantee related functions. The technique to find the minimum edit distance of a graph is used to find commonalities in different graphs and the k-means cluster is used to finally ascertain whether an application is malicious or not.

In deep learning [14], a convolutional neural network (CNN) is assessed for Android malware classification. To extract knowledge from call sequences, LSTM is integrated. Naive Bayes algorithm and n-gram SVM were compared with the CNN results extracted from LSTM. With raw opcode sequences, a deep sequence can be built as in [15]. A research work found that was focused on network parameter tuning [16], [17] with different architectures of the network. It was tested to reach the highest term of accuracy. A 20% dropout at every layer leads to reduction in overfitting in a CNN in [18]. Hyperparameters are experimentally calculated as deep learning classifiers are presented in [19]. The difference between system call transition between malware and benign apps is given in [20], where Markov processes are applied to find system call sequences. The encoding effect in system calls into features with relative application sizes was experimentally examined and dynamically Android malware was detected using system call tracking is evaluated in [21]. Ransomware detection in [22] was detected using RNNs.

## III. Dataset

The use of deep learning and neural networks mandates the requirement of a cleaned and processed input in the form of a dataset. The dataset should contain the APKs of various Android applications and it should be large enough to ensure accurate prediction and classification without any bias. The CICInvesAndMal2019 dataset satisfies all the aforementioned criteria and is hence used for the purpose of this work. It contains diverse malware types such as Adware, Scareware, Ransomware, SMS-Malware, and each type further comprises of 42 different families. This dataset constitutes a total of nearly 10,000+ apps in which 6,500 are benign and 4,300+ are malware.

## IV. Proposed Methodology

### A. Extraction of Call Graphs from APKs

Androguard is an established tool which was used by researchers to extract the Call Graph for a given Android APK. The call graph of an Android APK is defined as a structural representation of the function/method set of an application. It is represented by:

$$G \in (V, E) \tag{1}$$

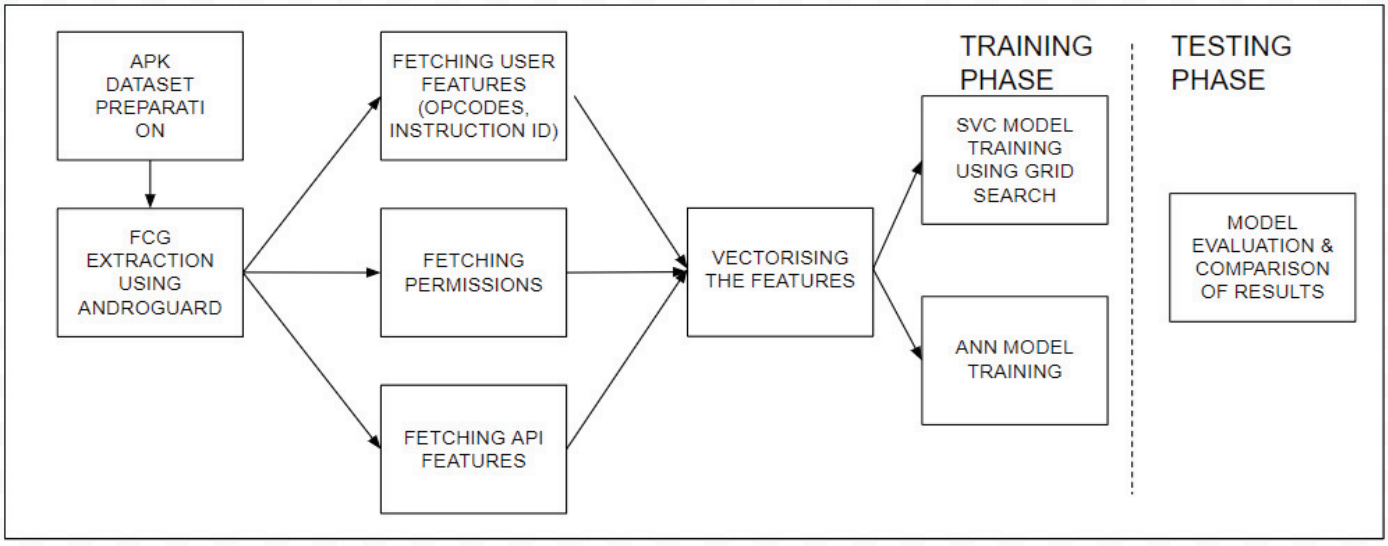



Fig 1. Flowchart of algorithm used



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:26:50 UTC from IEEE Xplore. Restrictions apply.

where G is the call graph composed of a node-set V and an edge-set E. The nodes are represented as:

$$V = \{v_j \mid j=1, 2, \ldots, n\} \quad (2)$$

The call-set between various methods of an APK are represented by:

$$E = \{(v_j, v_k) \mid v_j, v_k \in V\} \quad (3)$$

where the ordered pair represents the call from method $v_j$ to $v_k$. In this phase, the researchers analyzed the whole application and decompiled it, extracted its classes, functions and categorized them by the application of Androguard. This library was also used to produce a raw bytecode for a presented Android application. It can work while its other sessions are active. This means that when working over a big APK file the Androguard can delay the analysis of a particular APK in sessions and can be loaded back. And so, the researchers were able to perform bulk analysis with Androguard. This technique is commonly termed as *parallelization* which analyzes APKs via *multiprocessing*. Androguard is also useful in debugging broken apps. This usually happens with any APK if it uses corner cases or purposefully tries breaking the parsers which is common for malign applications. Once extracted, the next step was to visualize the call graph which is easily done with the help of an application called *GEPHI*. The command used in this work to generate *call graph* of an APK using Androguard is:

*Androguard cg -o* [OUTPUT PATH] [INPUT PATH]

### B. Feature Extraction

The next step followed by researchers involved the extraction of features from an APK *call graph* which are broadly obtained into two categories, user-level features and API-level features. These features consist of information about the nodes, edges, caller-callee relationships, class names, method names, descriptors, access flags, obfuscations, etc., in an Android application. Apart from these features, each node consists of one necessary feature known as an instruction type, or operational code, or simply *opcode*. The *opcode* mapping is a dictionary of values that point to different *opcodes* present in an APK. In this work, these mappings were extracted from the nodes of a call graph using the *Trie* data structure to a dictionary. Different *opcodes* result in different operations in an Android application. Table I shows all types of opcodes extracted mapped with their respective instruction IDs.

TABLE I. OPCODES OBTAINED WITH RESPECTIVE INSTRUCTION IDs.

| Opcode | Instruction IDs |
|---|---|
| 0x00 | Nop |
| 0x01 | Mov |
| 0x0E | return |
| 0x12 | const |
| 0x1D | monitor |
| 0x1F | check-cast |
| 0x20 | instanceof |
| 0x22 | New |
| 0x24 | Fill |
| 0x27 | Throw |
| 0x28 | goto/switch |
| 0x2D | cmp |

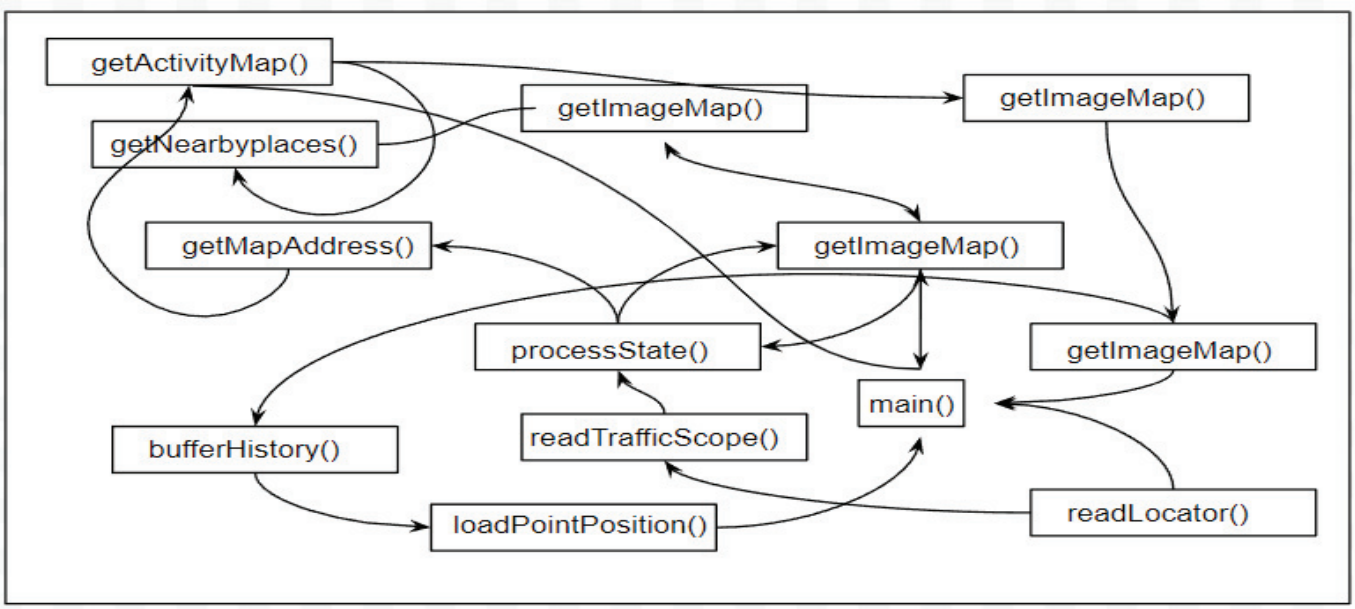



Fig 2. A sample representation of a call graph



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:26:50 UTC from IEEE Xplore. Restrictions apply.

The user-permissions and system-permissions are then extracted from the APK *call graph.* Some of the obtained permissions in various APKs are listed in Table II. Finally, all the aforementioned user-level and API-level features were consolidated and embedded into low-dimensional feature vectors and stored together in the final dataset. This dataset acts as the input for model training as well as model evaluation.

TABLE II. LIST OF APK PERMISSIONS.

| Permission | Description |
|---|---|
| BLUETOOTH | Connect with Bluetooth |
| RECORD_AUDIO | To record audio |
| READ_LOGS | To read log data |
| READ_SMS | To read SMS-messages |
| WRITE_SMS | To write SMS-messages |
| ACCESS_FINE_LOCATION | To access exact location using WiFi or GPS. |
| VIBRATE | Provide vibrator access |
| INTERNET | To open sockets of networking |

### C. Model Architecture and Training

Application of traditional machine learning algorithms over graphs is a tedious task because their time complexity is very high. A typical *call graph* structure consists of numerous different features as it represents the whole execution of an application, so this is a huge data to be processed. So, the primary step in this phase was to preprocess and vectorize the graph data, and this encoded feature-vector is stored in the final dataset as discussed in the previous step. After the conversion of the *call graph* frame to a lower-dimension, the researchers processed various learning algorithms onto it for classifying whether the given APK is benign or malware. A binary classification mechanism was used. The following classification methods were applied on the extracted feature vectors:

#### 1) Support Vector Classification (SVC)

A Support Vector Machine was used one of the binary classification model in this work. The basic approach followed was to maximize the linear classification margin in the given space of features. The partition between the two classes was described by the boundary between the nearest class instances, known as support vectors. This task can be simplified into a QP problem which is convex in nature and hence, the task of detecting Android malware also involves a binary classification model. After sufficient training of the classifier, a hyperplane was obtained as

$$w^x + b = 0 \quad (4)$$

where b indicates the bias, x indicates the input and w indicates weight-adjustable normal vector corresponding to the hyperplane. The hyperplane was a significant component of SVM as it splitted the feature space into two portions each representing different classes involved in the problem.

Once the training phase was completed, a decision function was obtained that was used to classify an unknown APK as malicious or benign.

#### 2) Artificial Neural Networks (ANN)

ANNs take less training time, are easy to deploy and require relatively small amount of APK dataset for training. In such type of networks, the weights are joined by different hubs. For the purpose of binary classification in this work, the most important step was to transpose the final-layer output to the labels of two classes. Hence, a simple network of fully connected layers using a sigmoid activation function was employed in this case. It can be computed in the following way:

$$y = \text{sigmoid}(\mathbf{W}_l x + b_l) \quad (5)$$

where $\mathbf{W}_l$ implies the matrix containing parameters and $b_l$ implies the bias-vector. A simple ANN usually contains three categories of layers i.e, input, hidden and output layer.

## V. Evaluation And Result

Once data preprocessing is performed and the CSV file containing feature vectors is obtained, Google Colaboratory is used to analyze it and perform predictions. As discussed, two major classification techniques are applied to the dataset obtained:

### A. Support Vector Classification (SVC)

A non-linear Support Vector Classifier is applied for detecting malign applications. As a primary step, the dataset is first divided into 80:20 ratio for training and testing respectively. Followed by this, hyperparameter tuning is performed with the help of Grid-Search, in which the range of three hyperparameters defined is as follows:

param_grid = {'C': [$10^{-1}$, $10^{0}$, $10^{1}$, $10^{2}$, $10^{3}$],
'gamma': [$10^{-4}$, $10^{-3}$, $10^{-2}$, $10^{-1}$, $10^{0}$],
'kernel': ['rbf']}

The optimal set of hyperparameters obtained is {'C': 10, 'gamma': $10^{-1}$, 'kernel': 'rbf'}. With these hyperparameters defined, the model is finally tested. The estimation of the correctness of a classification result is mainly done with the help of the following metrics:

#### 1) Accuracy

The ratio of correctly classified instances to the total number of instances is defined as accuracy.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

#### 2) Precision

The capacity of a classifier not to label a negative instance as a positive instance.

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

#### 3) Recall

The proportion of correctly classified benign APKs.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:26:50 UTC from IEEE Xplore. Restrictions apply.

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

*4) False Positive Rate (FPR):*

The proportion of incorrectly classified malicious APKs.

$$FPR = \frac{FP}{TN+FP} \quad (9)$$

*V. F1-Score*: This metric measures the model accuracy in its true sense as it takes the harmonic mean of precision and recall.

$$F1 - Score = \frac{2\,(Recall)(Precision)}{Recall+Precision} \quad (10)$$

*5) Support*

This metric defines true number of instances that actually belong to the specified class in the dataset.

*6) Macro-average*

It is simply calculated by taking the average of the metric for all classes in the dataset.

$$Macro - average = \frac{\sum_{i=1}^{n} M(i)}{n} \quad (11)$$

where M(i) denotes the metric value for the $i^{th}$ class and n denotes the number of classes.

*7) Weighted-average*

It is calculated by the average of the metric for each class in proportion to the number of instances belonging to that class.

$$Weighted - average = \frac{\sum_{i=1}^{n} M(i) \times n(i)}{n} \quad (12)$$

where M(i) denotes the metric value for the $i^{th}$ class, n(i) denotes the number of instances of the $i^{th}$ class and n denotes the number of classes.

The following classification results are hence obtained:

TABLE III. CLASSIFICATION REPORT

| | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Malign | 0.95 | 0.97 | 0.96 | 237 |
| Benign | 0.89 | 0.85 | 0.87 | 79 |
| Accuracy | | | 0.94 | 316 |
| Macro Average | 0.92 | 0.91 | 0.91 | 316 |
| Weighted Average | 0.94 | 0.94 | 0.94 | 316 |

In the testing phase, the model then predicts the nature of two sample APKs and its results are depicted in Table IV.

## B. Artificial Neural Networks (ANN)

In order to apply this deep learning algorithm, the dataset is again divided into 80:20 ratio for training and testing phase respectively. A simple 5-layered ANN model is then defined. The ANN architecture is as shown below:

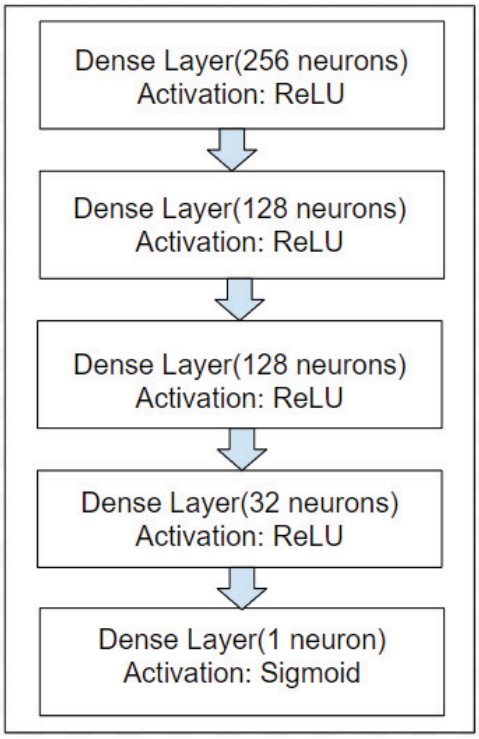



Fig. 1. ANN model architecture

After each dense layer, a 0.2 dropout layer ensures the avoidance of overfitting. The input and hidden layers are activated with ReLU function as it does not suffer from issue of vanishing gradient and hence ensures faster learning. The output layer of the network consists of just one unit activated with sigmoid function. The model is then compiled with SGD and Adam optimizers, applying the Binary Cross-entropy loss and a comparative performance analysis is done by training the model with 100, 200, 500 and 1000 epochs whose results are depicted in table V. In the testing phase, the models then perform predictions for two sample APKs whose results are nearly 100% accurate. From the accuracy and loss plots in table VI, one can conclude that the plots smoothen with increase in the number of epochs, especially in the case of SGD optimizer. The plots also become more concave in nature with increasing epochs. However, in case of Adam optimizer, the plots become inconsistent with increasing epochs, especially when the number of epochs is 1000. The loss in case of SGD and Adam increases with the number of epochs, but this increment is more abrupt in case of SGD. However, the accuracy increases with increasing epochs in SGD and nearly remains same in Adam. This confirms that SGD optimizer is better than Adam optimizer.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:26:50 UTC from IEEE Xplore. Restrictions apply.

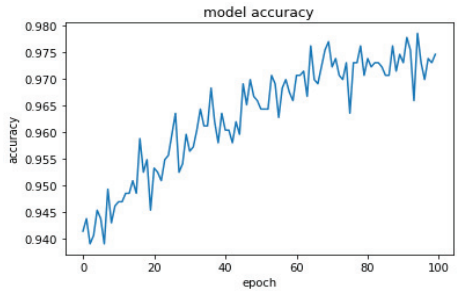
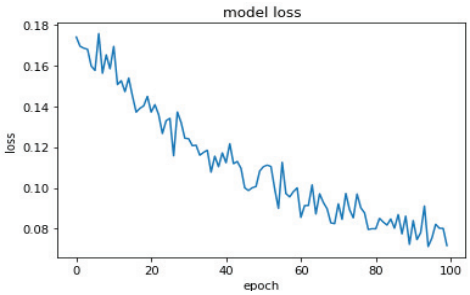
TABLE IV. PREDICTIONS BY SVC MODEL

| APKs | List of Permissions | Prediction |
|---|---|---|
| Ransomware/PornDroid/1c53e2c34d1219a2fae8fcf8ec872ac8.apk | ['android.permission.SYSTEM_ALERT_WINDOW', 'android.permission.WRITE_EXTERNAL_STORAGE', 'android.permission.READ_PROFILE', 'android.permission.READ_EXTERNAL_STORAGE', 'android.permission.READ_PHONE_STATE', 'android.permission.CAMERA', 'com.android.browser.permission.READ_HISTORY_BOOKMARKS', 'android.permission.VIBRATE', 'android.permission.WAKE_LOCK', 'android.permission.RECEIVE_BOOT_COMPLETED', 'android.permission.WRITE_SETTINGS', 'android.permission.ACCESS_NETWORK_STATE', 'android.permission.READ_CONTACTS', 'android.permission.GET_TASKS', 'android.permission.INTERNET', 'android.permission.GET_ACCOUNTS'] | [‘malign’] |
| benign/com.mobo.genius.market.apk | ['android.permission.SYSTEM_ALERT_WINDOW', 'android.permission.ACCESS_WIFI_STATE', 'com.google.android.c2dm.permission.RECEIVE', 'android.permission.READ_EXTERNAL_STORAGE', 'android.permission.WRITE_EXTERNAL_STORAGE', 'android.permission.NETWORK', 'android.permission.USE_CREDENTIALS', 'android.permission.WAKE_LOCK', 'android.permission.ACCESS_NETWORK_STATE', 'com.wearvn.testpush.permission.C2D_MESSAGE', 'com.mobo.genius.market.permission.C2D_MESSAGE', 'android.permission.INTERNET', 'android.permission.GET_ACCOUNTS'] | [‘benign’] |

TABLE V. COMPARATIVE PERFORMANCE ANALYSIS

| Model Name | Optimizer Used | Number of Epochs | Loss | Accuracy |
|---|---|---|---|---|
| ANN_sgd_100 | SGD | 100 | 28.65% | 91.77% |
| ANN_adam_100 | Adam | 100 | 87.06% | 92.72% |
| ANN_sgd_200 | SGD | 200 | 32.53% | 92.72% |
| ANN_adam_200 | Adam | 200 | 92.60% | 92.09% |
| ANN_sgd_500 | SGD | 500 | 48.28% | 92.09% |
| ANN_adam_500 | Adam | 500 | 160.82% | 90.82% |
| ANN_sgd_1000 | SGD | 1000 | 62.15% | 93.04% |
| ANN_adam_1000 | Adam | 1000 | 191.58% | 92.72% |

TABLE VI. COMPARISON USING PLOTS

| Model | Accuracy Plot | Loss Plot |
|---|---|---|
| ANN_sgd_100 |  |  |

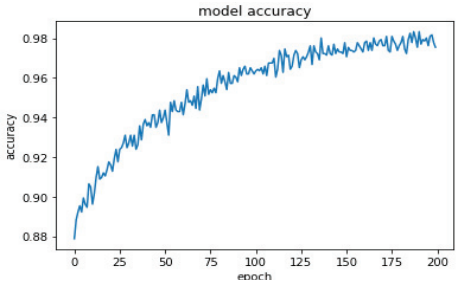


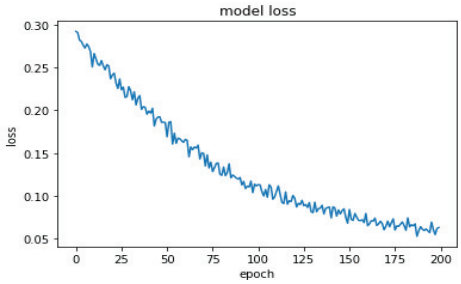




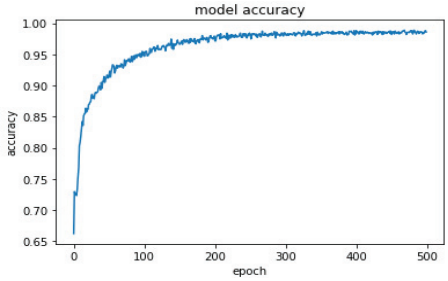
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:26:50 UTC from IEEE Xplore. Restrictions apply.

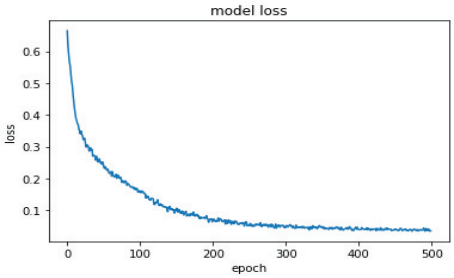
ANN_sgd_200



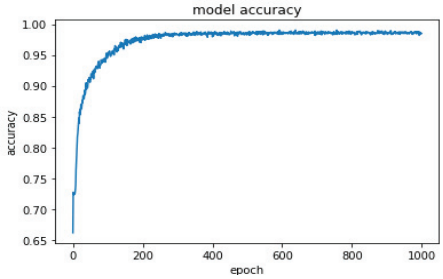



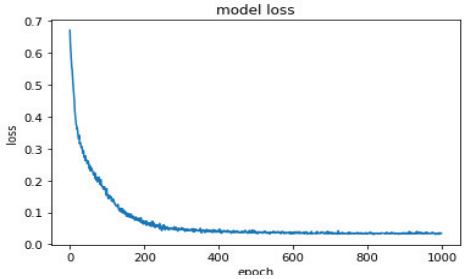

ANN_sgd_500



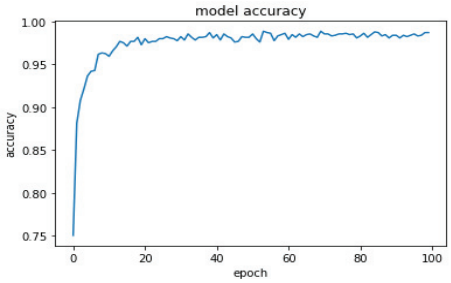



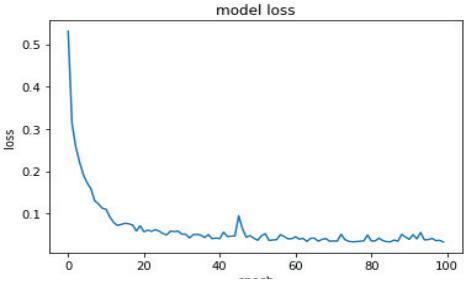

ANN_sgd_1000



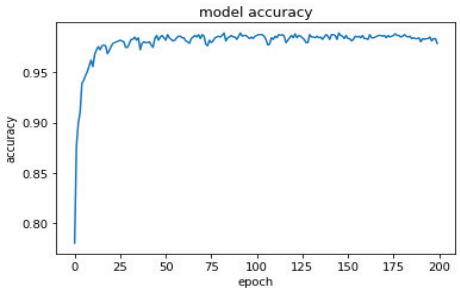



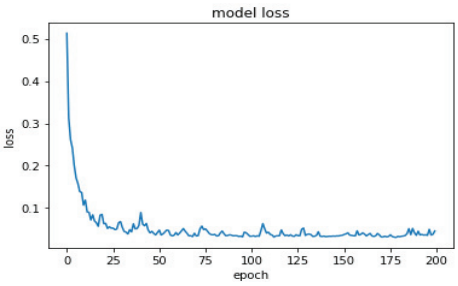

ANN_adam_100




ANN_adam_200

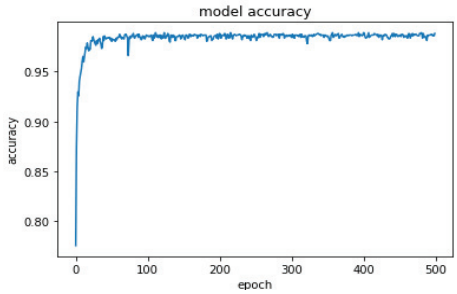


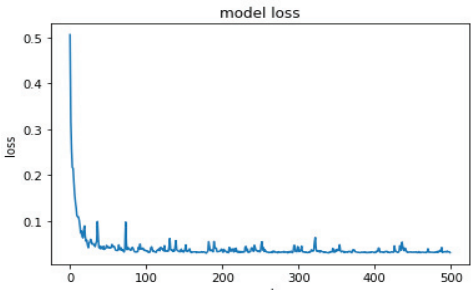





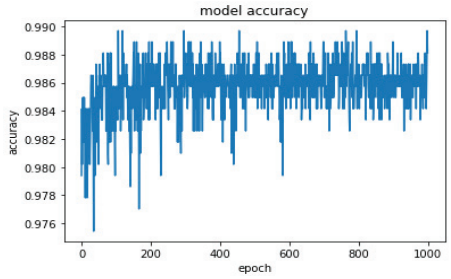

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:26:50 UTC from IEEE Xplore. Restrictions apply.

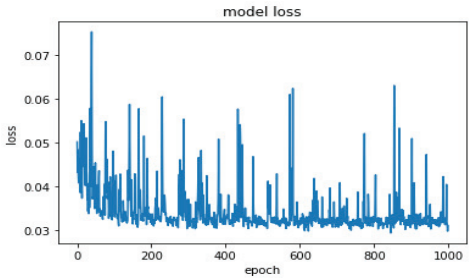
ANN_adam_500




ANN_adam_1000




## VI. Conclusions And Future Contribution

This work dispenses a novel approach based on APK dataset preparation by consolidating call-graphs feature, op-codes, and permissions extracted from the raw APKs. This dataset was trained and tested with Support Vector Classifier (SVC) and deep neural network and the results of the simulation are fairly accurate. This research will enable experimenters to be aware of the present condition, the advantages and disadvantages of each technique, and the future scope of development in this area. The primary contribution of this work is a depiction of the actualization of the approaches involved in Android malware analysis and regulations which can be self-regulated and associated with software development tools to aid analyses. However, there is scope of improvement in the techniques described in this paper. Numerous other features, as mentioned in [23, 24, 25, 26], once extracted, would enable the models to perform significantly better. NLP techniques can also be used to study sequences of various system-calls in an Android application. With the introduction of new malware families now and then, the training dataset needs to be continuously updated with more training data consisting of diverse malware families for more accurate detection.

## References


[1] Vij, Devyani, Vivek Balachandran, Tony Thomas, and Roopak Surendran. "Gramac: A graph based android malware classification mechanism." In *Proceedings of the tenth ACM conference on data and application security and privacy*, pp. 156-158. 2020.

[2] Tchakounté, Franklin. "Permission-based malware detection mechanisms on android: analysis and perspectives." *Journal of Computer Science* 1, no. 2 (2014).

[3] Geethanjali, D., Tan Li Ying, Chua Wan Jun Melissa, and Vivek Balachandran. "AEON: android encryption based obfuscation." In *Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy*, pp. 146-148. 2018.

[4] Chua, Melissa, and Vivek Balachandran. "Effectiveness of android obfuscation on evading anti-malware." In *Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy*, pp. 143-145. 2018.

[5] Cao, Shaosheng, Wei Lu, and Qiongkai Xu. "Deep neural networks for learning graph representations." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1. 2016.

[6] Borgwardt, Karsten M., and Hans-Peter Kriegel. "Shortest-path kernels on graphs." In *Fifth IEEE international conference on data mining (ICDM'05)*, pp. 8-pp. IEEE, 2005.

[7] Anderson, Blake, Daniel Quist, Joshua Neil, Curtis Storlie, and Terran Lane. "Graph-based malware detection using dynamic analysis." *Journal in computer Virology* 7, no. 4 (2011): 247-258.

[8] Yang, Chao, Zhaoyan Xu, Guofei Gu, Vinod Yegneswaran, and Phillip Porras. "Droidminer: Automated mining and characterization of fine-grained malicious behaviors in android applications." In *European symposium on research in computer security*, pp. 163-182. Springer, Cham, 2014.

[9] Kinable, Joris, and Orestis Kostakis. "Malware classification based on call graph clustering." *Journal in computer virology* 7, no. 4 (2011): 233-245.

[10] Wüchner, Tobias, Martín Ochoa, and Alexander Pretschner. "Robust and effective malware detection through quantitative data flow graph metrics." In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, pp. 98-118. Springer, Cham, 2015.

[11] Hashemi, Hashem, Amin Azmoodeh, Ali Hamzeh, and Sattar Hashemi. "Graph embedding as a new approach for unknown malware detection." *Journal of Computer Virology and Hacking Techniques* 13, no. 3 (2017): 153-166.

[12] Zeng, Zhiping, Anthony KH Tung, Jianyong Wang, Jianhua Feng, and Lizhu Zhou. "Comparing stars: On approximating graph edit distance." *Proceedings of the VLDB Endowment* 2, no. 1 (2009): 25-36.

[13] Xu, Xiaojun, Chang Liu, Qian Feng, Heng Yin, Le Song, and Dawn Song. "Neural network-based graph embedding for cross-platform binary code similarity detection." In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 363-376. 2017.

[14] Nix, Robin, and Jian Zhang. "Classification of Android apps and malware using deep neural networks." In *2017 International joint conference on neural networks (IJCNN)*, pp. 1871-1878. IEEE, 2017.

[15] McLaughlin, Niall, Jesus Martinez del Rincon, BooJoong Kang, Suleiman Yerima, Paul Miller, Sakir Sezer, Yeganeh Safaei et al. "Deep android malware detection." In *Proceedings of the seventh ACM on conference on data and application security and privacy*, pp. 301-308. 2017.




Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:26:50 UTC from IEEE Xplore. Restrictions apply.

[16] Yuan, Zhenlong, Yongqiang Lu, and Yibo Xue. "Droiddetector: android malware characterization and detection using deep learning." *Tsinghua Science and Technology* 21, no. 1 (2016): 114-123.

[17] Hou, Shifu, Aaron Saas, Yanfang Ye, and Lifei Chen. "Droiddelver: An android malware detection system using deep belief network based on api call blocks." In *International conference on web-age information management*, pp. 54-66. Springer, Cham, 2016.

[18] Nauman, Mohammad, Tamleek Ali Tanveer, Sohail Khan, and Toqeer Ali Syed. "Deep neural architectures for large scale android malware analysis." *Cluster Computing* 21, no. 1 (2018): 569-588.

[19] Martinelli, Fabio, Fiammetta Marulli, and Francesco Mercaldo. "Evaluating convolutional neural network for effective mobile malware detection." *Procedia computer science* 112 (2017): 2372-2381.

[20] Xiao, Xi, Zhenlong Wang, Qing Li, Shutao Xia, and Yong Jiang. "Back-propagation neural network on Markov chains from system call sequences: a new approach for detecting Android malware with system call sequences." *IET Information Security* 11, no. 1 (2017): 8-15.

[21] Dimjašević, Marko, Simone Atzeni, Ivo Ugrina, and Zvonimir Rakamaric. "Evaluation of android malware detection based on system calls." In *Proceedings of the 2016 ACM on International Workshop on Security And Privacy Analytics*, pp. 1-8. 2016.

[22] Rhode, Matilda, Pete Burnap, and Kevin Jones. "Early-stage malware prediction using recurrent neural networks." *computers & security* 77 (2018): 578-594.

[23] Zhang, Mu, Yue Duan, Heng Yin, and Zhiruo Zhao. "Semantics-aware android malware classification using weighted contextual api dependency graphs." In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pp. 1105-1116. 2014.

[24] Wei, Xuetao, Lorenzo Gomez, Iulian Neamtiu, and Michalis Faloutsos. "Profiledroid: Multi-layer profiling of android applications." In *Proceedings of the 18th annual international conference on Mobile computing and networking*, pp. 137-148. 2012.

[25] Yan, Lok Kwong, and Heng Yin. "{DroidScope}: Seamlessly Reconstructing the {OS} and Dalvik Semantic Views for Dynamic Android Malware Analysis." In *21st USENIX security symposium (USENIX security 12)*, pp. 569-584. 2012.

[26] Elish, Karim O., Xiaokui Shu, Danfeng Daphne Yao, Barbara G. Ryder, and Xuxian Jiang. "Profiling user-trigger dependence for Android malware detection." *Computers & Security* 49 (2015): 255-273.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:26:50 UTC from IEEE Xplore. Restrictions apply.

Environment, Development and Sustainability
https://doi.org/10.1007/s10668-022-02518-w

# Eco-design requisites for solar desaltification still augmented evacuated annular tube collectors with parabolic concentrator: an optimum-environ-economic viability

**Ashok Kumar Singh**[1,2] · **Samsher**[1]

Received: 25 September 2021 / Accepted: 9 June 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

**Abstract**
In this paper, an eco-design criterion for a novel solar desaltification setup (SDS) with evacuated annular tube collector (EATC) and a specific set of modified combination of parabolic concentrator has been examined for maximum performance with environmental and economic circumstances. This technique approaches to rectify the irregular utilization of EATC for optimum performance with environ-economic viabilities, which evidently satisfies the eco-design requirements and boosts the solar absorption capabilities uniformly along the periphery of vacuum tubes, and results improved thermo-syphon loom significantly more than in typical applications. The suggested unit is being refined and improved water temperature by 11.4% while keeping the basin lid and annular tubes at the same direction (30°). An incremental improvement in thermo-syphon circulation of 28.1% obtained through the present study. The average solar intensity of the respective clear day has been found as 401.8 kW and the overall energy and exergy efficiency on a daily basis are 50.8%, and 3.8%, correspondingly. At a minimal retail price of 0.07 $/l, and improved daily output by 12.3 kg per day than the typical SDS-EATC system taken for comparison into consideration are determined to be more satisfactory. 131.97 and 67.44 tons alleviates for $1318.36 and $673.77 from environmental generated money based on energy-exergy are there for $CO_2$, respectfully. The setup cost is noticeably reduced by 9.15% to the comparative system, and its productivity is determined to be 940.8% (> 100%), and this indicates that the present system is highly viable and appreciable for the feasible adaptation with the positive environ-economic possibilities.

**Keywords** Solar desaltification · Parabolic collector · Energy-exergy · Efficiency · Thermo-syphon

**List of symbols**

| | |
|---|---|
| $A_a$ | MCPC aperture area ($m^2$) |
| $A_g$ | Glass area ($m^2$) |

✉ Ashok Kumar Singh
agashok26@gmail.com

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

[2] Mechanical Engineering Department, Galgotias College of Engineering and Technology, Greater Noida, Uttar Pradesh 201306, India

Published online: 02 July 2022

| | |
|---|---|
| $A_b$ | Liner area ($m^2$) |
| $A_{rc}$ | Collector area of EATC ($m^2$) |
| $C_w, C_f$ | Water heat capacity (J/kg K) |
| $d$ | EATC interior diameter (m) |
| $E_e$ | Energy get (kWh) |
| $E_x$ | Exergy get (kWh) |
| $F'$ | Tube collector efficiency aspect |
| $G_r$ | Grashof No. |
| $h_{ba}$ | Net heat transport coefficient (liner-atmosphere) ($W/m^2$ K) |
| $h_{bw}$ | Convective heat transport coefficient (liner-water) ($W/m^2$ K) |
| $h_{t-g}, h_o$ | Thermal thrashing coefficient (glaze-atmosphere) ($W/m^2$ K) |
| $h_{t-wg}$ | Net heat transport coefficient (water-top glass) ($W/m^2$ K) |
| $h_{c-wg}$ | Convection heat transport coefficient (water-top glass) ($W/m^2$ K) |
| $h_{e-wg}$ | Evaporation heat transport coefficient (water-top glass) ($W/m^2$ K) |
| $h_{r-wg}$ | Radiation heat transport coefficient (water-top glass) ($W/m^2$ K) |
| $h_{r-ev}$ | Radiation thermal thrashing coefficient of EATC ($W/m^2$ K) |
| $h_{\text{saf}}$ | Net heat transport coefficient from EATC absorber to water ($W/m^2$ K) |
| $I_d(t)$ | Imitated radiation of ground ($W/m^2$) |
| $I_s(a)$ | Absorbed sun's ray energy ($W/m^2$) |
| $I_s(t)$ | Sun's ray energy of instance ($W/m^2$) |
| $I_b(t)$ | Irradiated solar beam radiation ($W/m^2$) |
| $K_g$ | Glaze conductivity (thermal) (W/m K) |
| $K_b$ | Lining conductivity (thermal) (W/m K) |
| $K_w$ | Medium (water) conductivity (thermal) (W/m K) |
| L | Heat latent (vaporization) (J/kg) |
| $L_{\text{EATC}}$ | Tube length (m) |
| $\dot{m}_{\text{ewg}}$ | Hour yielding (kg/h) |
| $\dot{m}_f$ | Water circulation rate (kg/s) |
| $m_w$ | Basin water amount (kg) |
| $m_f$ | EATC water amount (kg) |
| $n_o$ | Air index (refractive) |
| $n_g$ | Glass index (refractive) |
| $n_w$ | Water index(refractive) |
| $N_u$ | Nusselt No. |
| $N_{sh}$ | Sunshine hours |
| $P_r$ | Prandtl No. |
| $p_w$ | Fractional vapor pressure (water surface) (Pa, $N/m^2$) |
| $p_{gi}$ | Fractional vapor pressure (top glaze) (Pa, $N/m^2$) |
| $\dot{Q}_{N-\text{EATC}}$ | Heat gain rate from EATC (kJ/s) |
| $R_a$ | Rayleigh No. |
| $R_e$ | Reynold No. |
| $R_g$ | Glass reflectivity |
| $R_{i1}$ | Internal tube ID (m) |
| $R_{i2}$ | Internal tube OD (m) |
| $R_{o2}$ | External tube OD (m) |
| $R_{sc}$ | Absorber reflectivity |
| $T_a$ | Air temperature (°C) |
| $T_b$ | Lining temperature (°C) |
| $t_b$ | Lining thickness (m) |

| | |
|---|---|
| $t$ | Time (s) |
| $t_g$ | Glass thickness (m) |
| $T_{sa}$ | Temperature of EATC absorber (°C) |
| $T_f$ | EATC fluid temperature (°C) |
| $T_{fo1}$ | EATC outlet water (brackish) temperature (°C) |
| $T_w$ | Still water (brackish) temperature (°C) |
| $T_{go}$ | Outer glass facade temperature (°C) |
| $T_{gi}$ | Inner glaze facade temperature (°C) |
| $U_{saa}$ | EATC tube overall heat transmit coefficient |
| V | Averaged air speed (m/s) |

**Abbreviations**

| | |
|---|---|
| SDS | Solar desaltification scheme |
| EATC | Evacuated annulled tube collector |
| CPC | Combination of parabolic concentrator |
| OCPC | Orient combination of parabolic concentrator |
| MCPC | Modified combination of parabolic concentrator |
| CCPC | Cusp combination of parabolic collector |

**Greek Letters**

| | |
|---|---|
| $\nu$ | Viscosity (kinematic) ($m^2/s$) |
| $\beta'$ | Volumetric expansion (thermal) ($K^{-1}$) |
| $\alpha$ | Absorptivity (absorber) |
| $\mu$ | Viscosity (dynamic) ($N.s/m^2$) |
| $\alpha_g$ | Glass absorptivity |
| $\alpha_w$ | Water absorptivity |
| $\alpha_b$ | Liner absorptivity |
| $\eta_i$ | Instantaneous efficiency (%) |
| $\eta_e$ | Energy efficiency (%) |
| $\eta_x$ | Exergy efficiency (%) |
| $\tau$ | Transmissivity |
| $\sigma$ | Stefan–Boltzmann constant ($W/m^2\ K^4$) |
| $\rho$ | Density ($kg/m^3$) |
| $\varepsilon_{\text{eff}}$ | Effective emissivity |
| $\varepsilon_g$ | Glass emissivity |
| $\varepsilon_w$ | Water emissivity |

## 1 Introduction

The solar desaltification system (SDS) technique is inexpensive, atmosphere affable, and self-sustainable with insignificant maintenance outlay for commendable performance to produce potable water (Tiwari et al., 1992). On the other hand, a typical SDS has a greater heat loss of about 60% of net incident solar radiation (Malik et al., 1982). Singh et al. (2020) and Singh (2020) describe advances in this sector, demonstrating better yield with skillful pairings of supporting elements assisted in solar desalination models. Sato et al. (2012) also discovered that EATC is an effective idea in its thermal syphon form with homogeneous inter-mingling and temperature interaction to hot–cold

waterways. Morrison et al. (2005) investigated a range of 21 EATCs for radiant heat and assessed respective performance parameters, showing a favorable appearance in favor of thermo-syphon transit of EATC water volume. Budihardjo et al. (2007) used a vacuum tubular solar collector to investigate the efficient and responsible fluid heating factors such as the rectangular circuit of the thermo-syphon prototype for one Sun irradiation. Budihardjo and Morrison (2009) broadened the preceding lessons in the same method with 30 EATCs and linked the findings to FPC results. The conclusion revealed the photonic and temperature loss properties with a positive impact on the effectiveness of EATC for astral heating. Koffi et al., (2008a, 2008b) investigated the thermo-syphon technique for solar water heaters conceptually and empirically, with satisfactory results. Dev and Tiwari (2012) combined the EATC with an additional heat compartment mounted with SDS basin. The findings showed 1.9 L/day yielding and 30.1% efficiency without full utilization of the beneath section of EATCs (24). And, in an experiment of EATC (15 no.) with passive still, Sampathkumar et al. (2013) identified the yield production of 3.9 L/day with 43% efficiency. Singh et al. (2013) conducted a hypothetical analysis with passive still and associated EATC (10 nos.) and exposed daily yielding by 3.8 L/$m^2$ and 54.4% efficiency. Kumar et al. (2014) discovered that an enforced mode of operation deteriorates the performance of the system than thermo-syphon. Similarly, Yari et al. (2016) considered still with EATC (30nos.) and observed 58% decreased yielding. Issa & Chang (2017) tested a single steeply inclined (30) SDS system with EATC (5nos.) parallel to the ground, which slows the thermo-syphon loop. And Patel et al. (2019) performed experiments on a range of still with EATC setups to purify salty volumes of water and found that using EATC increased performance. Later, a plethora of other unique design options and effective combinations with various materials were shown with the recommendation of the special feature of EATC tubes used in solar desalination technology (Singh, 2020; Singh & Samsher, 2020, 2021a).

Furthermore, the use of solar energy in SDS results in a tiny quantity of harmful gases ($SO_2$, $CO_2$, NO, $NO_X$, etc.) being released into the environment as a consequence. With this, Tiwari et al. (2015) observed a better exergy value of $6.29 in a year for $14.5 in one ton of associated $CO_2$. Reddy and Sharon (2017) examined series annular flow with SDS for environmental and financial remarks and discovered a profitable setup with emissions of 1.6 ton $SO_2$, 221.8 ton $CO_2$ and 0.7 ton NO. Likewise, the $CO_2$ alleviate with other environmental and monetary findings for slanted still with wick arrangement have been analyzed and shown to significantly reduce carbon fouls (17.65 tons) from the model (Sharon et al., 2017). Yousef and Hassan (2019) investigated several factors for passive SDS, including exergo-environment-economy, and explored for higher carbon mitigation of 41.6 tons. In addition, a study for tubular SDS revealed a superior exergo-environment-economic value of $4.42 annually (Bait, 2019). And, Alayi et al. (2021) revealed the possibilities of solar heated water systems, especially for residential purposes and buildings and analyzed thermally and environmentally with appreciable results that show more than 70% energy saving through this system along with a noticeable energy saving of fossil fuels up to 237.3 kWh annually. Further, a study about the passive still augmented vacuum tubes collectors (10) was presented for economic, ecological, and energy matrices explanations with better mitigating 77.2 tones of $CO_2$ at a value of $771.23 in carbon credit gained money (Singh & Samsher, 2021b, 2022). Also, the similar studies for the double orientations of solar still systems were presented accordingly with positive results in embodied energy, yielding, and efficiencies (energy-exergy) performances as 1152.87 kWh, 16.94 kg/day, and 46.53%, 3.62%, correspondingly (Singh & Gautam, 2022; Singh & Samsher, 2021c).

The aforesaid studies suggest a preference for the use of EATC; however, the literature on the subject is lacking in terms of the best way to use EATC by neglecting the beneath area usefulness for getting uniform solar heat all over its circumference. Furthermore, the prior literature displays of heat syphon model (rectangular loop) other than the uniform heat syphon model, which is probable with the proposed hybrid configurations of SS-SDS augmented EATC-MCPC, which performs a cone-shaped thermal syphon with a superior circulation and consequences well for efficiency, economic growth, service life conversion, and clean water profitability in an eco-comradely way. In this way, the present communication of research work is centered to analyze the novel scheme and design requisites for clean water yielding through thermal modeling of the system for the recital of energy, exergy, efficiency, productivity, matrices, and profitability under respective summer (June) clear day conditions for 30 years of life span considerations.

The research problem that has been recognized as irregular utilization of a significant number of EATC (especially underneath and side segments), rectangle syphon circuit, which can be remedied by using modified parabolic reflectors in terms of novel design, operational conditions, and environ-economic perspectives. In view of the identified problems in this research area, this study explores on performance enhancement and economic design of system for the treatment of brackish/saline water through desaltification technique. The study relied on the work evaluation of the suggested mechanism (single directional slope solar still augmented parabolic concentrator assisted vacuum tube collector), which is a production efficient, performance metrics, economic assessments, as well as the system's impact on the green trail. The recommended scheme routine is compared through preceding studies. In the present study, the investigations aim to enhance the design requisites for clean water yielding and develop a thermal model for the analysis of the system recital for energy, exergy, efficiency, productivity, matrices, and profitability.

## 2 Projected system requisites

Figure 1 shows a schematic illustration of the current plan having the additional mountings of OCPC and CCPC (i.e., association of MCPC) and 15.91% increased basin water depth in comparison with the existing model as studied by Singh et al. (2013). The SS-SDS sink chamber is equipped with 0.78 W/m K thermal conductive window glass that is positioned at 30° (Tiwari, 2014; Tiwari & Tiwari, 2007). The basin lining is positioned over the floor of the basin with higher absorption potential (~0.97) for the gloss painted top (black) confirming the higher amount of solar heat grabbing. The whole basin area is equipped with a 0.05 m thick glass wool layer that ensures a minimum heat leakage from the wall surface and base to the air, with the exception of lid glazing (entrance of solar radiation to the still pool). Moreover, the internal face of basin walls (submerged part) is painted (matt black) surface to the level of water height (0.16 m) through which absorption of sun rays on the underwater wall surfaces is improved. The rest of the walls are glossy white coated that bounces incident sun rays toward the water that helps for temperature rising of basin water simultaneously. To increase the system's performance, a vacuum annular collector is added to the SDS setup (Fig. 2). At the aperture of the EATC, a valve provision is provided to check the reversing of water flow when there is an inadequate temperature for automatic thermo-siphon operating circumstances. The vacuum annular is spaced at a range of 0.079 m just above CCPC reflector, with a central gap of 0.314 m between them. Table 1 presents the parameters of the developed framework and parameters connected to its parts.

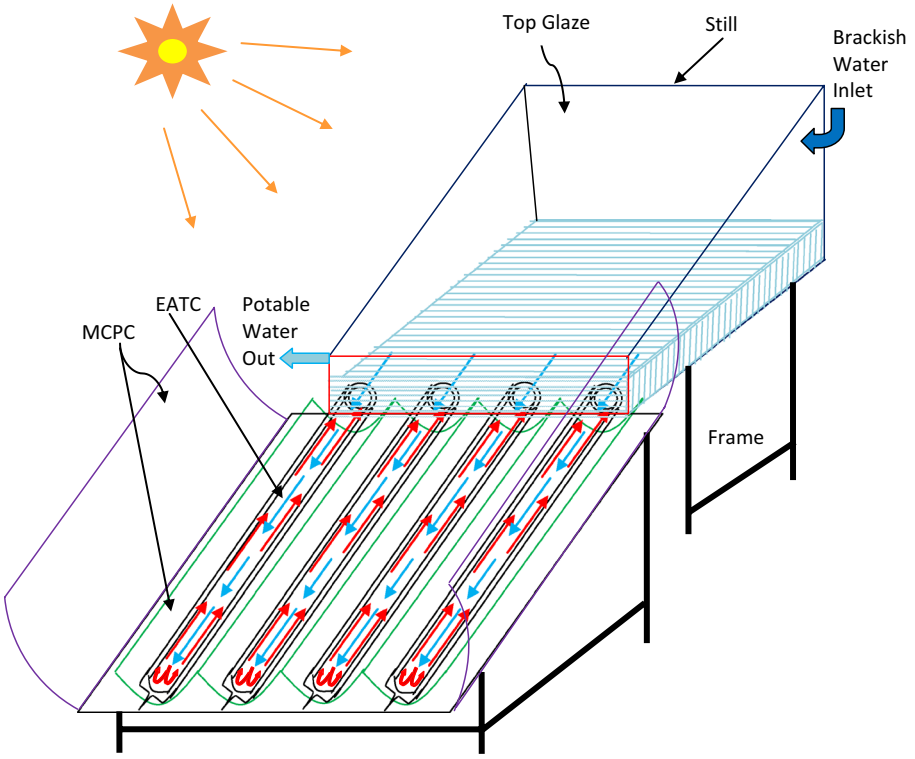



**Fig. 1** Representation of suggested scheme (single slopped desaltification model with vacuum annular collector and modified parabolic collector)

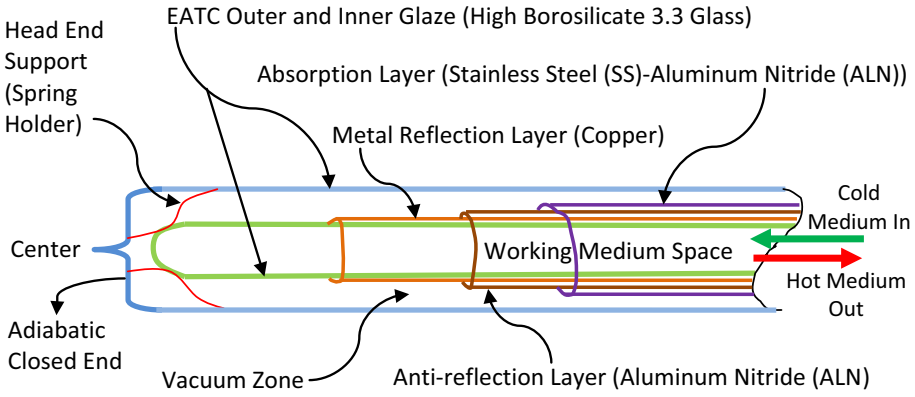



**Fig. 2** Scheme of evacuated borosilicate tube solar radiation collector

Furthermore, the parabolic modified reflector defines the cumulative influence of the cusp combination of the parabolic reflector (CCPC) and oriented combination of parabolic reflector (OCPC) assimilated with vacuum tube collectors to significantly enhance overall effectiveness, as no research study has previously used this innovative combination (edge

**Table 1** The suggested innovative system's specifications and analytical design considerations by following the various parameters and constant values of double slope solar still unit as presented by Singh and Gautam (2022)

| Single Sloped Solar Desaltification Model | | | |
|---|---|---|---|
| Element | Extent | Element | Extent |
| Basin length | 1 m | Frame substance | Mild Steel |
| Basin width | 1 m | Basin direction | South |
| Basin south height | 0.260 m | Water deepness | 0.16 m |
| Basin north height | 0.837 m | Liner depth | 0.05 m |
| Basin glass area | 1.155m$^2$ | Insulation depth | 0.05 m |
| Glass inclination | 30° | $m_w$(water mass) | 160 kg |
| Basin substance | FRP | $\alpha_g$(glass absorbtivity) | 0.05 |
| Insulating substance | Glass fiber | $\alpha_b$(liner absorptivity) | 0.97 |
| Cover substance | Window glass | $\alpha_w$(water absorptivity) | 0.34 |
| Cover depth | 0.004 m | $n_o$(air refraction index) | 1 |
| $\sigma$(Stefan-Boltzmann constant) | $5.6697 \times 10^{-8}$ W/m$^2$ K$^4$ | $n_g$(glass refraction index) | 1.5 |
| $\varepsilon_g$(glass emissivity) | 0.9 | $n_w$(water refraction index) | 1.3 |
| $\varepsilon_w$(water emissivity) | 0.9 | $K_g$(glass thermal conductivity) | 0.78 W/m K |
| C (plate constant) | 0.54 | $n$(plate number) | 0.25 |
| $K_b$(liner thermal conductivity) | 0.33 W/m K | $K_i$(insulation thermal conductivity) | 0.03 W/m K |
| Vacuum Annular Collector | | Parabolic Modified Reflector | |
| Element | Extent | Element | Extent |
| Tube length | 1.8 m | OCPC span | 1.8 m |
| Internal tube inner diameter | 0.047 m | OCPC partial opening | 0.935 m |
| Internal tube outer diameter | 0.049 m | OCPC net opening | 2.87 m |
| External tube inner diameter | 0.056 m | Reflector density | 2710 kg/m$^3$ |
| External tube outer diameter | 0.058 m | CCPC span | 1.8 m |
| Vacuum spacing | 0.007 m | CCPC opening | 0.314 m |

**Table 1** (continued)

| Vacuum Annular Collector | | Parabolic Modified Reflector | |
|---|---|---|---|
| Element | Extent | Element | Extent |
| Glass depth | 0.002 m | OCPC direction to tube | 40° |
| Absorber area | 0.0034 $m^2$ | OCPC approach to ground | 45° |
| Tube substance | Glass | CCPC direction | 45° |
| Central distance | 0.314 m | CPC thickness | 0.3 mm |
| $F'$(tube efficiency factor) | 0.968 | Reflectivity of CPC | 0.86 |
| $R_{sc}$(absorber reflectivity) | 0.05 | CPC substance | Al 1060 |
| α (absorptivity) | 0.95 | Reflector shade | Silver |
| τ (transmissivity) | 0.95 | Reflector to tube opening | 0.079 m |
| $h_{saf}$(net heat transfer coefficient) | 100 W/$m^2$ K | Reflector concentration fraction | 2 |

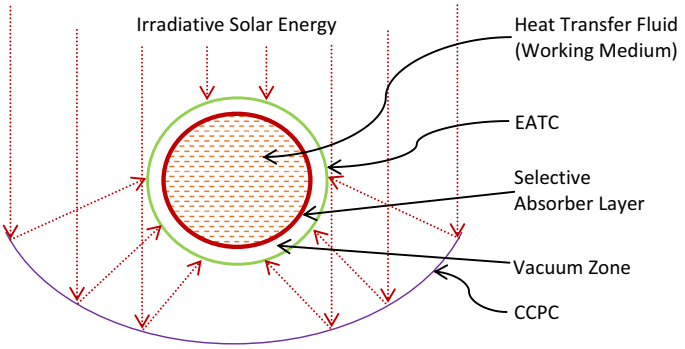



**Fig. 3** Scheme of EATC augmented CCPC wrapped underneath of EATC

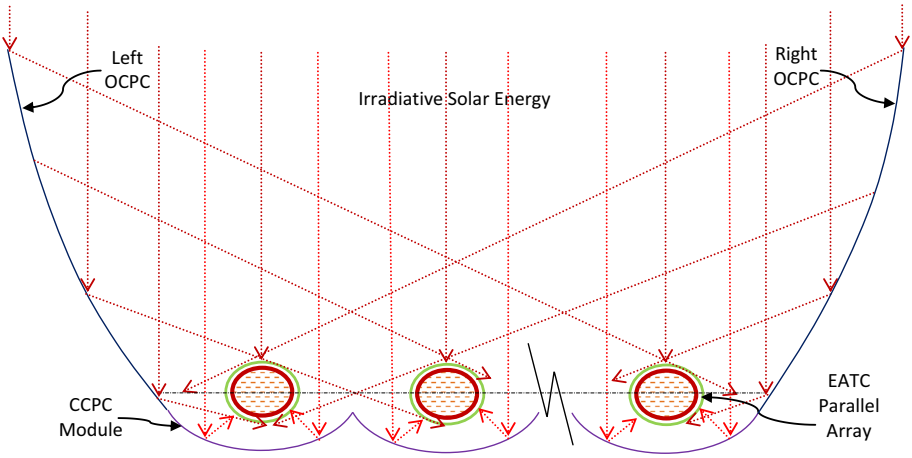



**Fig. 4** Scheme of EATC augmented MCPC wrapped sides as well as beneath to the EATC

and under sectional usage of vacuum tubes) and structure as confirmed in the presented system. Furthermore, each OCPC and CCPC measure has been thoroughly overlooked to ensure that no obstacles exist in receiving diffused and focused sunlight over the corresponding faces of the tube, as indicated in Fig. 4. As a result, the suggested coupling of side oriented reflector (OCPC) and underneath cusp reflector (CCPC) means MCPC exposes a creative association of design there in the domain of solar ray collectors aided with water treatment technology having the effective concentration range 2 of the parabolic profile for the projected model.

The following are the MCPC solar ray collector's significant achievements for the entire solar temperature distribution and associated mass flow from the tube to the basin along with the working of the model:

i. Because of the modest tube dimensions in relation to cusp diffusers, the underside circumferential region of the tube is unvaryingly lighted by cusp diffusers, with shadow impact over Cusp CPC being negligible (Fig. 3). The homogenous heat gain around the EATC perimeter is approximated by MCPC (Fig. 4).

ii. Due to the properties of colder water, the conical heat exchange between hot & cold water shifts lower, allowing for an increased flow of hot water movement to the basin. The intermixing of cold and hot streams is accomplished by an annular heat addition cycle with a longitudinal conical circuit.
iii. The solar desaltification model (Fig. 1) follows the close loop circulation of solar heated water that receives the heat from sun rays through the direct exposition of transparent top glass inclined covering and through the diffused-reflected cum concentrated radiations over cusp and side-oriented modified parabolas augmented with EATC tubes. This additional heat is further transmitted to the still basin chamber, increasing the water temperature, and thus improved vaporization enhances the overall yielding during the sunshine hours. The higher water depth supports the nocturnal yielding due to the availability of a large amount of sensible heat correspondingly.

## 3 Thermal model study

The focused thermal model development is directly dependent on the energy equilibrium equations trailed by the conservation of energy, 2nd law of thermodynamics with the aim of describing the operation principles and evaluating the system's performance. The assumptions behind the energy equilibrium equations are represented here accordingly.

i. The system is sealed, lumped, and quasi-steady with negligible heat absorbing abilities of the glaze, insulating material, and walls.
ii. The SS-SDS-EATC-MCPC components are flawlessly interconnected, with no temperature differential across the component thickness.
iii. The water channel interfaces (cold-hot) of EATC are ideally conical, however for clarity of computation and implicated as concentrically uniform and circular profiles across the EATC.
iv. There is no contact between neighboring tubes of EATC array, which is identical which made isolated during less or no light time.
v. With homogeneous and laminar merging of water channels (hot–cold), the thermal gradient in the EATC is unidirectional radial.

### 3.1 EATC-MCPC array

(a) Energy equilibrium equation for vacuum tube absorber

The absorber that has sandwiched layer of copper–aluminum-stainless steel, which is placed over the interior segment of the EATC tube, where sun rays are absorbed and moved to the surrounded water volume with a negligible heat loss toward the atmosphere. Further, the energy balance can be represented as follows (Tiwari et al., 2016; Singh and Gtam, 2022).

$$I_s(a) - U_{\mathrm{saa}}\left(T_{\mathrm{sa}} - T_a\right) = F^{'}.h_{\mathrm{saf}}\left(T_{sa} - T_f\right) \tag{1}$$

The term '$U_{saa}$' is the overall heat loss factor from EATC tube, and Eq. (1) can be expressed in the rearranged form to represent the temperature of EATC absorber ($T_{sa}$) and the corresponding constants and unknowns are referred through Table 1 and Appendix.

$$T_{\mathrm{sa}} = \left[(\propto \tau)_{\mathrm{eff}}.I_b(t) + F'h_{\mathrm{saf}}T_f + U_{\mathrm{saa}}T_a\right] / \left[F'h_{\mathrm{saf}} + U_{\mathrm{saa}}\right] \tag{2}$$

where $(\propto \tau)_{eff}$ represents the effective product of absorptivity and transmissivity of the respective annular tube.

(b) Energy equation of EATC working medium

The EATC is absolutely filled with brackish water and obtains heat energy from the sandwiched absorber of EATC. Furthermore, the equilibrium equation may be signified as (Singh & Gautam, 2022; Tiwari et al., 2016):

$$\dot{m}_f C_f \frac{dT_f}{dx}dx = F'\left[h_{\mathrm{saf}}\left(T_{sa} - T_f\right)\right]\left(2\pi R_{i1}\right)dx \tag{3}$$

$$h_{\mathrm{saf}}\left(T_{\mathrm{sa}} - T_f\right).F' + U_L\left(T_f - T_a\right) = F_1 I_b(t).(\propto \tau)_{\mathrm{eff}} \tag{4}$$

T rearranged form of the above expression can be represented as,

$$\dot{m}_f C_f \frac{dT_f}{dx} = \left(2\pi R_{i1}\right)\left[F_1 I_b(t).(\propto \tau)_{eff} - U_L\left(T_f - T_a\right)\right] \tag{5}$$

Following the boundary conditions,

$$T_f = \begin{cases} T_{fi}; \text{for} x \to 0 \\ T_{fo1}; \text{for} x \to L_{\mathrm{EATC}} \end{cases} \tag{6}$$

In the influence of boundary conditions, the above equation represents EATC water temperature as,

$$T_{fo1} = F_2 T_{fi} + \frac{\left(A_{rc}.F_r\right)}{\dot{m}_f C_f}\left[F_1(\propto \tau)_{\mathrm{eff}} I_b(t) + U_L T_a\right] \tag{7}$$

where, absorber area, i.e., $A_{rc} = 2\pi R_{o1} L_{EATC}$ and due to the identical EATC, the thermo-siphon loop is happening in a similar fashion to the other associative tubes. And, the expression for the Nth tube $T_{foN}$ is represented as written below.

$$T_{\mathrm{foN}} = T_{fi}.F_2 + \frac{A_{rc}.F_r}{\dot{m}_f C_f}\left[(\propto \tau)_{\mathrm{eff}} I_b(t).F_1 + U_L T_a\right] \tag{8}$$

Further, the equivalent temperature ($T_{N-fo}$) is attained by following the expressions,

$$\frac{1}{T_{N-fo}} \propto \frac{1}{T_{fo1}} + \frac{1}{T_{fo2}} + \ldots \frac{1}{T_{\mathrm{foN}}} \tag{9}$$

With the help of Eq. (9), one can get the net temperature $T_{N-fo}$ coming out from the parallel array as

$$T_{N-fo} = F_2 T_{fi} + \frac{N'.\left(A_{rc}.F_r\right)}{N.\dot{m}_f C_f}\left\{F_1(\propto \tau)_{\text{eff}} I_b(t) + U_L T_a\right\} \tag{10}$$

where $N'$ is the proportionality constant and approximately equal to $N$. And, heat gain for first and correspondingly Nth tube can be articulated as,

$$\dot{Q}_1 = \dot{m}_f C_f.\left(T_{N-fo} - T_{fi}\right) \tag{11}$$

$$\dot{Q}_N = \dot{m}_f C_f\left(T_{N-fo} - T_{fi}\right) \tag{12}$$

Further, the net thermal gain ($\dot{Q}_{N-EATC}$) is quantified by adding up the ake heat inputs and represented as,

$$\dot{Q}_{N-\text{EATC}} = N.\dot{m}_f C_f\left[\frac{\left(A_{rc}.F_r\right)}{N.\dot{m}_f C_f}\left\{F_1(\propto \tau)_{\text{eff}} I_b(t) + T_a U_L\right\} + T_{fi}F_2 - T_{fi}\right] \tag{13}$$

Further, the rearranged form specifically into its physical significance can be written as,

$$\dot{Q}_{N-\text{EATC}} = F_1(\propto \tau)_{\text{eff}}\left(F_r A_{rc}\right) I_b(t) - C_f \dot{m}_f.(N-1)T_{fi} - \left(A_{rc}.F_r\right)U_L\left(T_{fi} - T_a\right) \tag{14}$$

### 3.2 Solar desaltification scheme

(a) Top cover outer glass energy equation

The sun energy that strikes on the top of SDS (pointed south) follows the fractional reflection and absorption. The condensing face of the glass transmits heat to the atmosphere, and the corresponding equilibrium equation is represented as follows (Singh & Samsher, 2020; Tiwari et al., 2016):

$$\frac{K_g.A_g}{t_g}\left(T_{gi} - T_{go}\right) = \left(T_{go} - T_a\right)h_{t-g}A_g \tag{15}$$

From the above equation, outer glass temperature ($T_{go}$) can be written as

$$T_{go} = \left(T_{gi}\frac{K_g}{t_g} + h_{t-g}.T_a\right)/\left(h_{t-g} + \frac{K_g}{t_g}\right) \tag{16}$$

Furthermore, Eq. (15) can be reshuffled into an additional form as articulated,

$$\frac{K_g.A_g}{t_g}\left(T_{gi} - T_{go}\right) = (T_{gi} - T_a)U_c A_g \tag{17}$$

(b) Equilibrium equation for top cover inner glass

The condensing zone is heated in two distinct ways: first, it gets intense solar radiation (partial), and second, it gains heat from the basin water. There is also a conduction heat

loss from the top glass to outwards. All of this may be written as an equilibrium equation, which appears as:

$$\alpha_g^A.A_g.I_s(t) = \frac{K_g.A_g}{t_g}\left(T_{gi} - T_{go}\right) - h_{t-wg}.\left(T_w - T_{gi}\right).A_b \tag{18}$$

The terms '$\alpha_g^A$ and $h_{t-wg}$' are the fractional solar density taken by the top glass and net thermal transmission coefficient (water to glass), respectively (Cooper, 1973; Dunkle, 1961; Singh & Samsher, 2021a).

(c) Basin water equilibrium equation

By ingesting attenuated sunlight energy through convection, conduction form EATC and liner, the basin water column raises its temperature through sensible heating in addition to heat losses from fluid surface to glass. As a response, thermal equilibrium relation may be defined as,

$$\alpha_w^A.I_s(t).A_b + \dot{Q}_{N-EATC} + h_{bw}.A_b.\left(T_b - T_w\right) = m_w C_w.\frac{dT_w}{dt} + h_{t-wg}.A_b.\left(T_w - T_{gi}\right) \tag{19}$$

The terms '$\alpha_w^A, R_w and R_g$' is the fractional solar density, water reflectivity, and glass reflectivity, respectively. Further, '$h_{bw}$' refers to heat transmit coefficient (convective) from liner to basin medium, and articulated as (Tiwari, 2014),

$$h_{bw} = \left(K_w/t_b\right)\left\{C\left(R_a\right)^{1/n}\right\} \tag{20}$$

The terms of '$C$' and '$n$' are applicable for hot horizontal plates in an upward direction under various test interpretations (Tiwari, 2014). The formulas in Table 2 are utilized to determine the derived expressions in the thermal model, and the respective thermal properties of water are treated by the shown empirical relations accordingly.

(d) Lining energy equilibrium equation

Basin underneath contains the liner, which is the final destination for solar fluxes to be absorbed with maximum possibility due to matt black paint having 97% absorption

**Table 2** Thermo-physical property of water subjective to temperature deviation (IAPWS, 1996; Pissavi, 1982; IAPWS, 2008; Koffi, 2008)

| Constraint | Empirical Formulae |
|---|---|
| Specific thermal competence (J/kg K) | $C_w = 4226 - 3.244T + 0.0575T^2 - 0.0002656T^3$ |
| Water density (kg/m$^3$) | $\rho = 1001 - 0.003417T^2 - 0.08832T$ |
| Heat conductivity (W/m K) | $K_w = 0.557 - 0.00000708T^2 + 0.002198T$ |
| Volumetric thermal extension (K$^{-1}$) | $\beta' = \left(0.3 + 0.116T - 0.0004T^2\right).10^{-4}$ |
| Viscosity (kinematic) (m$^2$/s) | $\nu = \left[(0.5155 + 0.0192T)^{-1} - 0.12\right].10^{-6}$ |
| Viscosity (dynamic) (N.s/m$^2$) | $\mu = 4.284.10^{-5} + \left\{0.517(T + 64.993)^2 - 91.296\right\}^{-1}$ |
| Vaporization latent heat (J/kg) | $L = 2.5.10^6 - 2.37.10^3T + 0.268T^2 - 8.103.10^{-3}T^3$ |

of solar irradiance. The liner transfers heat to the volume of water while causing minimal losses to the air, as shown by the thermal equilibrium equation.

$$\alpha_b^A.I_s(t).A_b - h_{bw}A_b\left(T_b - T_w\right) = h_{ba}A_b\left(T_b - T_a\right) \tag{21}$$

The terms '$\alpha_b^A, \& h_{ba}$' are the fractional solar density and net heat loss coefficient. Equation (21) can be rearranged for $T_b$ as follows,

$$T_b = \left\{\alpha_b^A.I_s(t) + h_{bw}T_w + h_{ba}T_a\right\}/\left(h_{bw} + h_{ba}\right) \tag{22}$$

The above equation can be represented in different forms also by subtracting $T_w$ to both the sides as,

$$h_{bw}\left(T_b - T_w\right) + U_{bwa}\left(T_w - T_a\right) = \alpha_{beff}I_s(t) \tag{23}$$

Following Eqs. (17) and (18), $T_{gi}$ can be expressed as

$$T_{gi} = \left(\alpha_g^A I_s(t)A_g + U_cA_gT_a + h_{t-wg}A_bT_w\right)/\left(U_cA_g + h_{t-wg}A_b\right) \tag{24}$$

Above equation can be represented in different forms also by subtracting $T_w$ to both the sides as,

$$h_{t-wg}\left(T_w - T_{gi}\right) + h_1'.\alpha_g^A.I_s(t) = U_{ta}\left(T_w - T_a\right) \tag{25}$$

The elaborative form of Eq. (19) can be given as,

$$\begin{aligned} m_wC_w.\frac{dT_w}{dt} + \left\{U_{ta}\left(T_w - T_a\right)A_b - h_1'.\alpha_g^A.I_s(t).A_b\right\} &= \alpha_w^A.I_s(t).A_b + \left\{\alpha_{beff}I_s(t)A_b - U_{bwa}\left(T_w - T_a\right)A_b\right\} \\ &+ \left\{F_1(\propto\tau)_{eff}I_b(t)\left(A_{rc}.F_r\right) - \dot{m}_fC_f.T_{fi}(N-1) - \left(A_{rc}.F_r\right)U_L\left(T_{fi} - T_a\right)\right\} \end{aligned} \tag{26}$$

As the SDS water is the inlet for EATC, so above equation can be represented in different form also by putting $T_{fi} = T_w$ as,

$$\begin{aligned} &\frac{dT_w}{dt} + \frac{T_w}{m_wC_w}\left\{U_{bwa}A_b + U_{ta}A_b + \left(A_{rc}F_r\right)U_L + \dot{m}_fC_f(N-1)\right\} = \frac{1}{m_wC_w} \\ &\left[I_s(t).A_b\left(\alpha_w^A + \alpha_{beff} + h_1'.\alpha_g^A\right) + \left\{F_1(\propto\tau)_{eff}I_b(t)\left(A_{rc}.F_r\right)\right\} + T_a\left\{A_b(U_{bwa} + U_{ta}) + \left(A_{rc}F_r\right)U_L\right\}\right] \end{aligned} \tag{27}$$

The above equation can be given in terms of simple linear differentials as,

$$\frac{dT_w}{dt} + a.T_w = f(t) \tag{28}$$

The boundary states of affairs are at $t = 0, T_w = T_{wo}$

$$T_{wo} = \frac{f(t)}{a} + C \tag{29}$$

The resulting expression for $T_w$ is as given,

$$T_w = T_{wo}.e^{-at} + \frac{\bar{f}(t)}{a}\left(1 - e^{-at}\right) \tag{30}$$

(e) Mass Flow Examination of medium into EATC

The heat gain/m$^2$ ($\dot{q}$) of EATC water column be obtained by subsequent expressions as given,

$$\dot{q} = \dot{Q}_{N-\mathrm{EATC}} / \left(N.A_{rc}\right) \tag{31}$$

$$\mathrm{Re} = \frac{4\dot{m}_f}{\pi d.\mu} \tag{32}$$

The above equation can be reshuffled as,

$$\dot{m}_f = 1.57.(\mathrm{Re}.\mu.R_{i1}) \tag{33}$$

Following the expression for '*Re*' referring to Reynolds number for different EATC tubes for siphon circulation at different thermodynamic circumstances can be represented as (Budihardjo & Morrison, 2009),

$$\mathrm{Re} = 0.1914\left\{\frac{Nu.Gr}{Pr}\left(\frac{L_{EATC}}{2R_{i1}}\right)^{1.2}.\mathrm{Cos}\theta\right\}^{0.4084} \tag{34}$$

Now, the above equation can be written as,

$$\dot{m}_f = 0.6643\left\{\frac{\left(g.\beta'.\rho^2.\dot{q}\right)}{\left(C_f.\mu^{0.55}\right)}.\left(L_{\mathrm{EATC}}\right)^{1.2}.{R_{i1}}^{5.25}.\mathrm{Cos}\theta\right\}^{0.4084} \tag{35}$$

Finally, the resulting appearance of the above equation can be symbolized after following Eqs. (31) and (14) as,

$$\dot{m}_f = 1.52\left[\frac{\left[\left(A_{rc}.F_r\right)\left\{F_1(\propto\tau)_{eff}I_b(t) - U_L\left(T_{fi}-T_a\right)\right\} - \dot{m}_f C_f.T_{fi}(N-1)\right]}{N.A_{rc}}.\left(\frac{L_{EATC}}{d}\right)^{1.2}.\frac{g.\beta'.\rho^2.{R_{i1}}^{6.45}}{C_f.\mu^{0.55}}.\mathrm{Cos}\theta\right]^{0.41} \tag{36}$$

here, '$T_{fi} = T_w$' is applied in the above equation and utilized as contributing medium to the EATC tube.

## 4 Co-design performance observations

### 4.1 Yielding analysis

The potable water yielding is depending on the temperature divergence ($T_w - T_{gi}$) which can be written as follows,

$$\dot{m}_{\mathrm{ewg}} = \frac{\left\{h_{e-wg}.A_b(T_w - T_{gi})\right\}}{L} \times 3600 \tag{37}$$

The above equation for the daily yielding can be expressed as written,

$$\dot{M}_{\text{ewg}} = \sum_{t=1}^{t=24} \left[ \frac{\left\{ A_b \left( T_w - T_{gi} \right) . h_{e-wg} \right\}}{L} \right] \times 3600 \tag{38}$$

### 4.2 Energy analysis observations

The potable water-dependent hourly energy gain can be represented as

$$(E_e)_{\text{hr}} = (\dot{m}_{\text{ewg}} . L) \times \frac{1}{3600} \tag{39}$$

Similarly, the representation of daily energy gain as

$$(E_e)_{\text{daily}} = \frac{1}{3600} \cdot \sum_{t=1}^{t=24} (h_{e-wg} . A_b (T_w - T_{gi}) \tag{40}$$

Now, to represent energy efficiency, one can state as

$$\eta_i = \frac{h_{e-wg} . A_b \cdot \left[ \left\{ \frac{\bar{f}(t)}{a} (1 - e^{-at}) + T_{wo} . e^{-at} \right\} - T_{gi} \right]}{I_b(t) . A_a + I_s(t) . A_b} \times 100 \tag{41}$$

Now, its expanded form can be represented as

$$\eta_i = \left[ \frac{h_{e-wg} . A_b}{I_s(t) . A_b + I_b(t) . A_a} \right] . \left[ \left\{ T_{wo} . e^{-at} + \frac{\left[ \left\{ I_s(t) A_b . \alpha_{eff}^A + T_a . U_{eff} + F_1 . (\propto \tau)_{eff} . \left( A_{rc} F_r \right) I_b(t) \right\} / \left( m_w C_w \right) \right]}{\left[ \left\{ U_{eff} + \dot{m}_f C_f (N - 1) \right\} / \left( m_w C_w \right) \right]} . \left( 1 - e^{-at} \right) \right\} - T_{gi} \right] \times 100 \tag{42}$$

Further, the efficiency (hourly and daily energy) of the projected system can be articulated similarly,

$$\left( \eta_e \right)_{\text{hr}} = \frac{(E_e)_{\text{hr}}}{\left\{ I_s(t) . A_b + \dot{Q}_{N-\text{EATC}} \right\}} \times 100 \tag{43}$$

$$\left( \eta_e \right)_{\text{daily}} = \frac{(E_e)_{daily}}{\sum_{t=1}^{t=24} \left\{ I_s(t) . A_b + \dot{Q}_{N-\text{EATC}} \right\}} \times 100 \tag{44}$$

### 4.3 4.3 Exergy analytical observation

Following steady-state affairs of exergy equilibrium of the scheme that can be articulated as,

$$(\dot{E}_x)_{\text{in}} - (\dot{E}_x)_{\text{des}} = (\dot{E}_x)_{\text{out}} \tag{45}$$

where $(\dot{E}_x)_{\text{des}}$ refers to the energy (destructed) due to irreversibility and $(\dot{E}_x)_{\text{hr-out}}, (\dot{E}_x)_{\text{daily-out}}$ represents hourly and daily exergy outputs of the projected scheme, respectively, and further can be expressed by following Cengel and Boles (2013).

$$(\dot{E}_x)_{\text{hr-out}} = h_{e-wg}.A_b\left\{\left(T_w - T_{gi}\right) - T_a.ln\left(\frac{T_w}{T_{gi}}\right)\right\} \tag{46}$$

Moreover, the hourly exergy $\left(\dot{E}_x\right)_{\text{hr-in}}$ to the scheme that directly associates the solar energy $\left(\dot{E}_{x_{\text{sun}}}\right)_{\text{hr-in}}$ with incoming exergy $\left(\dot{E}_{x_{\text{EATC}}}\right)_{\text{hr-in}}$ through EATC array as followed by (Petela, 2003),

$$\left(\dot{E}_x\right)_{\text{hr-in}} - \left(\dot{E}_{x_{\text{EATC}}}\right)_{\text{hr-in}} = \left(\dot{E}_{x_{\text{sun}}}\right)_{\text{hr-in}} \tag{47}$$

$$\left(\dot{E}_{x_{\text{sun}}}\right)_{\text{hr-in}} = I_s(t).A_b\left\{1 + \frac{1}{3}\left(\frac{T_a}{T_s}\right)^4 - \frac{4}{3}\left(\frac{T_a}{T_s}\right)\right\} \tag{48}$$

Also, the incoming exergy $\left(\dot{E}_{x_{EATC}}\right)_{hr-in}$ through EATC array can be given as,

$$\left(\dot{E}_{x_{\text{EATC}}}\right)_{\text{hr-in}} = C_f\dot{m}_f\left\{\left(T_{N-fo} - T_{fi}\right) - T_a.\ln\left(\frac{T_{N-fo}}{T_{fi}}\right)\right\} \tag{49}$$

So, Eq. (47) can be given in hourly and daily exergy inputs as written,

$$(\dot{E}_x)_{\text{hr-in}} = \dot{m}_f C_f\left\{\left(T_{N-fo} - T_{fi}\right) - T_a.ln\left(\frac{T_{N-fo}}{T_{fi}}\right)\right\} + 0.933.I_s(t).A_b \tag{50}$$

$$\left(\dot{E}_x\right)_{\text{daily-in}} = \sum_{t=1}^{t=10}\left[\left(\dot{E}_{x_{\text{EATC}}}\right)_{\text{hr-in}} + 0.933.I_s(t).A_b\right] \tag{51}$$

And finally, hourly and daily exergy efficiency can be articulated as

$$\left(\eta_x\right)_{\text{hr}} = \frac{\left(\dot{E}_x\right)_{\text{hr-out}}}{\left[\left(\dot{E}_{x_{\text{EATC}}}\right)_{\text{hr-in}} + \left(\dot{E}_{x_{\text{sun}}}\right)_{\text{hr-in}}\right]} \times 100 \tag{52}$$

$$\left(\eta_x\right)_{\text{daily}} = \frac{\sum_{t=1}^{t=24}\left[h_{e-wg}.A_b\left\{\left(T_w - T_{gi}\right) - T_a.ln\left(\frac{T_w}{T_{gi}}\right)\right\}\right]}{\sum_{t=1}^{t=10}\left[\dot{m}_f C_f\left\{\left(T_{N-fo} - T_{fi}\right) - T_a.ln\left(\frac{T_{N-fo}}{T_{fi}}\right)\right\} + 0.933.I_s(t).A_b\right]} \times 100 \tag{53}$$

## 5 Energy-exergy-based matrix observations

### 5.1 Pay-off Period

The pay-off period corresponds to the compensated comparable values against energy and exergy for the devoted embodied energy by the structure and should be optimum for the betterment of the system. The pay-off time period $(EP)_T$ can be symbolized as expressed in terms of energy and exergy,

$$\left(EP_e\right)_T = E_{\text{in}}/\left(E_e\right)_{\text{yr}} \tag{54a}$$

$$\left(EP_x\right)_T = E_{\text{in}}/\left(E_x\right)_{yr-\text{out}} \tag{54b}$$

'$E_{\text{in}}$' shows the consumed embodied energy associated with the system from the beginning to the final preparation of components or the system (Tiwari et al., 2016). The bifurcated embodied energy of the proposed scheme is in Table 4. However, the comparative performance of different SDS systems has been presented in Table 3.

### 5.2 Pay-off part

The pay-off part ($(EP)_F$) of the model is the reciprocal of pay-off time period (energy-exergy) and should be tending to one for its betterment as mentioned; $(EP)_F \rightarrow 1, for (EP)_T = 1$; the system is measured to be meaningful.

### 5.3 Life cycle convertible efficiency

The convertible efficiency for life span regarding its usefulness aligned with the incident astral energy incident to the proposed arrangement with respect to the energy and exergy perceptions is termed as life conversion efficiency, and following Singh and Samsher (2021c) , it can be articulated as,

**Table 3** Proportional performance of different SDS systems in amalgamation with solar collectors

| Study references | Solar collector | Avg. Yielding (kg/m$^2$·h) | Working condition |
|---|---|---|---|
| Tiwari et al. (2003) | Water medium | 0.12 | Conventionally natural |
| Tiwari et al. (2003) | Water medium and FPC | 0.26 | Forced by hydraulic pump |
| Sampathkumar et al. (2013) | Water basin and heat pipe type EATC | 0.54 | Forced by hydraulic pump |
| Singh et al. (2013) | Water medium and EATC | 0.56 | Natural and thermo-siphon |
| Kumar et al. (2014) | Water medium and EATC | 0.58 | Forced by hydraulic pump |
| Proposed Scheme | Water medium, MCPC mounted EATC | 0.67 | Natural and thermo-siphon |

$$\left(\eta_e\right)_{\mathrm{LCC}} = \frac{n.\left(E_e\right)_{\mathrm{yearly}} - E_{\mathrm{in}}}{\left(E_e\right)_s} \tag{55a}$$

where '$\left(E_e\right)_s$' indicates the net occurrence of solar energy for 30yrs life span and articulated as,

$$\left(E_e\right)_s = n \sum_{i=1}^{i=365} \left\{I_s(t).A_b + I_b(t).A_a\right\} \tag{55b}$$

Similarly, the exergy efficiency and $\left(E_x\right)_s$ for the respective life span of the projected model can be articulated as,

$$\left(\eta_x\right)_{\mathrm{LCC}} = \frac{n.\left(E_e\right)_{\mathrm{yearly}} - E_{\mathrm{in}}}{\left(E_x\right)_s} \tag{55c}$$

$$\left(E_x\right)_s = 0.93.n. \sum_{i=1}^{i=365} \left\{A_a.I_b(t) + A_b.I_s(t)\right\} \tag{55d}$$

## 6 Economic analysis

### 6.1 Economics of projected model

The complete study is conducted all through the way of 30 years of the system's operation. Total cost annually (TAC) is the consequential capital value under the impact of fixed yearly cost (FAC). The yearly maintenance value (AMC) and salvage value per annum basis (ASC) are all regulating criteria in the present analytical method.

(a) Yield Cost

Further, the yield cost ($C$) and respective energy productions (energy-exergy) are based on present monetary observations and expressed by the following (Reddy et al., 2018; Singh & Samsher, 2021c),

$$C = \frac{(TAC)}{\left(M_{ewg}\right)_a} \tag{56a}$$

$$C_e = \frac{TAC}{\left(E_e\right)_{\mathrm{yearly}}} \tag{56b}$$

$$= \frac{TAC}{\left(E_x\right)_{\mathrm{yr-out}}} \tag{56c}$$

$$\mathrm{TAC} = \mathrm{AMC} + \mathrm{FAC} - \mathrm{ASC} \tag{56d}$$

$$\mathrm{FAC} = \mathrm{CRF} \times \mathrm{PCC}\ \mathrm{ASC} = \mathrm{SFF} \times S \tag{56e}$$

The typical interest rate for appreciation has been set at 10% during the whole approach to economic indicators, predicated on the drifting expansion of the public markets. Since the unit is self-sufficient, presently, 5% of FAC is being used in the research. Furthermore, the reserve or salvage fund (S) following the model's expected service term is calculated through 20% of the principal investment costs (PCC) that indicates individual component costs, manufacturing costs, installation costs, and commissioning costs. These can be expressed in the following way:

$$\mathrm{AMC} = (\mathrm{FAC}) \times 0.05 \tag{56f}$$

$$S = (\mathrm{SFF}) \times 0.2 \tag{56g}$$

$$\mathrm{SFF} = (i)/\{(1+i)^n - 1\};\ CRF = \{i(1+i)^n\}/\{(i+1)^n - 1\} \tag{56h}$$

The capital regain factor (CRF) refers a compounded appreciation to the respective system over a typical span of time for primary investment costs and total value known as fixed yearly cost that is necessary to carry on the model operating. Additionally, to keep the installation running effectively, maintenance costs are a requirement. Additionally, the model's salvaging outlay (S) at the time of termination generates extra returns and is determined by the graded to be sinked fund factors (SFF) useful to the articulated interest rate.

(b) System Production

The productivity per annum $\left(\eta_P\right)_{\text{yearly}}$ of the planned scheme can be articulated as (Benson, 1952; Cox, 1951),

$$\left(\eta_P\right)_{\text{yearly}} = \frac{\left[C_s \times \left(M_{\text{ewg}}\right)_{yr}\right]}{TAC} \times 100 \tag{57}$$

'$(C_s)$, and $\left(M_{ewg}\right)_{yearly}$' shows the selling worth of yield (0.07$/l) and respective production yield per annum for output commencing from the anticipated system.

c) Exergy-based Economic Study

The presentation of the concern can be expected in more suitable and cost effectual way through exergy-based economic investigation (Tsatsaronis & Park, 2002). Additionally, the exergy-based economic factor $\left(E_x\right)_{EF}$ represents the relation of yearly exergy output to the net yearly charge as offered,

$$\left(E_x\right)_{\text{EF}} = \left(E_x\right)_{yr-\text{out}}/\text{TAC} \tag{58}$$

## 6.2 Environmental economic study

(a) Pollutants equivalent emission

The pollutant ($CO_2$: $(\text{EEq})_C$, $SO_2$: $(\text{EEq})_S$, and NO: $(\text{EEq})_N$) equivalent emission for the respective system can be estimated by the following (Reddy et al., 2018; Singh & Gautam, 2022),

$$(\mathrm{EEq})_C = E_{\mathrm{in}} \times 1.58 \tag{59a}$$

$$(\mathrm{EEq})_S = E_{\mathrm{in}} \times 0.012 \tag{59b}$$

$$(\mathrm{EEq})_N = E_{\mathrm{in}} \times 0.005 \tag{59c}$$

(b) Pollutant mitigation

The following pollutant mitigates from the projected model during its entire working life, one can express the mitigate amount produced based on carbon credits under the influence of energetic and exergetic perspectives, respectively, as (Mittal et al., 2014; Singh & Samsher, 2021c; Singh & Gautam, 2022),

$$\left(\mathrm{MEq}_e\right)_C = 1.58\left\{\left(E_e\right)_{yr} \times n - E_{in}\right\} \tag{60a}$$

$$\left(\mathrm{MEq}_e\right)_S = 0.012\left\{\left(E_e\right)_{yr} \times n - E_{in}\right\} \tag{60b}$$

$$\left(\mathrm{MEq}_e\right)_N = 0.005\left\{\left(E_e\right)_{yr} \times n - E_{in}\right\} \tag{60c}$$

(c) Carbon credit returns

The dematerialized form of amount through carbon credit returns in the global market based on the standard pricing system is recovered yet (Roome, 2019; Singh & Samsher, 2021c). Accordingly, the energy, exergy-based ecological outlay may be represented as $\left(E_{ce}\right)_C$ and $\left(E_{\mathrm{cx}}\right)_C$ depending on the carbon mitigates of the suggested system throughout its life as expressed here,

$$\left(E_{\mathrm{ce}}\right)_C = 9.99\left[1.58\left\{\left(E_e\right)_{\mathrm{yr}} \times n - E_{\mathrm{in}}\right\}\right] \tag{61a}$$

$$\left(E_{\mathrm{cx}}\right)_C = 9.99\left[1.58\left\{\left(E_x\right)_{\mathrm{yr-out}} \times n - E_{\mathrm{in}}\right\}\right] \tag{61b}$$

## 7 Methodology

The methodological steps of the presented thermal analysis, result discussion, and conclusions have been made sequentially, as mentioned here:

*First:* The first step involved the conversion of received solar radiation over a flat parallel to a horizontal surface (ground) into the inclined response of those solar radiations (Liu & Jordan, 1960) to the SDS system as well as EATC array. The raw irradiance data have been procured from IMD, Pune meteorological department of India.

*Second:* With the help of MATLAB analytical tool, the unknown parameters have been revealed with reference to the known and available data, such as intensive solar radiation, different constants, and standard values accordingly through the derived formulations into the thermal model by following the equations such as Eq. (24) for $T_{gi}$, Eq. (30) for $T_w$, Eq. (37) for potable water yields, and Eq. (36) for mass circulation rate through thermo-siphon correspondingly.

*Third:* Consequently, the energy-exergy output on an hourly basis and corresponding efficiencies are analyzed with the help of derived equations in the thermal model by following Eq. (39), Eq. (51), and Eq. (42), respectively. Further, the daily-based outputs have been measured accordingly on energy-exergy basis by tracking the results for 24 h of the particular archetypal clear day observations, respectively.

*Fourth:* Finally, the energy-exergy matrices, environmental revenue, etc., have been analyzed by following the derived expressions as given in Eqs. (54)–(55), Eq. (61), respectively. And the analyzed observations have been further discussed, compared with the earlier researches, and concluded accordingly with fare sense of motivation.

## 8 Result and discussion

Figure 5 confirms the hourly distinctions of diverse solar responsive elements (i.e., diffuse, beam, global, and reflected radiations) over the horizontal plane (source radiation data and ambient temperature received from IMD Pune), which have been converted for the

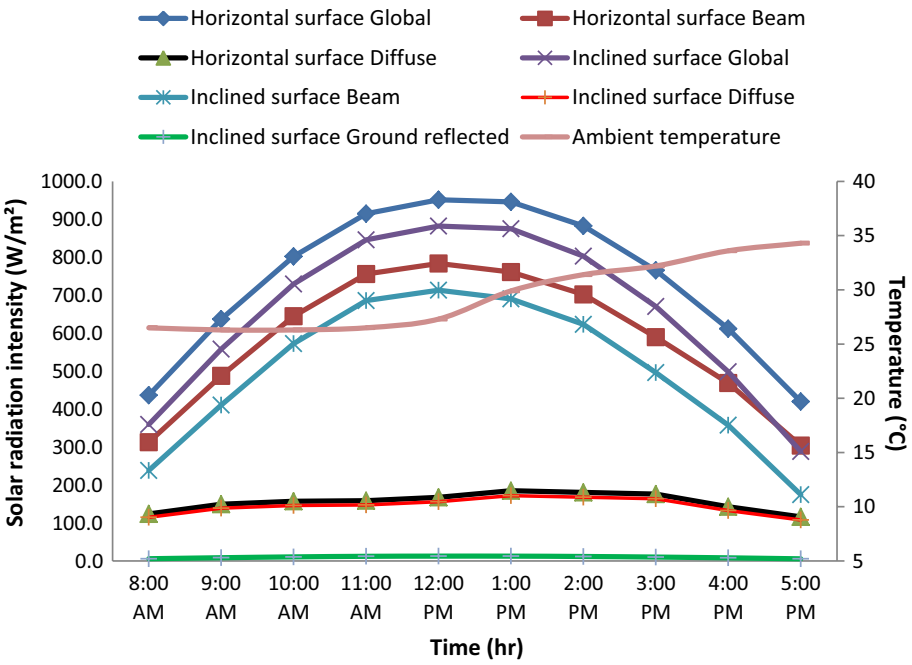



**Fig. 5** Hourly solar irradiative representation along with air movement and atmospheric temperature

equivalent response over inclined SDS and EATC surfaces for the particular apparent day. As the figure depicts the decrement in inclined solar radiation findings over the projected system; however, the inclined interfaces of solar recipients are at an angle of approximately equal to the latitude angle of respective location, which not only allows the direct incidence of solar radiation perpendicularly into the basin and EATC both with more than 95% intake value of the total incident radiations. Also, the air movement and corresponding temperature variations have been seen up to 4.11 m/s and 24 °C to 35 °C, respectively, that supports the yielding throughout the operation.

### 8.1 Thermodynamic model relative perception

The present model offers an original amalgamation of integration with different components (single sloped SDS with EATC and mounted MCPC) that has not been accounted in any research work; however, the validation up to SS-SDS system along with EATC without MCPC has been compared with similar work (Singh et al., 2013) for still's water temperature, basin cover glass temperature (inner), yielding, and mass flow circulations (thermosiphon) as portrayed in Figs. 6, 7. The realizations are quite similar where the MCPC and water level have little impact. The influence of MCPC accelerates the circulating rate of the thermal syphon cycle into the tube (particularly at elevated solar times), and as a response, basin fluid heats up fast, and increased water level (i.e., 0.16 m improves sensible heating capacity better) greatly promotes these shifts. The proposed methodology has successful outcomes, with a 315% yield increment by rise in water level up to 433%, and corresponding temperature increment in basin water up to 10.8%, also 28% in water circulation. The present system derived thermal paradigm validates its reliability and may be used to estimate system's performance for the given mathematical constraints of the model, as evidenced by the above validated observations of the findings.

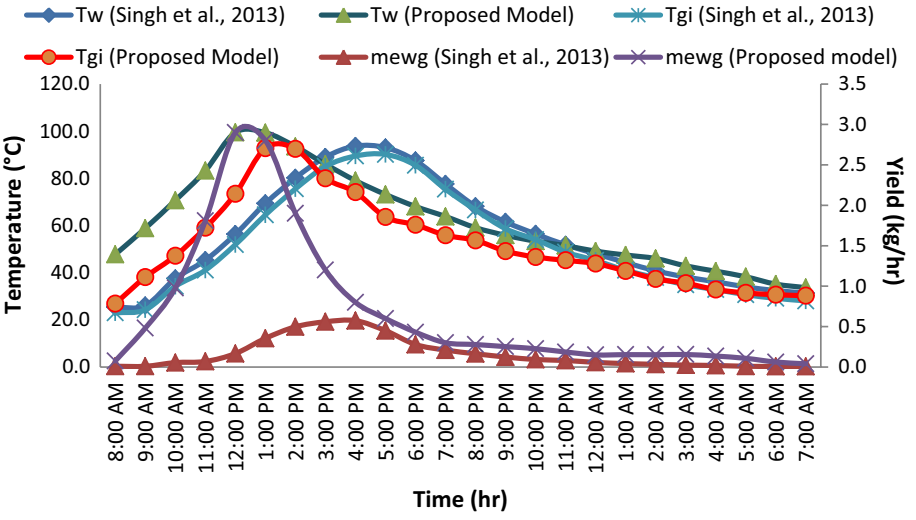



**Fig. 6** Observation of thermal models results with pre and present systems

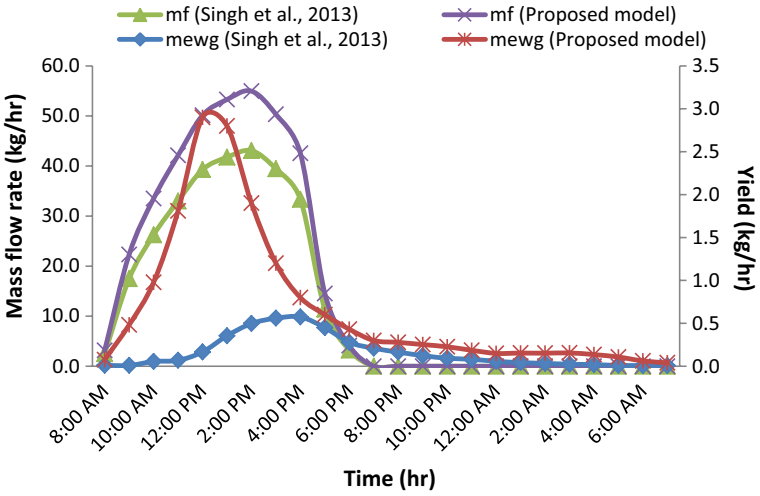



**Fig. 7** Hourly disparity in the water stream speed and yield creation

## 8.2 Projected structure optimization

Owed to the constraints of thermodynamic modeling that doesn't pertain to heat up the water to boiling level or more, so the present model fixes the optimal number of EATC array up to its best possible level. As a result, the assessment is conducted out for a diverse number of tubing along with different water levels in SDS under the prescribed plan parameters, and it finalizes the 4 numbered EATC tube with 0.16 m brackish water thickness achieves 99.6 °C optimal hotness into the basin, as shown by Fig. 8. Considering, Figs. 9a,b validates the general design criteria that are taken into account for the proposed methodology and also referred to be preferable to the other correlative configurations. The proper use of the elements allows for the still water temperature as 99.6 °C and corresponding better yield as compared to other models. Further, as sound as a considerable quantity of water is there for collecting a superior amount of heat (sensible form) and preserving increased temperature for a prolonged duration, which is good for nighttime yielding.

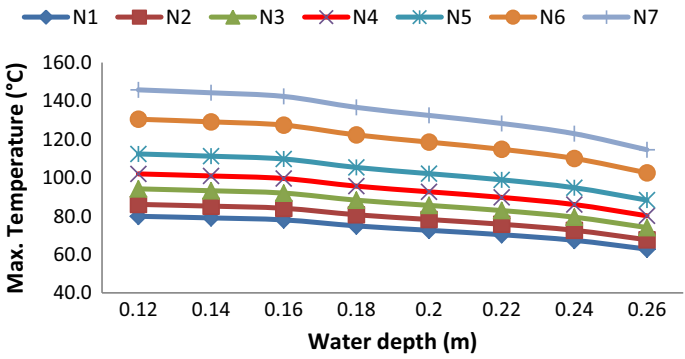
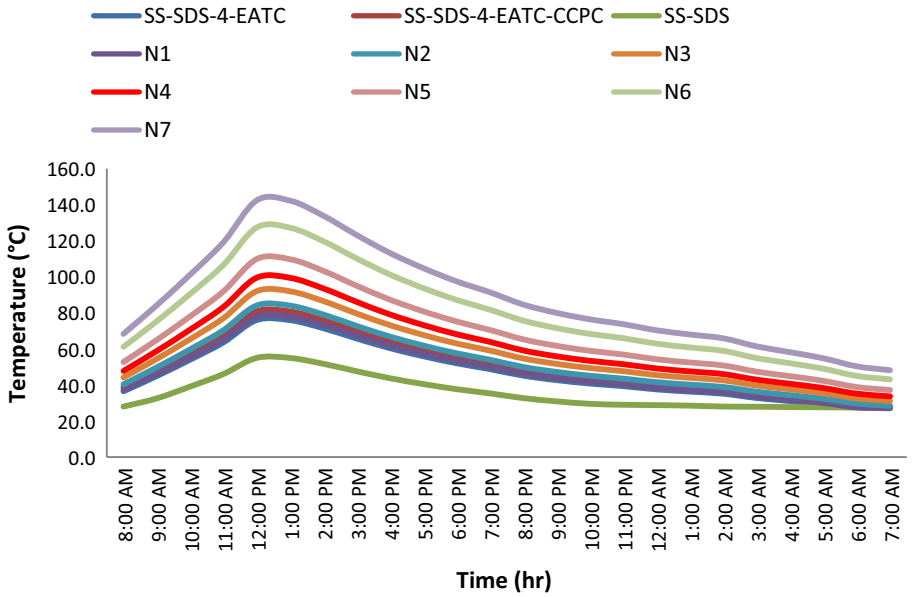
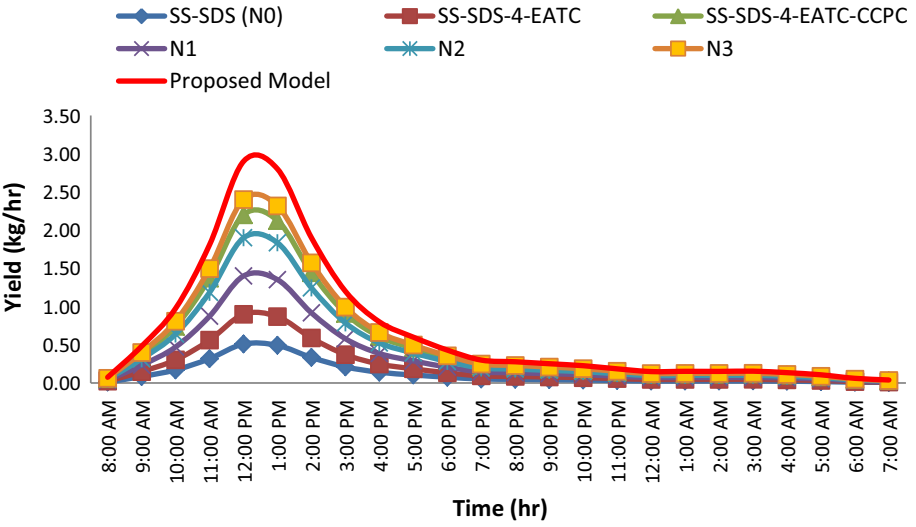



**Fig. 8** Maximum temperature gain of basin water with respect to variable water levels under the definite EATC numbers




**(a)** Corresponding temperature rise in basin water with respect to definite water level under the diverse EATC numbers.




**(b)** Comparative yielding of different models under similar design and work conditions with respect to time.

**Fig. 9** *a* Corresponding temperature rise in basin water with respect to definite water level under the diverse EATC numbers. *b* Comparative yielding of different models under similar design and work conditions with respect to time

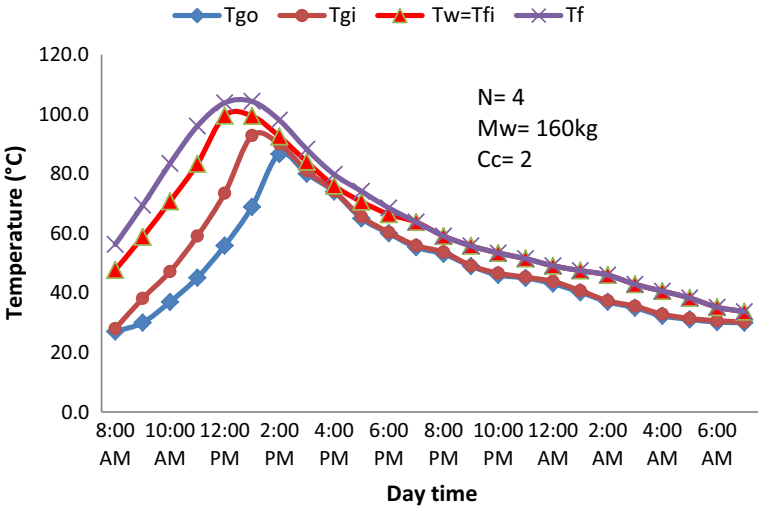



**Fig. 10** Temperature disparity in the model under the given constraints of the projected model

### 8.3 Energetic exergetic performance analysis

For a respective clear day in the summer of June, the presented Fig. 10 exhibits temperature change in various regions of still scheme. The EATC fluid heat appropriately regulates the temperature in the sink, with significant growth (1 p.m.) and wider heat disparities, foremost to a high stream pace. The thermal syphon movements into the tube are articulated by the functional thermal gradient, which is reachable up to evening around 7 p.m. Also, a momentous temperature divergence ($T_w - T_{gi}$) becomes visible (2 p.m.) then ostensible temperature variations are preserved for a prolonged time during the day, which supplies day-night yields correspondingly. In count, Fig. 11 portrayed the various vacillations in

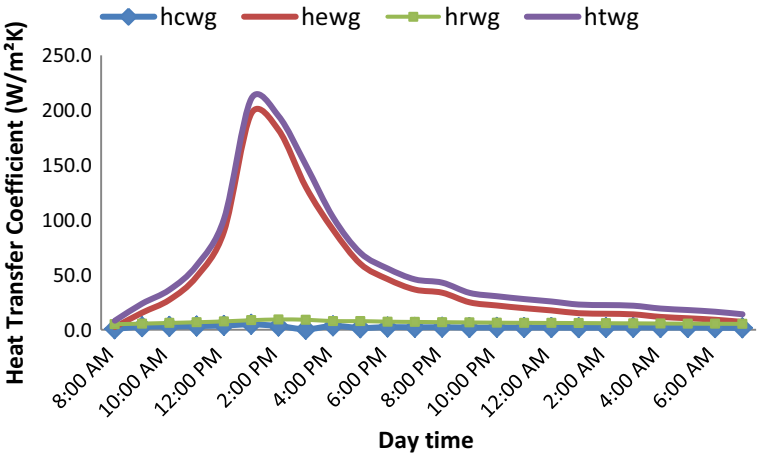



**Fig. 11** Coefficients of heat transfer deviations in the model in the summer of June at $N=4$

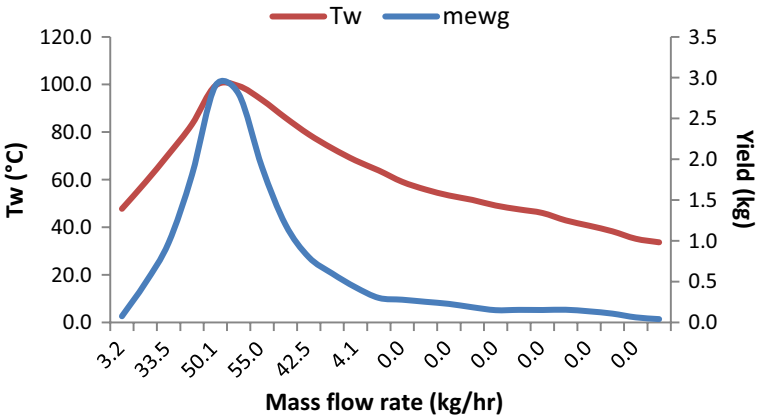



**Fig. 12** Yielding at different temperatures and equivalent flow rate for June month operating conditions

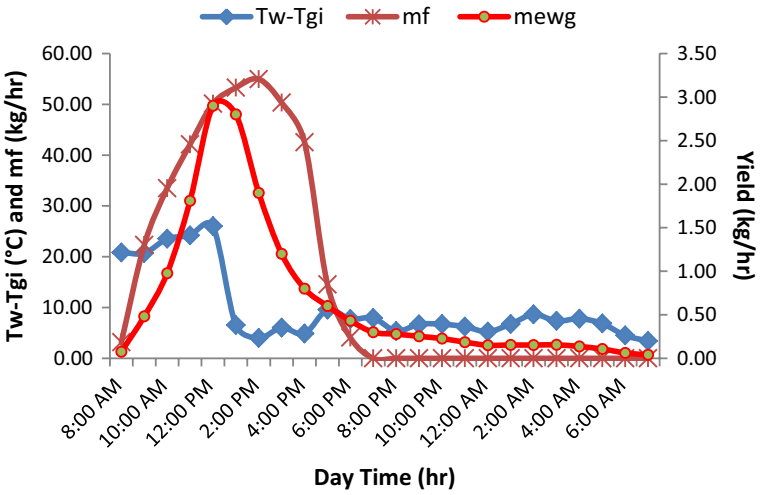
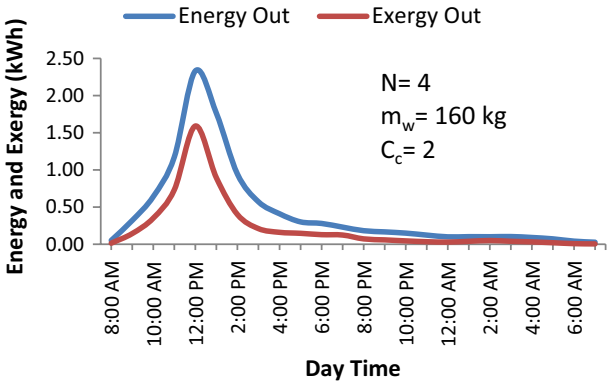



**Fig. 13** Mass flow, yield, and $(T_w - T_{gi})$ representation in the summer month of June, archetypal sunny day

coefficient of heat transfer of still model. It has shown the evaporative mode of heat transmission as an important measure that heats the basin glass cover correspondingly, whereas the irradiative mode impacts less due to its lower temperature; however, it works better for more than 1000 °C. The most notable gains are visible in the mid of the selected day, while the sun strength is the premier.

The creation of potable water under the influence of respective thermo-syphon at different water temperatures, i.e., $T_w$ and differences, i.e., $T_w - T_{gi}$ for the whole day as made known in Figs. 12, 13. The highest capitulate (2.9 kg) offered by the system at nearly 1 p.m. at $T_w - T_{gi}$ (26.02 °C), and the rate of circulation is possibly higher (55 kg/hr) at about 2 p.m. because of maximum $T_w - T_{gi}$ is attained that increases water temperature faster, which leads to subsequent evaporation and hence condensation. So, the average output of



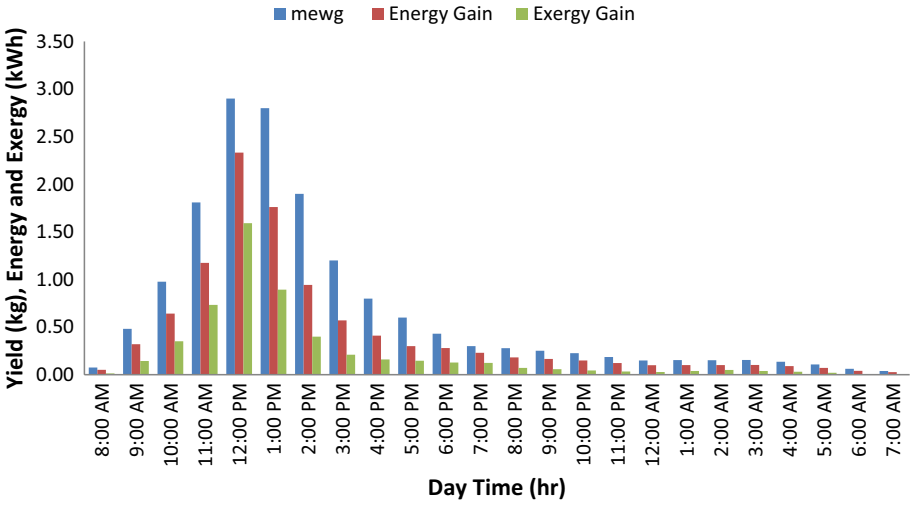




**Fig. 14** Energy-exergy gain of the model for the suggested clear day circumstances in the summer month of June

**Fig. 15** Yielding on hourly basis as per energy and exergy gains of the proposed model

yield appeared better (0.67 kg/m$^2$) for the proposed system in contrast to other still models for similar working backgrounds, as depicted in Table 3.

Figures 14, 15 correspond to the energy-exergy-based gains for the selected sunny clear day working conditions with optimal results, i.e., 2.33 kWh and 1.59 kWh, in that order which appeared at the climax radiation hours and followed by the modeled Eqs. (40) and (46) prove the above results. Also, the sensitivity of such efficiencies is presented in Fig. 16, which is a cumulative effect of SDS and EATC both. These efficiencies depend on various temperature distinctions like ($T_{N-fo} - T_{fi}$), ($T_{N-fo} - T_w$), and ($T_w - T_{gi}$) further on incoming radiations, energy-exergy transfer, etc., as demonstrated in Eq. (44). It is also seen with the decrease in temperature difference efficiency reflects high (7.75%-94.63%), especially in the noontime. On the contrary, efficiencies further increase in the evening also due to lowered temperature differences and solar radiation accordingly.

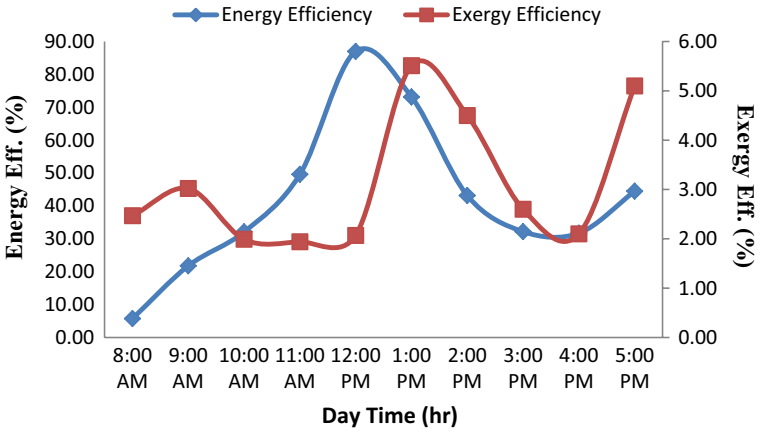



**Fig. 16** Representation of efficiencies (energy-exergy)

**Table 4** Component utilization to construct the proposed arrangement as per present market standards and energy density

| Model components | Measurement | Price ($) | $E_{in}$(kWh) |
|---|---|---|---|
| Glass | 1.16 kg, 4 mm | 12.56 | 51.6 |
| Basin material (FRP) | 21.17 kg | 140.9 | 542.8 |
| EATC | 4 Nos | 21.16 | 106.84 |
| MS stand | 20 kg | 17 | 190 |
| MS clamp | 2 kg | 1.7 | 19 |
| MS frame | 5 kg | 4.3 | 47.5 |
| Nozzle | 0.100 kg | 2.2 | 2.13 |
| Liner | 250 ml | 2 | 6.3 |
| Gasket | 98.9 m, 2.1 kg | 3 | 69.09 |
| Reflector (parabola) | 2.8 kg | 10.1 | 168.5 |
| Net Value | | 214.92 | 1203.76 |

## 8.4 Energy-based matrix analysis

Table 4 illustrates several components, including their parameters, embodied energy, and price of the stock, which are recommended to estimate various appearances of matrices (energy-exergy). Table 5 demonstrates the fallout of the recommended system energy-based matrices, particularly for attending to 275 clear days for each year up to 30 years of service with an average sun's ray potency of 401.8 kW (averaged intensity).

Figure 17 indicates the project model matrices analysis (energy-exergy) over a 30yrs operating life. Its energy recovery term is 0.42 years, while its exergy payoff term is 0.82yrs due to the corresponding exergy destructions linked with the system as well, so the respective exergy payoff period goes longer. The related payoff ratios (energy and exergy) are 2.36, and 1.22, respectfully, which reflects the still model capacity and abilities, and the larger payoff value indicates the system's quality correspondingly. These outputs are primarily controlled by the state's energy of embodying and

**Table 5** Energy-exergy matrices with related results of the planned solar still scheme ($n = 30$yrs.)

| Constraint | | Result observation |
|---|---|---|
| Yearly yielding | | 4455 kg |
| Daily yielding | | 16.2 kg/day |
| Life yielding | | 133,650 kg |
| System existence | | 30 yrs |
| Embodied input | | 1203.76 kWh |
| Average each day sun's ray strength | | 401.8 kW |
| Constraint | Exergy | Energy |
| Overall life energy-out | 39,847.5 kWh | 85,222.5 kWh |
| Existence instant astral energy | 138,011.7 kWh | 147,922.5 kWh |
| Annually solar energy | 4600.4 kWh | 4930.8 kWh |
| Annual energy exposed from still | 1328.3 kWh | 2840.8 kWh |
| Pay off Period | 0.82 yrs | 0.42 yrs |
| Pay off part | 1.22 | 2.36 |
| Efficiency (Conversion Life Succession) | 0.31 | 0.57 |

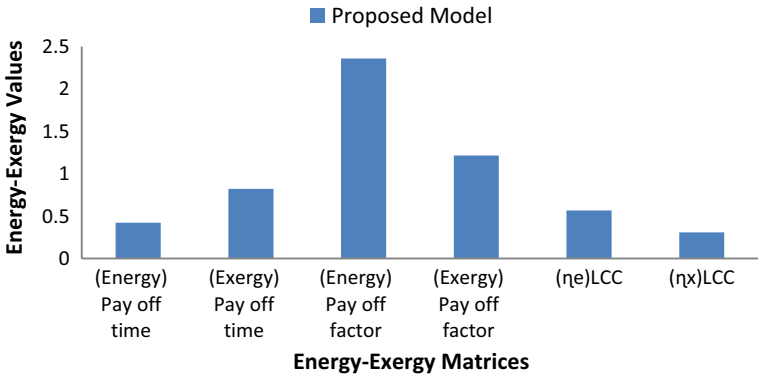



**Fig. 17** Energy-exergy matrices division of the anticipated representation ($n = 30$yrs.)

yearly yielding (energy/exergy), which is portrayed as a longer payoff time or factor (energy/exergy) for more yearly production or a decreased energy of embodying for the model. The efficiency of life conversion (energetic/exergetic) of the model is represented in Fig. 18, which reflects its goodness for the respective efficiencies (i.e., $\rightarrow 1$), as in the current plan. The conversion efficiency of the system gets better with the length of time, as seen by 0.57, 0.31 for energy, exergy, respectively, over the first 30 years of operation.

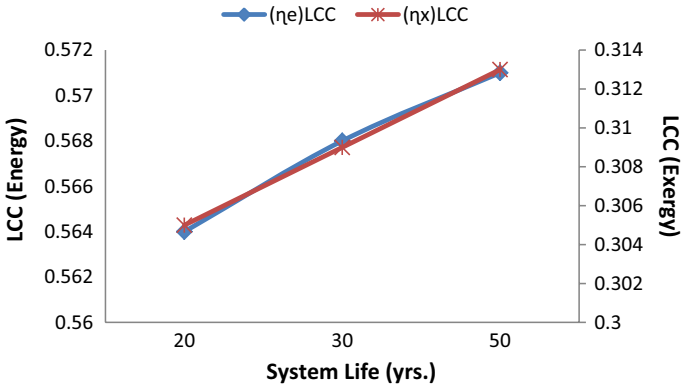



**Fig. 18** Depiction of conversion efficiency for specified period of time of the system

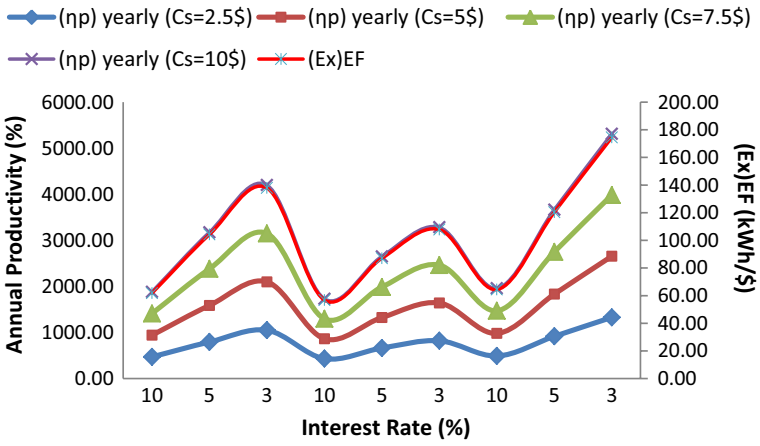



**Fig. 19** Annual productivity and exergetic economy under the respective interest rates

### 8.5 Requisites for economic considerations under environmental perspective

The yearly productivity and exergetic economy vacillations of the current plan in the impact of the changeable pace of interest and yielding sale value are depicted in Fig. 19. As inflation rises, the denominator quantity in the succeeding Eqs. (57–58) rises, lowering the entire output and exergetic economy of the scheme. The method is significantly practicable for its long term survival in the competitive industry for the provision of treated water since the yearly yield for studied variables is larger than 100%. Intended for a 10% inflated pace and a 0.07$/l sales worth of treated water, the figures for productivity and monetary variables are 940.8% & 61.8%, respectfully.

Furthermore, the new plan is extremely efficient for mitigating $CO_2$, NO, $SO_2$ due to substantial yielding throughout the operational life, with mitigates (energy-exergy),

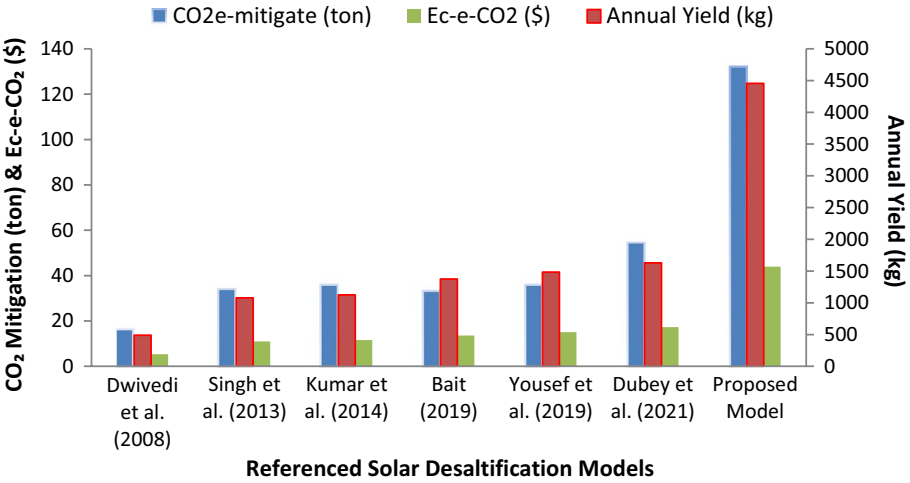



**Fig. 20** Distinctive observations different solar still systems under the yielding and environomical perspectives

i.e., 131.97, 64.44 tones for $CO_2$, 1.00, 0.51 tones for $SO_2$, 0.42, 0.21 tones for NO, in that order. As well as, energetic-exergetic environmental cost generated in terms of carbon credit income is $1318.4 & $673.8, correspondingly. Figure 20 shows a relative observation of different still systems in terms of yield production and environomic. To evaluate the performances on a comparable basis of multiple systems under the same life span (i.e., 30 years) is 1.58 kg which represents mitigation of $CO_2$, and corresponding credits are 9.99$/ton. Hence, the existing topological relationship is finer to the additional models deliberated.

## 9 Conclusion and future suggestions

The present analysis of the thermodynamic model represents the essential requirement of MCPC to enhance the EATC tube to improve the SS-SDS water treatment system in its optimum capacity in comparison with the existing studies of the respective model. Further, the focused conscious conclusions have been framed accordingly as follows:

i. The combination of MCPC-EATC with SS-SDS works well and produces 99.5 °C maximum water temperature during the mid of the day. Larger amount of water with a novel association and concentration ratio of 2 enhanced the water temperature (11.4%) significantly and improved the daily output by 12.3 kg per day than the typical SDS-EATC system taken for comparison into consideration. And, the system efficiencies are 3.5% and 50.8% for exergy, energy, respectively.
ii. The energetic-exergetic yielding prices are relatively noticeable, i.e., 0.008–0.016 $/kWh, respectively, for the considered life (30 years). Energy matrices referred to remarkable results with the payback of less than one year time and have better conversion efficiency as 0.57 and 0.31 on an energy and exergy basis, in that order.

iii. $CO_2$ emissions (energy-exergy) are quite effective as 131.97–64.44 tones, respectively, with better earned credits (energy-exergy) as $1318.4-$673.8, respectively. So, the system is substantially practical for its future survival in the bazaar place to generate drinkable water.

iv. As the suggested model produces excessive heat due to the optimum association of EATC tubes during the peak hours of sun's ray incidence, which is not effectively synchronized with the yield generation, so one can utilize additional condensing chambers, etc. to cope up the left part of vapors for the further noticeable yielding purpose that may also be validated by respective experimentation of the suggested model.

## Appendix

Relationships employed to resolve Eq. (1), (2) (Duffie & Beckman, 2006; Singh & Gautam, 2022)

$$U_{saa} = \left( \frac{R_{o2} \ln\left(\frac{R_{i2}}{R_{i1}}\right)}{K_g} + \frac{1}{h_{r-v}} + \frac{R_{o2} \ln\left(\frac{R_{o2}}{R_{o1}}\right)}{K_g} + \frac{1}{h_o} \right)^{-1}$$

$$I_s(a) = I_b(t).(\propto \tau)_{eff}$$

$$(\propto \tau)_{eff} = R_{sc} \alpha \tau^2 \left(A_a / A_{rc}\right)$$

$$h_o = h_{c-wg} + h_{r-wg} = 3.8V + 5.7$$

$$h_{r-v} = \varepsilon_{eff}.\sigma \left[ \left(T_f + 273.15\right)^2 + \left(T_{sa} + 273.15\right)^2 \right] \times \left[ T_f + T_{sa} + 546.30 \right]$$

$$\varepsilon_{\text{eff}} = \left(1/\varepsilon_g + 1/\varepsilon_w - 1\right)^{-1}$$

Terminologies utilized in Eqs. (7), (10), and (17) as stated,

$$F_1 = \frac{F'.h_{\text{saf}}}{\left(F' h_{\text{saf}} + U_{\text{saa}}\right)}$$

$$U_L = \frac{F'.h_{\text{saf}}.U_{\text{saa}}}{(F' h_{\text{saf}} + U_{\text{saa}})}$$

$$F_2 = 1 - \frac{\left(A_{rc}.F_r\right)U_L}{\dot{m}_f C_f}$$

$$F_r = \frac{\dot{m}_f C_f}{A_{rc}.U_L}\left\{1 - \exp\left(-\frac{2\pi R_{i1} L_{\mathrm{EATC}}.U_L}{\dot{m}_f C_f}\right)\right\}$$

$$h_{t-g} = 5.7 + 3.8V$$

$$U_c = h_{t-g}/\left(1 + \frac{h_{t-g}}{K_g/t_g}\right)$$

Relationships utilized for Eq. (18) as expressed (Cooper, 1973; Dunkle, 1961; Singh & Gautam, 2022; Singh & Samsher, 2020),

$$\alpha_g^A = \left(1 - R_g\right)\alpha_g$$

$$R_g = 1 - \frac{4n_o n_g}{\left(n_o + n_g^2\right)(1 + n_o)}$$

$$h_{t-wg} = h_{r-wg} + h_{c-wg} + h_{e-wg}$$

$$h_{r-wg} = \varepsilon_{\mathrm{eff}}.\sigma\left\{\left(T_w + 273.15\right)^2 + \left(T_{gi} + 273.15\right)^2\right\}\left\{T_w + T_{gi} + 546.30\right\}$$

$$h_{c-wg} = 0.884\left\{\left(T_w - T_{gi}\right) + \frac{\left(P_w - P_{gi}\right)\left(T_w + 273.15\right)}{268.9 \times 10^3 - P_w}\right\}^{1/3}$$

$$h_{e-wg} = 16.273 \times 10^{-3} h_{c-wg}\left\{\frac{P_w - P_{gi}}{T_w - T_{gi}}\right\}$$

$$P_w = \exp\left(25.32 - \frac{5144}{T_w + 273.15}\right)$$

$$P_{gi} = \exp\left(25.32 - \frac{5144}{T_{gi} + 273.15}\right)$$

Equations used in Eq. (20) as,

$$\alpha_w^A = \left\{\left(1 - R_g\right)\left(1 - \alpha_g\right)(1 - R_w)\alpha_w\right\}$$

$$R_w = \left[1 - \left(4n_o n_w\right)/\left\{\left(n_o + n_w^2\right)(1 + n_o)\right\}\right]$$

$$R_a = \frac{g\beta' \rho^2 X^3 C_w \Delta T}{\mu K_w}$$

$$G_r = \frac{g\beta' \rho^2 X^3 \Delta T}{\mu^2}$$

$$P_r = \frac{\mu C_w}{K_w}$$

Equations used in Eq. (22) as written (Tiwari, 2014),

$$h_{ba} = \left\{ \left( t_b / K_b \right) + 0.357 \right\}^{-1}$$

$$\alpha_{beff} = \left( \alpha_b^A . h_{bw} \right) / \left( h_{ba} + h_{bw} \right)$$

$$U_{bwa} = \left( h_{bw} . h_{ba} \right) / \left( h_{bw} + h_{ba} \right)$$

Relationships involved in Eqs. (25), (28), and (34) as,

$$U_{ta} = \left( h_{t-wg} . A_g . U_c \right) / \left( A_b . h_{t-wg} + U_c . A_g \right)$$

$$h_1' = \left( A_g . h_{t-wg} \right) / \left( h_{t-wg} . A_b + A_g . U_c \right)$$

$$a = \left\{ U_{eff} + \dot{m}_f C_f (N-1) \right\} / \left( m_w C_w \right)$$

$$U_{\text{eff}} = \left\{ \left( U_{ta} + U_{bwa} \right) A_b + \left( A_{rc} F_r \right) U_L \right\}$$

$$\alpha_{\text{eff}}^A = \left( \alpha_w^A + \alpha_{beff} + h_1' . \alpha_g^A \right)$$

$$f(t) = \left\{ I_s(t) A_b . \alpha_{eff}^A + T_a . U_{eff} + F_1 . (\propto \tau)_{eff} . \left( A_{rc} F_r \right) I_b(t) \right\} / \left( m_w C_w \right)$$

$$(Gr.Nu)/Pr = \left( g . d^4 . \dot{q} . \beta' \right) / \left( \nu^2 . C_f . \mu \right)$$

$$Nu = \frac{h.X}{K_w}$$

$$\nu = \mu / \rho$$

**Author contributions** AKS: Conceptualization, Data curation, Formal analysis, Validation, Investigation, Methodology, Software, Writing—original draft. SG: Supervision, Resources, Visualization, Project administration.

**Funding** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Data availability** The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

**Ethical approval** Not applicable.

**Consent to participate** Not applicable.

**Consent to publish** Not applicable.

## References


Alayi, R., Khalilpoor, N., Heshmati, S., Najafi, A., & Issakhov, A. (2021). Thermal and environmental analysis solar water heater system for residential buildings. *International Journal of Photoenergy, 6838138*, 1–9.

Bait, O. (2019). Exergy, environeeconomic and economic analyses of a tubular solar water heater assisted solar still. *Journal of Cleaner Production, 212*, 630–646.

Benson, F. (1952). Further notes on the productivity of machines requiring attention at random intervals. *Journal of the Royal Statistical Society: Series B (Methodological)*, *XIV*, 200–210.

Budihardjo, I., & Morrison, G. L. (2009). Performance of water-in-glass evacuated tube solar water heaters. *Solar Energy, 83*, 49–56.

Budihardjo, I., Morrison, G. L., & Behnia, M. (2007). Natural circulation flow through water-in- lass evacuated tube solar collectors. *Solar Energy, 81*, 1460–1472.

Cengel, X. A., & Boles, M. A. (2013). *Thermodynamics, an engineering approach*. McGraw-Hill Education Pvt. Ltd.

Cooper, P. I. (1973). The maximum efficiency of single-effect solar stills. *Solar Energy, 15*, 205–217.

Cox, D. R. (1951). The productivity of machines requiring attention at random intervals. *Journal of the Royal Statistical Society: Series B (Methodological) XIII*, 65–82.

Dev, R., & Tiwari, G. N. (2012). Annual performance of evacuated tubular collector integrated solar still. *Desalination Water Treatment, 41*, 204–223.

Duffie, J. A., & Beckman, W. A. (2006). *Solar engineering of thermal processes*. Hoboken- Wiley.

Dunkle, R. V. (1961). Solar water distillation: the roof type solar still and a multiple effect diffusion still, International Developments in Heat Transfer ASME 895–902.

IAPWS (1996). International association for the properties of water and steam, for the thermodynamic properties of ordinary water substance for general and scientific use, 1996.

IAPWS (2008). International association for the properties of water and steam, the viscosity of ordinary water substance, 2008.

Issa, R. J., & Chang, B. (2017). Performance study on evacuated tubular collector coupled solar stillin west texas climate. *International Journal of Green Energy*. https://doi.org/10.1080/15435075.2017.1328422

Koffi, P. M. E., Andoh, H. Y., Gbaha, P., Toure, S., & Ado, G. (2008a). Theoretical and experimental study of solar water heater with internal exchanger using thermosiphon system. *Energy Conversion and Management, 49*, 2279–2290.

Koffi, P. M. E., Andoh, H. Y., Gbaha, P., Toure, S., & Ado, G. (2008b). Theoretical and experimental study of solar water heater with internal exchanger using thermosiphon system, Energy Convers. *Manage, 49*, 2279–2290.

Kumar, S., Dubey, A., & Tiwari, G. (2014). A solar still augmented with an evacuated tube collector in forced mode. *Desalination, 347*, 15–24.

Liu, B. Y. H., Jordan, R. C., (1960). The interrelationship and characteristic distribution of direct, diffuse and total solar radiation, *Solar Energy, 4*(3).

Malik, M. A. S., Tiwari, G. N., Kumar, A., & Sodha, M. S. (1982). *Solar distillation—a practical study of a wide range of stills and their optimum design, construction and performance*. Pergamon Press.

Mittal, M. L., Sharma, C., & Singh, R. (2014). Decadal emission estimates of carbon dioxide, sulphur dioxide and nitic oxide emissions from coal burning in electric power generation plants in India. *Environmental Monitoring and Assessment, 186*, 6857–6866.

Morrison, G. L., Budihardjoa, I., & Behnia, M. (2005). Measurement and simulation of flow rate in a water-in-glass evacuated tube solar water heater. *Solar Energy, 78*, 257–267.

Patel, J., Markam, B. K., & Maiti, S. (2019). Potable water by solar thermal distillation in solar salt works and performance enhancement by integrating with evacuated tubes. *Solar Energy, 188*, 561–572.

Pissavi, P. (1982). Modelling of the dynamic behavior of a tank for solar storage with internal exchanger. *Rev´ Ge´n Thermique, 246–247*, 521–35.

Reddy, K. S., & Sharon, H. (2017). Energy–environment–economic investigations on evacuated active multiple stage series flow solar distillation unit for potable water production. *Energy Conversion and Management, 151*, 259–285.

Reddy, K. S., Sharon, H., Krithika, D., & Philip, L. (2018). Performance, water quality and enviro-economic investigations on solar distillation treatment of reverse osmosis reject and sewage water. *Solar Energy, 173*, 160–172.

Roome, J. (2019). *State and Trends of Carbon Pricing (2019)*. World Bank Group.

Sampathkumar, K., Arjunan, T., & Senthilkumar, P. (2013). The experimental investigation of a solar still coupled with an evacuated tube collector. *Energy Sources, Part A433A: Recovery, Utilization, and Environmental Effects+A465, 35*, 261–270.

Sato, A. I., Scalon, V. L., Padilha, A., (2012). Numerical analysis of a modified evacuated tubes solar collector. International Conference on Renewable Energies and Power Quality (ICREPQ'12).

Sharon, H., Reddy, K. S., Krithika, D., & Philip, L. (2017). Experimental performance investigation of tilted solar still with basin and wick for distillate quality and enviro-economic aspects. *Desalination, 410*, 30–54.

Singh, A. K., Samsher (2021c). Tech-en-econ-energy-exergy-matrix (T4EM) observations of evacuated solar tube collector augmented solar desaltification unit: A modified design loom, Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2021c.09.088

Singh, A. K. (2020). An inclusive study on new conceptual designs of passive solar desalting systems. *Heliyon, 7*, e05793.

Singh, A. K., & Gautam, S. (2022). Optimum techno-eco performance requisites for vacuum annulus tube collector-assisted double-slope solar desaltification unit integrated modified parabolic concentrator. *Environmental Science and Pollution Research*. https://doi.org/10.1007/s11356-021-18426-x

Singh, A. K., Samsher (2020). Analytical study of evacuated annulus tube collector assisted solar desaltification system: A review, *Solar Energy, 207*: 1404–1426.

Singh, A. K., Samsher (2021a). A review study of solar desalting units with evacuated tube collectors. *Journal of Cleaner Production 279*: 123542.

Singh, A. K., Samsher (2021b). Material conscious energy matrix and enviro-economic analysis of passive ETC solar still. *Materials Today: Proceedings*, *38*, 1–5.

Singh, A. K., Samsher (2022). Techno-environ-economic-energy-exergy-matrices performance analysis of evacuated annulus tube with modified parabolic concentrator assisted single slope solar desalination system. *Journal of Cleaner Production, 332*:129996.

Singh, A. K., Yadav, R. K., Mishra, D., Prasad, R., Gupta, L. K., & Kumar, P. (2020). Active solar distillation technology: A wide overview. *Desalination, 493*, 114652.

Singh, R. V., Kumar, S., Hasan, M., Khan, M. E., & Tiwari, G. (2013). Performance of a solar still integrated with evacuated tube collector in natural mode. *Desalination, 318*, 25–33.

Tiwari, G. N., Raj, K., Maheshwari, K. P., Sawhney, R. L. (1992). Recent advances in solar distillation. In: International Journal of Solar Energy and Energy Conversion. Wiley Eastern, New Delhi, pp. 32e149 (Chapter 2).

Tiwari, G., Shukla, S., & Singh, I. (2003). Computer modeling of passive/active solar stills by using inner glass temperature. *Desalination, 154*, 171e185.

Tiwari, G. N., Tiwari, A., Shyam (2016). *Handbook of Solar Energy Theory Analysis and Applications*, Springer Nature

Tiwari, A. K., & Tiwari, G. N. (2007). Annual performance analysis and thermal modelling of passive solar still for different inclination of condensing cover. *International Journal of Energy Research, 31*(4), 1358–1382.

Tiwari, G. N. (2014). *Solar Energy: Fundamentals, Design*. Narosa Publishing House.

Tiwari, G. N., Yadav, J. K., Singh, D. B., Al–Helal, I. M., Abdel–Ghany, A. M. (2015). Exergoeconomic and enviroeconomic analyses of partially covered photovoltaic flat plate collector active solar distillation system. *Desalination, 367*, 186–196.

Tsatsaronis, G., & Park, H. (2002). On avoidable and unavoidable exergy destructions and investment costs in thermal systems. *Energy Conversion and Management, 43*(9–12), 1259–1270.
Yari, M., Mazareh, A. E., & Mehr, A. S. (2016). A novel cogeneration system for sustainable water and power production by integration of a solar still and PV module. *Desalination, 398*, 1–11.
Yousef, Md. S., & Hassan, H. (2019). Assessment of different passive solar stills via exergoeconomic, exergoenvironmental, and exergoenviroeconomic approaches: A comparative study. *Solar Energy, 182*, 316–331.

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

Physica Scripta

**ACCEPTED MANUSCRIPT**

# Eco-friendly green synthesis of stable ZnO nanoparticles using citrus limon: x-ray diffraction analysis and optical properties

To cite this article before publication: Anamika Kem *et al* 2022 *Phys. Scr.* in press https://doi.org/10.1088/1402-4896/ac7efb

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is **© 2022 IOP Publishing Ltd**.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.
As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by-nc-nd/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 14.139.251.98 on 15/07/2022 at 10:23

# Eco-friendly Green Synthesis of stable ZnO nanoparticles using citrus limon: X-Ray Diffraction Analysis and Optical Properties

**Anamika Kem[a], Mohd Rehan Ansari[a], P. Prathap[b], M. Jayasimhadri[c] and Koteswara Rao Peta[a]***

*[a]Department of Electronic Science, University of Delhi South Campus, Benito Juarez Road, New Delhi-110021*
*[b]CSIR-National Physical Laboratory, Dr. KS Krishnan Road, New Delhi-110012*
*[c]Luminescent Materials Research Lab (LMRL), Department of Applied Physics, Delhi Technological University, Delhi 110 042, India*

Eco-friendly stable ZnO nanoparticles (NPs) were synthesized by sol-gel method using citrus limon as a chelating agent. The structural analysis and optical properties of ZnO NPs annealed in air ambient at different temperatures ranging from 600 to 900 ℃ are reported. A detailed X-ray diffraction (XRD) analysis of ZnO NPs has been discussed. Morphological, and optical properties of the NPs were investigated using Field Emission Scanning Electron Microscopy (FE-SEM), XRD, Diffuse Reflectance Spectroscopy (DRS), Photoluminescence (PL), Fourier Transform Infrared spectroscopy (FTIR) and Raman Spectroscopy techniques. The FE-SEM reveals that the average particle size of ZnO NPs increased from 62 to 77 nm due to agglomeration of particles or Ostwald ripening which usually happens at high temperature. Physical and microstructural properties such as crystallite size, strain, stress and dislocation density of ZnO NPs were obtained from XRD spectra studies and compared the estimated parameters by various models. It is observed that, the crystallite size increased after annealing and the strain present in NPs is due to defects. Kubelka-Munk function is used to obtain bandgap of NPs from DRS spectra. A slight decrease in the bandgap of ZnO NPs is due to the increase in particle size. PL spectra reveals that the emission around 402 nm near the UV region is attributed to the near band edge emission (NBE) and emission peaks in a visible region is due to the transition of electrons from traps in bandgap to the valance band.

**Keywords: Zinc oxide,** Nanoparticles, Sol-gel Technique, Limon, X-ray Diffraction, Optical properties.

*Corresponding author: krao@south.du.ac.in ; drkoti2010@gmail.com(K. R. Peta)
First Author: anamikakem@south.du.ac.in (A. Kem)

Accepted Manuscript

## 1. Introduction:

In recent years, there has been considerable attention of the researchers on multifunctional ZnO material because of its hard, rigid structure, high chemical stability, low toxicity, high electrochemical coupling coefficient, high photo stability and broad range of UV rays absorption [1]. Particularly, ZnO nanoparticles (NPs) have been using for various applications such as solar cells[2], biomedical [3], chemical sensors [4], gas sensors [5] and photo catalysis [6]. Because, ZnO has intrinsic n-type wide band gap (3.37 eV at room temperature) [7], and 60 meV exciton-binding energy [8] which provides chemical, electrical and thermal stability [9].

The two main approaches to get nano dimensions of materials are top-down and bottom-up. Under these two mentioned categories there are numerous methods to synthesize/obtain ZnO NPs. The authentic methods which are mentioned in the existing literature are hydrothermal method [10], co-precipitation method [11], sol-gel method [12], spray pyrolysis [13] and microwave method [14]. However, in these approaches, the production of ZnO NPs includes toxic chemicals, which are harmful for our environment and also not cost efficient. Hence, at present more focus is on green synthesis of ZnO NPs which helps to counter the undesired and harmful production of chemicals. The production of these NPs through green synthesis route is also known as biogenic NPs [15]. Moreover, in the process of annealing ZnO NPs temperature affects the size distribution of ZnO NPs, crystal quality and alters the optical properties[16] due to the absorption of atoms from the environment by dangling bonds or zinc vacancy ($V_{Zn}$) or oxygen vacancy ($V_O$) present in high surface area of NPs. Hence, stability of ZnO NPs is most important for various applications. Aljawfi et. al suggested thermal annealing of ZnO NPs in air ambient can improve the stability of ZnO NPs in terms of crystalline quality and healing of oxygen vacancies [17].

Currently, naturally occurring toxic free and ecological friendly chelating agents have been gaining much attention for the production of highly crystalline ZnO NPs. In 2017, Agarwal et al. gave the review paper on different plant mediated synthesis of ZnO NPs [18]. In 2015, Thema et *al*. [19] have synthesized the highly crystalline ZnO NPs by using Agathosma betulina as an effective chelating agent and it was demonstrated for the first time. Mayedwa et al. [20] have reported an exceptionally simple, cheap and well-founded method for the synthesis of ZnO NPs using Sageretia thea which acts as an effectual chelating agent. The synthesized ZnO NPs has the hexagonal wurtzite structure with the particle size of 28.09 nm ± 5 nm. After the heat treatment the ZnO NPs showed significant crystallization at 300 °C



Accepted Manuscript

and 500 °C. Soares et al. [21] have successfully synthesized the ZnO NPs using whey as chelating agent via sol-gel method. In their study, TEM images show spherical and highly agglomerated ZnO NPs. The structural analysis namely X-ray diffraction (XRD) and Raman spectroscopy confirm that ZnO NPs exhibited wurtzite crystalline structure and their size increased from 18 to 89 nm with the increase of calcination temperature range from 400 ºC to 1000 ºC. Recently, Valian et al. [22] have delineated the synthesis of ceramic nanocomposites of $DyMnO_3$-ZnO using Vitis vinifera juice, Hibiscus sabdariffa and rhus as natural chelating agents. The produced nano composites were successfully applied for the determination of atenolol (ATN) in biological samples. These green nanocomposites also reveal the potential application in drug analysis of biological samples. Out of all the different green synthesis routes, plant extracts have been considered to be the easiest and the simplest methods. Particularly, citrus fruits contain bioactive compounds such as carotenoids, limonoids, flavonoids and coumarins[23][24]. Compared to the other conventional synthetic chelating agent, the lemon juice is considered toxic free, inexpensive, easily accessible and biodegradable [25].

In this work, because of the simple steps and cost-efficient bottom-up approach sol-gel process has been accustomed to prepare the optically and structurally stable ZnO NPs with the use of lemon juice. Characteristics and properties of the synthesized NPs of ZnO were analysed by techniques namely X-ray diffraction (XRD), Field emission scanning electron microscopy (FE-SEM), Diffuse reflectance Spectroscopy (DRS), Photoluminescence spectroscopy (PL), Fourier Transform Infrared spectroscopy (FTIR) and Raman Spectroscopy techniques.

## 2. Experimental details

### 2.1 Materials

Zinc nitrate hexahydrate ($Zn(NO_3)_2.6H_2O$) of 99.9 % purity was acquired from the Sigma-Aldrich and was used as received in the experiment. The lemon juice ($C_{36}H_{32}Cl_4N_6O_4$) of pH value ~2.6 extracted from lemon fruit which can be obtained from the local market.

### 2.2 Synthesis of ZnO NPs

A sol-gel method is one of the extensively used methods for the preparation of ZnO NPs/nanostructures. Figure.1 is the schematic representation of stratagem for the synthesis of ZnO nano particles using citrus limon. In the first step, the precursor $Zn(NO_3)_2.6H_2O$ of about 20 g was dissolved in 40 ml of milli-q water with ~18.2 MΩ $cm^{-1}$resistance. Then this



Accepted Manuscript

cationic solution was continuously stirred for 2 hours using the magnetic stirrer at room temperature to attain the homogeneous solution. The ratio of Zinc precursor to the lemon juice (chelating agent) was maintained as 1:3. In the second step, a yellow pigment gel was obtained by adding lemon juice (chelating agent) to the cationic solution. The solution was dehydrated at 100 ºC by continuous stirring at 800 rpm. A fluffy kind of mass was obtained when the yellow pigment gel obtained from the above process was dried at 170 ℃ for 10 hours using vacuum oven. In the final step, after cooling it down to room temperature, it was grounded for 30 minutes using agate mortar to procure ZnO NPs. Finally, ZnO NPs was sintered at different temperatures 600, 700, 800 and 900 ℃ for one hour in the air and again grounded for 1 hour to obtain ZnO NPs.

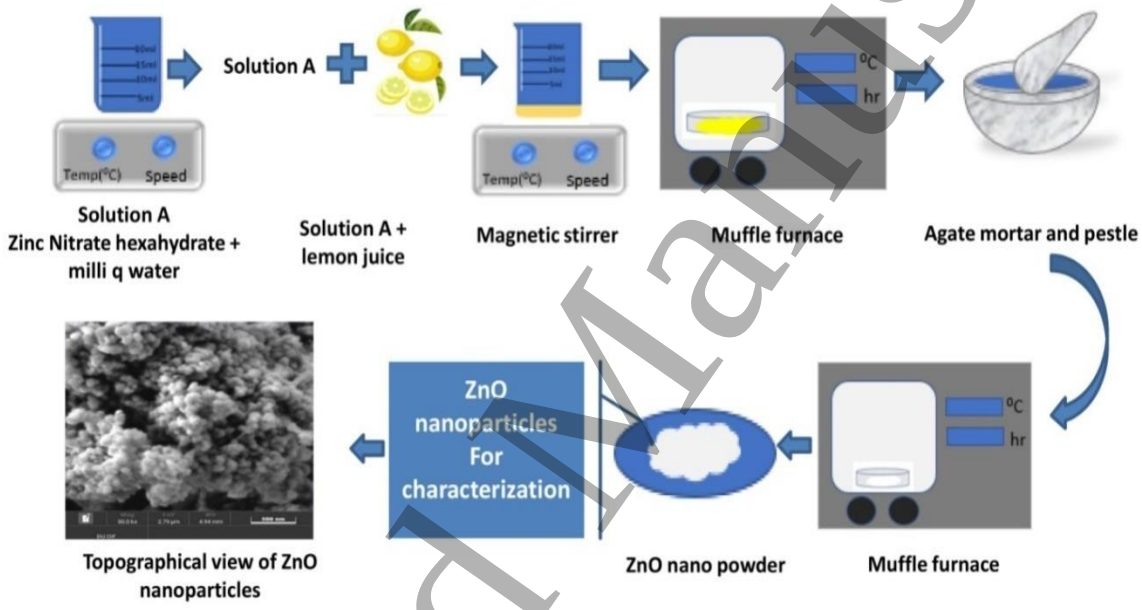



**Figure1.** Schematic diagram to show the synthesis process of ZnO nanoparticles by Sol gel method by using lemon as a chelating agent.

### 2.3 Characterization

The morphological properties of the ZnO NPs were characterized by FE-SEM (Model No. TESCAN S8152). The crystallinity of the ZnO NPs was examined by XRD instrument of $\lambda = 1.54$ Å for Cu-$K_\alpha$ radiation (model: Rigaku mini-600). Bandgap of ZnO NPs estimated from measured reflectance spectra using UV/Vis/NIR spectrophotometer (PerkinElmer, Model No. LAMBDA 1050) and the PL emission spectra of the ZnO NPs were measured at room temperature using fluorescence spectroscopy (PerkinElmer, Model No. LS-55). FTIR (Thermo Scientic Nicolet iS50 Spectrometer) was carried out to explore various functional groups of the prepared ZnO NPs. The Raman spectra were measured using RENISHAW Raman microscope enclosure RE 02, model: InVia and SERIAL No. H32320 with excitation



Accepted Manuscript

wavelength 532 nm, exposure time 90 sec, resolution 1 $cm^{-1}$ and laser spot area is 1 µm. The muffle furnace used for annealing was Thermotech TIC-4000.

## 3. Results and discussion

### 3.1 FE-SEM studies

The surface morphologies, particle size of the 700 ºC and 900 ºC annealed in air ambient of ZnO NPs were examined by FE-SEM studies. The morphological characteristic images for both samples are depicted in figure 2(a) and (c) respectively and inset shows the particle size distribution histograms of ZnO NPs. Spherical and oval shapes of NPs for both annealed samples were observed. The average particle size is evaluated using java based image processing software (Image J) and the calculated average size of the particle for 700 ºC sample is 64 nm and for 900 ºC is 77 nm. These results commensurate with XRD outcomes as well. The particle size is comparatively greater than the crystallite size which proves that the particle consists of an agglomeration of many crystallites [26] arranged in different phases. The increase in particle size with annealing temperature from 600 to 900 ºC is ascribed to the agglomeration of smaller particles to larger particles which generally happens at higher temperatures and also due to Oswald ripening thermal effect. The molecules on the surface of a particle are comparatively less energetically stable than the particles which are well aligned and internally packed. This fact leads to the conclusion that larger particles are more energetically favoured than smaller particles. The surface area to volume ratio decreases as the radius of the particle increases so the large particles have lower surface area to volume ratio which results in lower energy state and lower surface energy. Considering that the system tries to minimize its total energy so the molecules on the small particle surface will diffuse; continue to shrink and add to the surface of the larger particles and hence continue to grow in size [27]. The energy-dispersive X-ray analysis (EDAX) of the green synthesized ZnO nanoparticles samples shown in figure 2(b) and 2(d) revealed elemental identification of Zinc (Zn) and Oxygen (O) with quantitative compositional information. The two prominent peaks in the spectrum were found to have weight percentages of 71.86 of Zinc and 28.14 of oxygen for 700 ℃ and 71.61 of Zinc and 28.39 of oxygen for 900 ℃ respectively. This analysis inveterate the formation of ZnO NPs by the sol gel method. Table.1. shows the comparison of particle size of previous work from different synthesis method with the present work. Furthermore, the size and shape of ZnO NPs synthesized by sol gel technique using lemon as a chelating agent are stable during thermal annealing in air ambient.

Accepted Manuscript

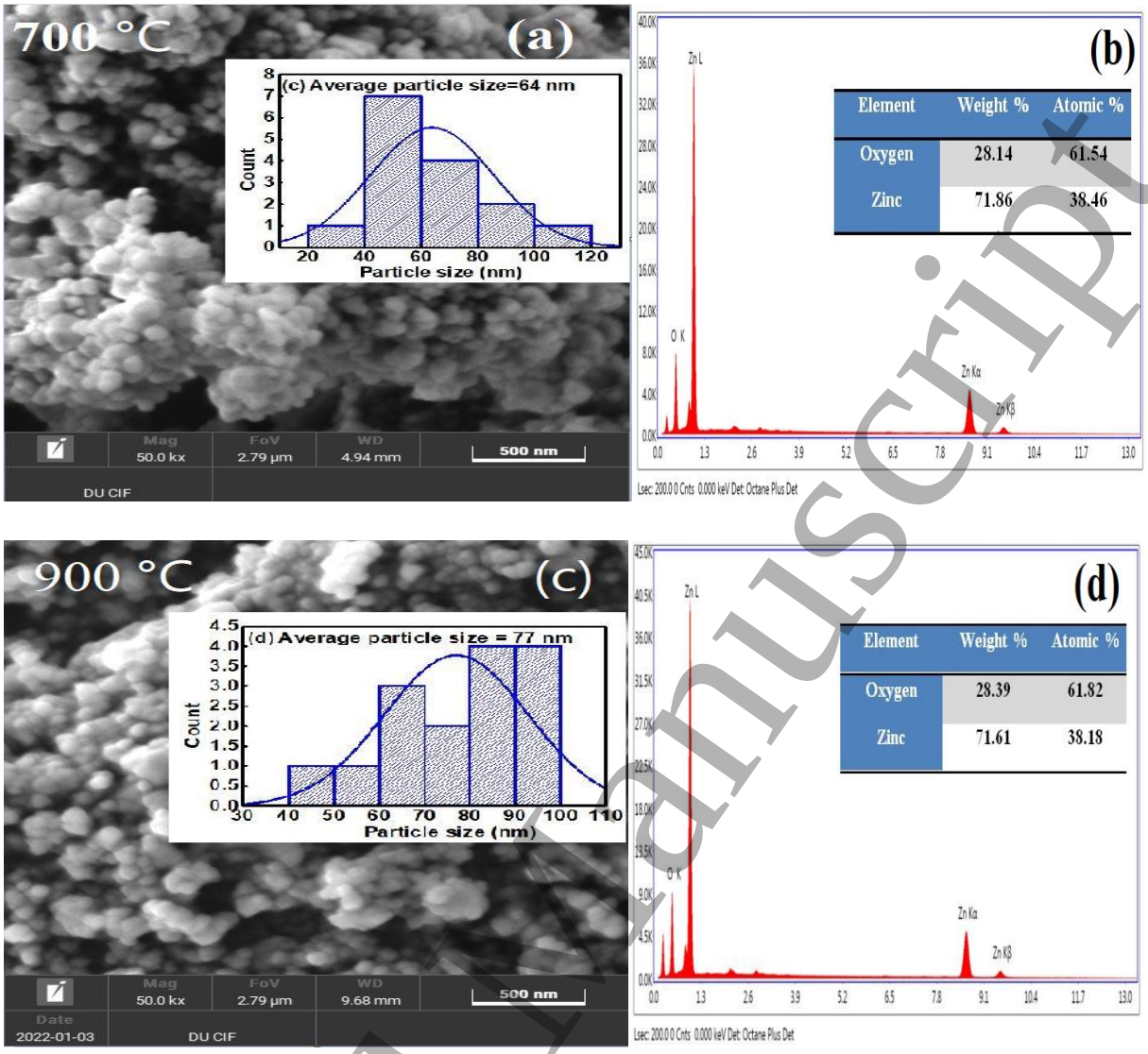



**Figure 2.** FE-SEM images and particle size distribution histograms of ZnO NPs (a, c) sintered at 700 ºC and 900 ºC temperatures respectively and (b, d) represents the EDAX of 700 ºC and 900 ºC ZnO NPs.

Table 1: Comparison of particle size in present work with the previous work with different synthesis method.

| S.No. | Synthesis Method | Materials used | Morphology | Particle Size (nm) | Ref |
|---|---|---|---|---|---|
| 1. | Sonochemical | Zinc acetate dihydrate, Ethanol | Irregular shape | 200–250 | [28] |
| 2. | Microwave-assisted | Zinc nitrate dehydrate Sigma Aldrich) and hexamethylenetetramine | Spherical shape to hexagonal rods | 25 nm - micro/sub micro sizes | [29] |

Accepted Manuscript

| | | | | | |
|---|---|---|---|---|---|
| 3. | Micro emulsion | Sodium hydroxide, emu oil, Span 80 Tween 80, de-ionized water, acetone Hexane, zinc acetate dihydrate | Elongated and rounded | 31.2 | [30] |
| 4. | Solvothermal | Solvothermal zinc nitrate hexahydrate, sodium hydroxide 20 to 30 gluconic acid, oleic acid, tween 80 ,DPPH, ABTS, (Sigma Aldrich), phosphate buffer hydrogen peroxide, potassium persulphate, methanol, Double distilled water | Nano granular, flower-like nano rod and nano flakes | 20 to 30 | [31] |
| 5. | Flash combustion | Zinc nitrate hexahydrate, citric acid | Roughly spherical oval morphologies | 120 | [32] |
| 6. | Sol gel | Zinc nitrate, aqueous extract of Mueller-Hinton Agar (M173), DPPH, potato dextrose agar, streptomycin quarts distilled water, Periconium sp. | quasi spherical | 16–78 | [33] |
| 7. | Sol gel | Zinc nitrate hexahydrate, lemon juice milli q water | Spherical and oval | 62-77 | Present work |

### 3.2 X-ray diffraction analysis

The crystallinity and phase of our ZnO NPs were determined by X-ray diffraction pattern studies. Figure 3 shows the X-ray diffraction patterns of the ZnO NPs as a function of annealing temperature. The diffraction pattern shows peaks at the angles 31.74°, 34.39°, 36.21°, 47.47°, 56.55°, 62.81°, 66.34°, 67.89°, 69.05°, 72.51° and 76.91° corresponding to the planes (100), (002), (101), (102), (110), (103), (200), (112), (112), (004) and (202). It is found that, these phases correspond with the pattern of the Joint Committee on Powder Diffraction Standards (JCPDS) card No.36-1451 of ZnO. Other compound related peaks were not identified in the spectra which attributed to the purity of our ZnO NPs as there would be

Accepted Manuscript

no complex products or residue of precursors and our results are good in accordance with the results [34] reported by Vijayaprasath et al.

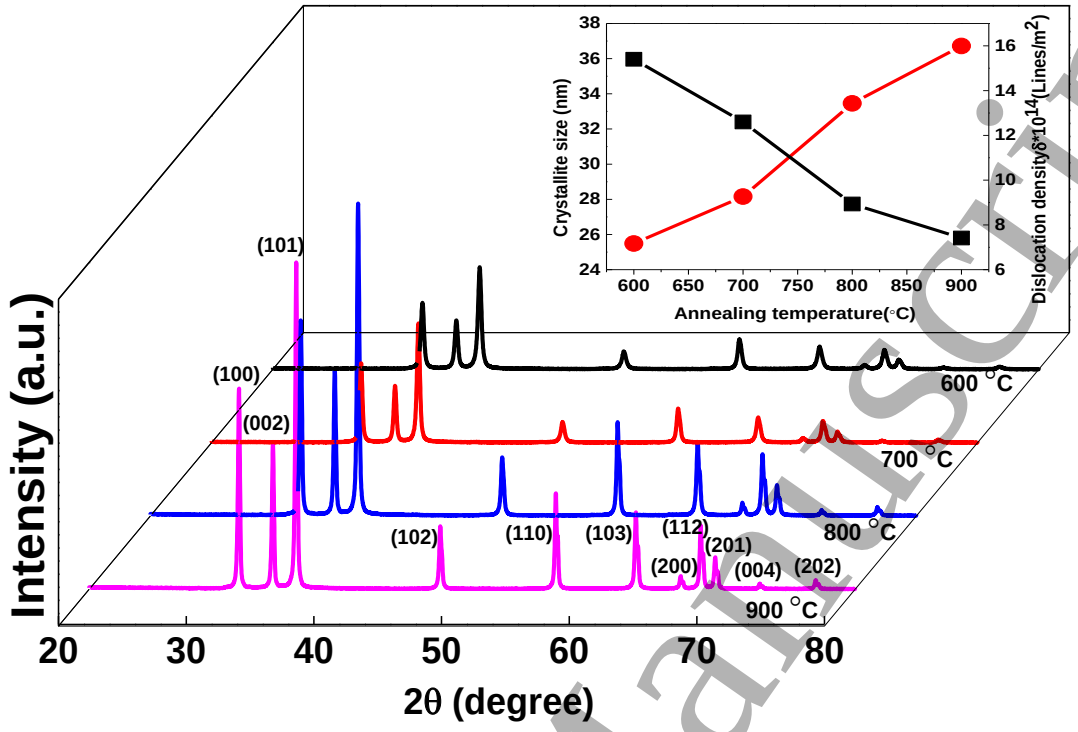



**Figure 3.**X-ray diffraction spectra of annealed ZnO NPs at various temperatures in air (600 ºC, 700 ºC, 800 ºC and 900 ºC) and inset is the variation of crystallite size and dislocation density as a function of annealing temperature.

The X-ray diffraction pattern (figure.3) was used to evaluate many lattice parameters such as distance between adjacent planes, distortion factor, cell volume, bond length and positional parameter of using the following equations.

$$\frac{1}{d^2_{(hkl)}} = \frac{4}{3}\left(\frac{h^2+hk+k^2}{a^2}\right) + \frac{l^2}{c^2} \quad (1)$$

where d is the plane spacing of miller indices h, k and l. The constant distance between unit cells in a crystal lattice 'a' and 'c' can be estimated using $\frac{\lambda}{\sqrt{3}\sin\theta}$ and $\frac{\lambda}{\sin\theta}$ respectively [35][36].

The Zn–O bond length '$L$' can be determined using [37]

$$L = \sqrt{\left(\frac{a^2}{3} + \left(\frac{1}{2} - u\right)^2 c^2\right)} \quad (2)$$

Where '$u$' is the positional parameter and can be determined as $u = \frac{a^2}{3c^2} + 0.25$, it measures the amount by which each atom is dislocated with the adjacent along the preferential 'c' axis.



Accepted Manuscript

Simultaneously, the cell volume is estimated using the relation $V = \frac{\sqrt{3}a^2c}{2}$ respectively [35]. The obtained physical parameters of ZnO nano crystallites such as distortion factor, cell volume, Zn-O bond length from XRD spectra for different temperatures are given in the Table.2.

Texture coefficient is an important parameter which tells about the degree of preferred orientation and can be estimated from the Harris's analysis [38]

$$TC_{hkl} = \frac{I_{hkl}}{I_{0(hkl)}}\left[\frac{1}{N}\sum_1^N \frac{I_{hkl}}{I_{0(hkl)}}\right]^{-1} \quad (3)$$

Where $TC_{hkl}$ is the texture coefficient of the plane, $I_{(hkl)}$ is peak intensity from experimental spectra $I_{O(hkl)}$ is peak intensity of X-ray diffraction from JCPDS 36-1451, and N is the number of reflections taken into account. The estimated lattice parameters of ZnO NPs for distinct annealing temperatures are given in Table.1. It is perceived that, the texture coefficient value is higher than unity which means the higher degree of growth in all preferred directions [39].

Table 2. Lattice Parameters, Distortion factor, Cell Volume, Zn-O bond length, Positional parameter and Texture Coefficient for annealed ZnO nanoparticles at different temperatures

| S.No | Temperature | Lattice parameters | | Distortion Factor | Cell Volume | Zn-O | Positional parameter | Texture coefficient |
|---|---|---|---|---|---|---|---|---|
| | (°C) | a (A°) | c (A°) | (c/a) | $(A°)^3$ | L (A°) | (u) | (TC) |
| 1 | 600 | 3.2556 | 5.2151 | 1.60188 | 47.86 | 1.9301 | 0.3798 | 1.11 |
| 2 | 700 | 3.2556 | 5.2151 | 1.60203 | 47.87 | 1.9814 | 0.3798 | 1.13 |
| 3 | 800 | 3.2478 | 5.2028 | 1.60194 | 47.52 | 2.6877 | 0.1299 | 1.16 |
| 4 | 900 | 3.2529 | 5.2101 | 1.60167 | 47.74 | 1.9794 | 0.3799 | 1.12 |

### 3.2.1. Scherrer method

Scherrer equation $D = \frac{K\lambda}{\beta \cos\theta}$ [40] is used to determine the crystallite size of ZnO NPs. Where D is the crystallite size, k is Scherrer's constant of value 0.9, λ is the wavelength of the X-rays source (1.54 A° for Cu $K_\alpha$ radiation), β is the full width at half-maximum (FWHM) and θ is Bragg's angle of the diffraction peak corresponding to planes. Dislocation density ($\partial$) is the inverse of square of crystallite sizeand quantify the defects in the NPs [41]. The estimated crystallite size and dislocation density are mentioned in Table 3. It is observed that the crystallite size is increased and dislocation density decreased after annealing.



Accepted Manuscript

### 3.2.2. Williamson – Hall method

There is shift in diffraction peaks towards the higher angle (figure 4 and inset) attributed to lattice strain [42] in ZnO NPs for 800 °C annealed sample. Hence, Williamson–Hall (WH) method can be used to evaluate the intrinsic lattice strain whereas Scherrer method can only give information about the crystallite size. According to WH method the peak broadening of the XRD pattern eventuate due to both crystallite size and intrinsic strain [35][43]. Hence FWHM of diffraction pattern defined as $\beta_{hkl} = \beta_{size} + \beta_{strain}$[44]. In this study, crystallite size, strain, energy density and stress were estimated from all modified WH-methods namely Uniform Deformation Model (UDM), Uniform Strain Deformation Model (USDM), Uniform deformation energy density model (UDEDM) and results are compared in the Table 3.

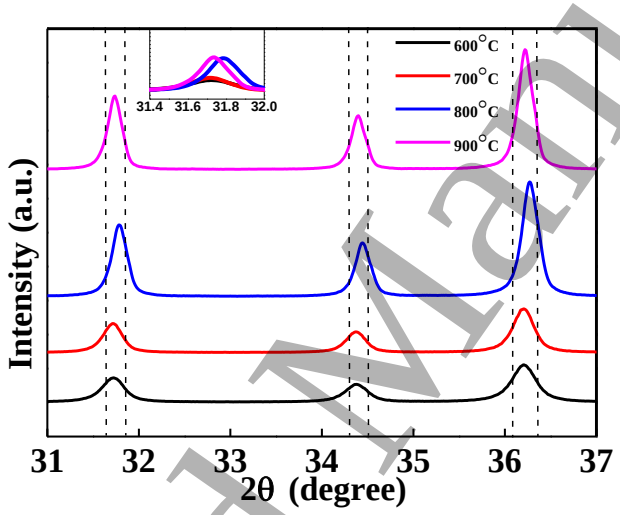



**Figure 4.** Shifting of XRD peaks towards higher angle for 800 °C annealed ZnO NPs.

**(a) UDM**

In this model, crystal is believed to be isotropic in nature as all physical properties show uniformity in all directions. Crystal imperfections and distortion causes the strain and strain is given by $\varepsilon \approx \frac{\beta_s}{tan\theta}$. Consequently, the total peak broadening relation can also be written as $\beta_{hkl}\cos\theta = \frac{K\lambda}{D} + 4\varepsilon\sin\theta$ [45] and the plot of $\beta_{hkl}\cos\theta$ vs. $4\sin\theta$ is linear (supplementary file.1). The estimated strain and crystallite size from the slope and the Y-intercept of linear fit and are given in Table 3.

Accepted Manuscript

**(b) USDM**

The assumption made from the UDM that the strain is homogenous in all growth directions of the crystal is not realistic in numerous cases. Accordingly, modification has been done in the WH equation by considering the anisotropic strain in the crystal. In the modified model USDM, lattice deformation stress is supposed to be homogenous in every crystallographic direction and consists of small micro-strain [36]. According to the law of physics, Hooke's law states [46] that the extent to which material stretches or compresses is in a linear relationship with applied stress as $\sigma = \varepsilon^* Y_{hkl}$, where σ is the stress applied, ε is the anisotropic strain to which extent material compresses or stretches and $Y_{hkl}$ being Young's modulus or elasticity of modulus. In this approximation, the W-H equation is remodelled as $\beta_{hkl}\cos\theta = \frac{K\lambda}{D} + 4\frac{\sigma}{Y_{hkl}}\sin\theta$. This equation is known as USDM where it considers the stress to be uniform in every crystallographic direction. Here, $Y_{hkl}$ being the Young's modulus or elasticity of modulus which can be expressed for the hexagonal structure [43] as

$$Y_{hkl} = \frac{\left[h^2 + \frac{(h+2k)^2}{3} + \left(\frac{al}{c}\right)^2\right]}{s_{11}\left(h^2 + \frac{(h+2k)^2}{3}\right)^2 + s_{33}\left(\frac{al}{c}\right)^4 + (2s_{13} + s_{44})\left(h^2 + \frac{(h+2k)^2}{3}\right)\left(\frac{al}{c}\right)^2} \quad (4)$$

Here, the elastic compliances of ZnO are $s_{11}$, $s_{13}$, $s_{33}$, $s_{44}$withvalues of 7.858 x $10^{-12}$, 2.206 x $10^{-12}$, 6.940 x $10^{-12}$, 23.57 x $10^{-12}$ $m^2$ $N^{-1}$additionally 'a' and 'c' are lattice parameters. The obtained Young's modulus for hexagonal ZnO NPs by substituting all these values in the above equation is 127 GPa [47]. In this model, stress and crystallite size of NPs were obtained from slope and intercept of linear fitting of curve $\beta_{hkl}\cos\theta$versus $\frac{4\sin\theta}{Y_{hkl}}$ plot (supplementary file.2) and are shown in Table 3.

**(c) UDEDM**

This is another modified model of WH method called UDEDM model which mention the uniform anisotropic micro strain in every crystallographic direction and it helps to evaluate the energy density of the crystal. In consonance with Hooke's law, energy density (u) is related to the strain and is given by the relation$u = \varepsilon^2\frac{Y_{hkl}}{2}$and stress and strain are linked as $\sigma = \varepsilon^* Y_{hkl}$. So, the strain related with energy density as $\varepsilon = \sigma\sqrt{2u/Y_{hkl}}$, where $Y_{hkl}$ is Young's modulus or elasticity of modulus. Using these the peak broadening expression [35] can be written as



Accepted Manuscript

$$\beta_{hkl}\cos\theta = \frac{K\lambda}{D} + 4\sigma\sin\theta\sqrt{\frac{2u}{Y_{hkl}}} \quad (5)$$

This equation is known as the UDEDM equation. By linear fitting of data which is obtained from $4\sin\theta\sqrt{\frac{2u}{Y_{hkl}}}$ versus $\beta_{hkl}\cos\theta$ plot (Supplementary file-3), the energy density and crystallite size can be evaluated from the slope and intercept of Y-axis and the obtained values are given in the Table 3.

### 3.2.3. Size-strain plot (SSP)

In WH method, the effect of size and strain induced broadening causes the total broadening of peaks in accordance of diffraction angle. But there are various models, which deal with the analysis of peak profile. In terms of isotropic line broadening, the SSP method provides better results compared to the previously mentioned method because this method is more significance to low angle reflections and these low angles give more explicit information about the size and strain. SSP is a type of model that deals with the analysis of peak profile as a combined function of Lorentzian and Gaussian method, in which size broadened XRD profile is explained by Lorentz function and strain broadened profile is explained by the Gaussian function [35][45]. So, the final broadening of the diffraction peak can be given as $\beta_{hkl} = \beta_L + \beta_G$, where $\beta_L$ and $\beta_G$ are the peak broadening given by Lorentz and Gaussian function. Accordingly, the SSP calculation is done using the equations given by

$$(d_{hkl}\beta_{hkl}\cos\theta)^2 = \frac{K\lambda}{D}(d_{hkl}^2\beta_{hkl}\cos\theta) + \frac{\varepsilon^2}{4} \quad (6)$$

Where $d_{hkl}$ represents lattice distance between the (hkl) planes. The calculated crystallite size and strain values from slope and intercept of linear fit to Y-axis of the graph (supplementary file-4) which was drawn between $(d_{hkl}\beta_{hkl}\cos\theta)^2$ versus $d_{hkl}^2\beta_{hkl}\cos\theta$of ZnO NPs annealed at different temperatures are given in table3. The estimated crystallite size and strain from all models increases with the increase of annealing temperature.

Table 3: Geometrical physical parameters of ZnO NPs using Scherrer, W-H and size- strain plot methods.

| S.No. | Annealing Temperature | Scherrer Method | Williamson-Hall Method | | | | | | | | | | Size-Strain Plot | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | UDM | | USDM | | | UDEDM | | | | | |
| | | Size D | Dislocation Density (δ) $10^{14}$ | Size D | Strain | Size D | Strain | Stress σ | Size D | Strain | Energy | Stress σ | Size D | Strain |
| | (°C) | (nm) | lines/m$^2$ | (nm) | ($10^{-3}$) | (nm) | ($10^{-3}$) | (MPa) | (nm) | ($10^{-3}$) | kJm$^{-3}$ | (MPa) | (nm) | ($10^{-3}$) |
| 1 | 600 | 25 | 15.4 | 29 | 0.84 | 29 | 0.84 | 109.46 | 29 | 0.84 | 44.91 | 106.94 | 24 | 0.496 |
| 2 | 700 | 28 | 12.6 | 37 | 0.94 | 29 | 0.94 | 109.46 | 37 | 0.93 | 55.50 | 119.37 | 29 | 0.603 |
| 3 | 800 | 33 | 8.93 | 60 | 1.18 | 60 | 1.18 | 153.38 | 60 | 1.17 | 87.63 | 149.81 | 42 | 0.718 |

Accepted Manuscript

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 900 | 36 | 7.42 | 49 | 1.05 | 63 | 1.05 | 136.59 | 63 | 1.05 | 70.05 | 133.39 | 43 | 0.614 |

### 3.2.4 Kinetic growth of ZnO nanoparticles

It is scrutinized that, the crystallite size of the ZnO NPs increases with annealing temperature in air ambient. The kinetics of the crystallite growth is described by [48]

$$G^n = k_o t e^{\frac{-Q}{RT}} \quad (7)$$

Where G is the average crystallite size, n is the kinetic growth exponent, $K_o$ is a constant, t is time, Q is the activation energy, R represents the gas constant and T is the absolute temperature. The kinetic growth exponent was evaluated from the slope of the fitted line of log G vs. log t [48]and the activation energy which was calculated from the slope of the fitted line of ln G vs. $T^{-1}$ (figure 5.) came out to be 115.68 kJ $mol^{-1}$ which is required to accept the electrons by the acceptor levels from the valence band or to hoist the electron from the donor levels respectively. The estimated activation energy of ZnO NPs is in good agreement with [49][50] reported values by Bhargava et al. and Zhang et al. respectively.

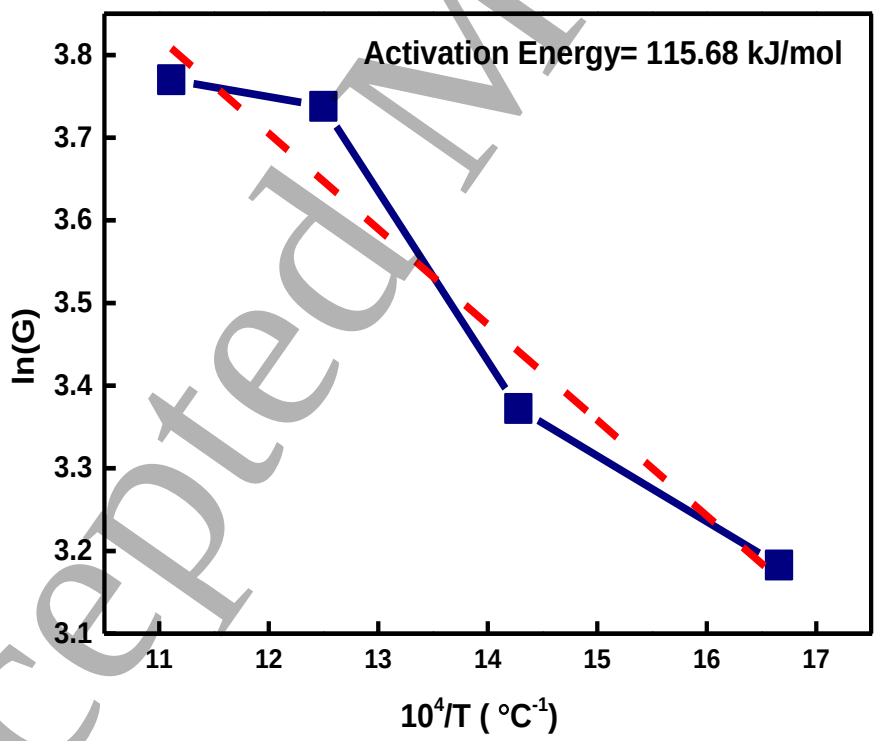



**Figure 5.** Arrhenius plot of the crystallite growth rate for green synthesised ZnO nanoparticles.

Accepted Manuscript

### 3.3. Optical properties

#### 3.3.1. UV-Vis diffuse reflectance studies

The optical band gap of the ZnO NPs were evaluated from diffuse reflectance spectra by plotting square of the Kubelka–Munk function against energy ($E_g$) and exploiting the linear part to zero (figure 6). Here, Kubelka–Munk function is given as [51][52]

$$F(R_\infty) = \frac{\alpha}{S} = \frac{(1-R)^2}{2R} \quad (8)$$

Where "α" represents the absorption coefficient, S is denoted as the scattering coefficient**,** R is the reflectance and use of $n = \frac{1}{2}$ for direct transition. The bandgap energy of the ZnO NPs is obtained from the intercept of the line. The inset figure 5 shows the variation in the optical band gap of the ZnO NPs in accordance of annealing temperature. It is observed that, there is no significant variation in band gap which attributes the stability of our ZnO NPs with annealing temperature. The optical band gap of ZnO NPs formed by the sol-gel method in the presence of citrus limon as a chelating agent is in good agreement with reported values. According to Davis et al.[53], the optical band gaps exhibiting by ZnO structures can be aligned from 3.10 eV to 3.37 eV formed in different solution techniques. Very recently, Sone et al. [54] reported the band gap value of 3.14 – 3.27 eV for green synthesised ZnO NPs. Ali et al [55] identified the band gap increasing in Ni/Co co-doped ZnO NPs by varying concentration formed by sol-gel technique. In our case the bandgap of ZnO NPs minutely decreased in order of meV with annealing temperature. It was put forwarded that the optical band gap decreased owing to increase in particle size [56].

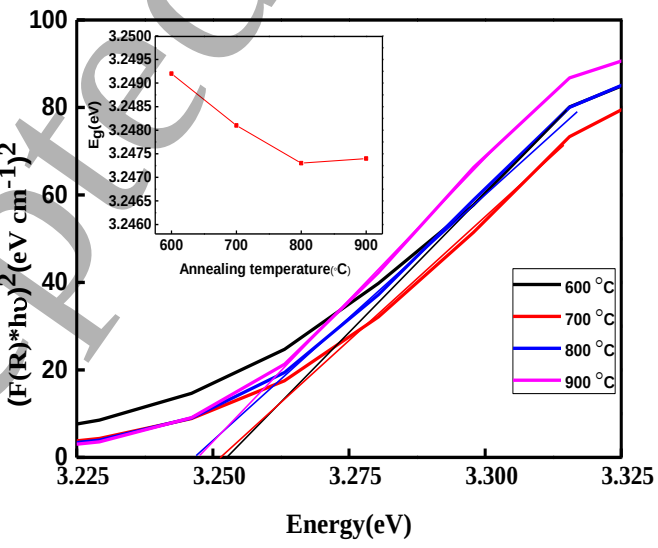



**Figure 6.** $(F(R) * h\vartheta)^2$ Vs. Energy band gap of ZnO nanoparticles annealed at various temperature.



Accepted Manuscript

### 3.3.2. Photoluminescence studies

Figure 7 shows the room temperature photoluminescence spectra of ZnO NPs synthesized by sol-gel method using citrus limon as a function of annealed temperature. The intensity of PL spectra was found to escalate with annealing temperature and can be related to the size and shape of the particles [57][58]. For easy electron transfer from valance band to conduction band, 340 nm excitation wavelength was used which helped for the excitation of defect level within the bandgap.

The PL emission spectra of most ZnO structures showed a peak near the UV band called near band-edge emission (NBE) and followed by emission peaks in visible band by mid bandgap states luminescence [59]. These higher wavelength peaks originated from various defects regulated by the growth technique of ZnO and can suppress the efficiency of UV emission while increasing the lasing action. In our case, the emission peak at 402 nm is due to free excitons recombination via an exciton-exciton collision process [60]. ZnO nanostructure known to have various defects such as zinc vacancy, oxygen vacancy, zinc antisite, oxygen antisite, zinc interstitial and oxygen interstitial with different charged states namely, neutral, single-charged and double charged [61]. These defects cause the broad emission in visible region. The conspicuous emission peaks in the visible region are at $\lambda$ = 420 nm, 486.5 nm, 529 nm and 603 nm and their energy transitions are shown in energy band diagram figure 7 inset. The peak at 420 nm arises from the electron transition from Zinc interstitial to the valence band. The origin of two aforementioned peaks reported violet emission in the visible region. The most intense peak at 486.5 nm emanates from the transition of an electron from Zinc interstitial to Zinc vacancies and gives blue emission in the visible region [62][63]. Oxygen vacancies have three types of charge states chronologically $Vo^0$, $Vo^+$, and $Vo^{2+}$ and these oxygen vacancies are located underneath the bottom of the conduction band. The emission peak at 529 nm can be analogous to singly ionised oxygen vacancy. The recombination of a photogenerated hole with a singly ionised charged state of the specific defect gives green emission [61][62][64]. The Zinc vacancy and Oxygen vacancy are positioned at 0.3 eV and 0.4 eV above the top of the valence band and Zinc interstitial is located at 0.5 eV underneath the bottom of the conduction band. The peak at 603 nm gives yellow emission and it is related to Zinc and Oxygen interstitial [65].

Accepted Manuscript

### 3.3.3 FTIR studies

Presence of surface-bound species on the ZnO NPs can be scrutinized using Fourier transformed infrared (FTIR) spectroscopy. The vibrational mode at 1444 $cm^{-1}$could be analogous to hydroxyl group (O–H) bending modes.[66]. In figure8, the peak at 1116 $cm^{-1}$ [67] ascribed to the C−C stretch in aromatic ring. The peak at around 887 $cm^{-1}$ [66] is associated to substitutional hydrogen at oxygen site (HO) which is bound to the Zn site and this hydrogen may behave as a shallow donor in ZnO. The peak at 840 $cm^{-1}$ could be assigned to amine groups of C-N stretching [68]. The obtained peaks in the bracket of 400-700 $cm^{-1}$ assigned the stretching mode of ZnO. [69].

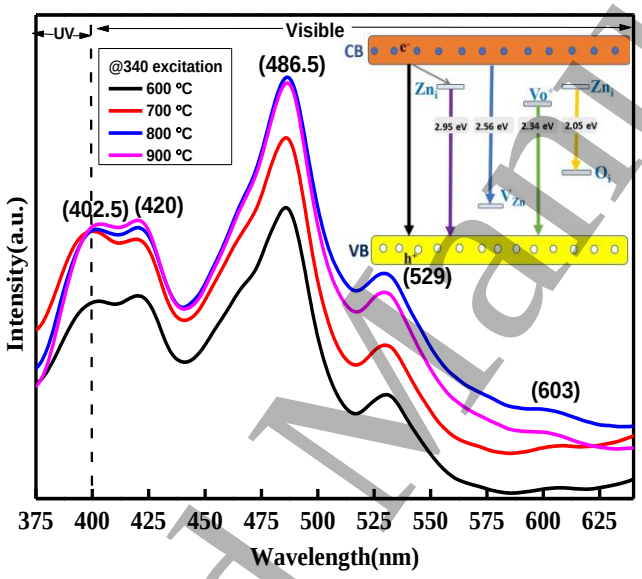



**Figure 7.** Photoluminescence emission spectra of ZnO nanoparticles annealed at distinct temperatures for 340 nm excitation wavelength and energy level diagram showing principal defects level for ZnO nanoparticles (inset).

In the FTIR spectra (Fig. 8), the prominent peak around 1444 $cm^{-1}$ showed a slight shift towards the low wave number and also decrement in their intensities, putting forwarded that the biomolecules obtained from the extract were bonded to the surface of the ZnO NPs. Furthermore, the decrease in intensity with increasing annealing temperature is due to reduction of OH groups with annealing temperature [70].

Accepted Manuscript

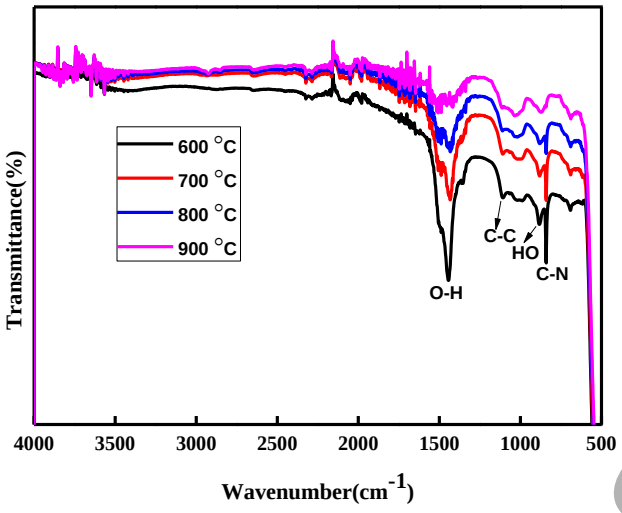



**Figure 8.** FTIR spectra of annealed ZnO NPs synthesized by sol-gel method using citrus limon as a chelating agent.

### 3.3.4 Raman Studies

Raman spectroscopy measurements were carried out in order to analyse the lattice disorder- induced in the ZnO lattice, crystallite behaviour and the defects in the ZnO NPs. Fig. 9 shows the Raman spectra which was measured at room temperature. In consonance with the group theory, the optical phonons of wurtzite ZnO appertain to the C4 6v space group and the irrevocable representation of optical phonon is given by $\Gamma_{opt} = A_1+2B_1+E_1+2E_2$ [71]. From this representation, $A_1$ and $E_1$ denote the polar modes and cleave into transverse-optical (TO) and longitudinal-optical (LO) phonons, whereas the $B_1$ represents the Raman inactive modes. Furthermore, the prominent peak observed at 98 $cm^{-1}$ and at 437 $cm^{-1}$ascribed the $E_2$ low and $E_2$ high frequency branch of Non polar $E_2$ modes, which are Raman active. Exhaustively, the two wavenumbers $E_2^{low}$ and $E_2^{high}$ are analogous to the motion of oxygen atom and zinc sub lattice, respectively. Strong $E_2^{low}$ and $E_2^{high}$ phonon modes suggests characteristic of the Wurtzite lattice and indicates good crystallinity quality. As can be seen from fig. 9, $E_2^{low}$ and $E_2^{high}$ modes improve with the increase in temperature, which can be attributed to the better crystal quality [72].

The multi phonon scattering process at 330 $cm^{-1}$is ascribed to $E_2^{high}$–$E_2^{low}$ mode[73]. This second order mode increases scantily with increasing temperature; this fact may suggest the improvement in the crystallization quality. In addition, a weak peak at 380 $cm^{-1}$ is also observed, which is assigned to $A_1$ (TO) mode, respectively. In consonance with the theory of polar optical phonons basically in wurtzite nanocrystals, the 1LO phonon mode in ZnO NPs



Accepted Manuscript

should be between 574 and 591 $cm^{-1}$ [74]. The frequency at 574 $cm^{-1}$ corresponds to the $A_1$ (LO) phonon and it observed only when the c-axis of Wurtzite ZnO is parallel to the sample surface. Whereas, the $E_1$ (LO) (591 $cm^{-1}$) phonon is observed when the *c*-axis is perpendicular to the sample surface. So, the peak around 583 $cm^{-1}$ is contributed to the overlaying of $A_1$ (LO) and $E_1$ (LO). The weak and broad peak at 583 $cm^{-1}$ in the spectrum is due to the formation of impurities and structural defects such as oxygen vacancies and zinc interstitials. As the temperature increases the peak at 583 $cm^{-1}$ does not enhance much so it can be understood that the ZnO nanoparticles is stable in terms of oxygen vacancies and zinc interstitials [72].

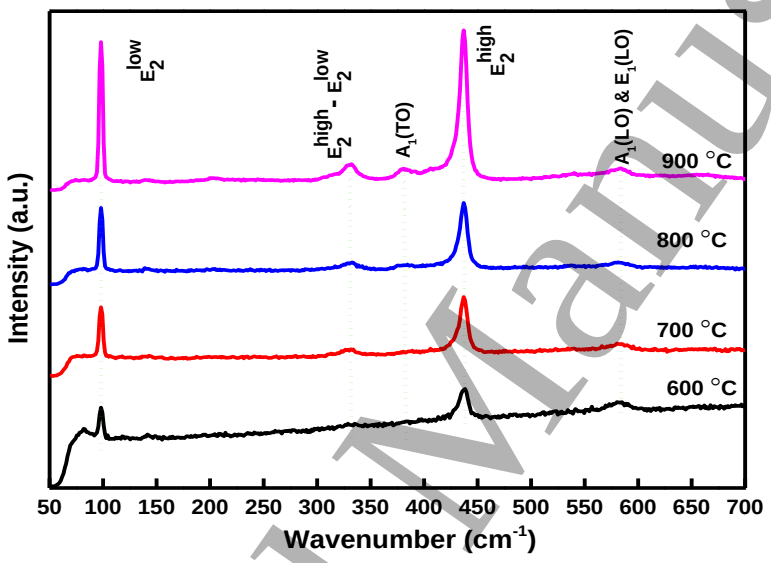



**Figure 9**. Room-temperature Raman spectra of ZnO NPs annealed in air ambient as a function of temperature.

## Conclusion

ZnO NPs were synthesised in simple and cost-effective way by sol-gel technique using lemon as a chelating agent. The results reported and analysed in this work, demonstrate that lemon is an ecologically friendly and efficient chelating agent which anticipated the formation of extremely crystalline ZnO NPs. FE-SEM revealed that, the ZnO NPs are in spherical shape with the combination of many crystallites. The X-ray diffraction analysis confirms that the ZnO NPs annealed in the range of 600 ºC to 900 ºC manifested wurtzite crystalline structure. The required activation energy came out to be 115.68 kJ $mol^{-1}$. Diffuse reflectance spectroscopy showed that the band gap of ZnO NPs was decreased in meV range with an increase in annealing temperature. The emissions in visible band such as violet, blue, green

Accepted Manuscript

and orange regions are due to defects present in the ZnO NPs. Raman-scattering spectra approved that the synthesized NPs were wurtzite hexagonal phase and owned good crystal quality with very few defects at four different annealing temperatures.

**Acknowledgment**

This research work was funded by Institution of Eminence (IoE), University of Delhi under faculty research programme (IoE/2021/12/FRP). The first author (Miss. Anamika Kem) would like to expresses her gratitude to University of Delhi for providing fellowship.

**References**

[1] M. Benhaliliba, ZnO a multifunctional material: Physical properties, spectroscopic ellipsometry and surface examination, Optik (Stuttg). 241 (2021) 167197. https://doi.org/10.1016/j.ijleo.2021.167197.

[2] S. Angaiah, S. Arunachalam, V. Murugadoss, G. Vijayakumar, A Facile Polyvinylpyrrolidone Assisted Solvothermal Synthesis of Zinc Oxide Nanowires and Nanoparticles and Their Influence on the Photovoltaic Performance of Dye Sensitized Solar Cell, ES Energy Environ. (2019) 59–65. https://doi.org/10.30919/esee8c280.

[3] N.P. Shetti, S.D. Bukkitgar, K.R. Reddy, C.V. Reddy, T.M. Aminabhavi, ZnO-based nanostructured electrodes for electrochemical sensors and biosensors in biomedical applications, Biosens. Bioelectron. 141 (2019) 111417. https://doi.org/10.1016/j.bios.2019.111417.

[4] V. Galstyan, E. Comini, C. Baratto, G. Faglia, G. Sberveglieri, Nanostructured ZnO chemical gas sensors, Ceram. Int. 41 (2015) 14239–14244. https://doi.org/10.1016/j.ceramint.2015.07.052.

[5] S. Bhatia, N. Verma, R.K. Bedi, Ethanol gas sensor based upon ZnO nanoparticles prepared by different techniques, Results Phys. 7 (2017) 801–806. https://doi.org/10.1016/j.rinp.2017.02.008.

[6] S. Bhatia, N. Verma, Photocatalytic activity of ZnO nanoparticles with optimization of defects, Mater. Res. Bull. 95 (2017) 468–476. https://doi.org/10.1016/j.materresbull.2017.08.019.

[7] Y.W. Chen, Y.C. Liu, S.X. Lu, C.S. Xu, C.L. Shao, C. Wang, J.Y. Zhang, Y.M. Lu, D.Z. Shen, X.W. Fan, Optical properties of ZnO and ZnO:In nanorods assembled by sol-gel method, J. Chem. Phys. 123 (2005). https://doi.org/10.1063/1.2009731.

[8] K. Gautam, I. Singh, P.K. Bhatnagar, K.R. Peta, The effect of growth temperature of seed layer on the structural and optical properties of ZnO nanorods, Superlattices Microstruct. 93 (2016) 101–108. https://doi.org/10.1016/j.spmi.2016.03.001.

[9] A. Wei, X.W. Sun, C.X. Xu, Z.L. Dong, M.B. Yu, W. Huang, Stable field emission from hydrothermally grown ZnO nanotubes, Appl. Phys. Lett. 88 (2006) 1–4. https://doi.org/10.1063/1.2206249.

[10] H.S. Wasly, M.S.A. El-Sadek, M. Henini, Influence of reaction time and synthesis temperature on the physical properties of ZnO nanoparticles synthesized by the hydrothermal method, Appl. Phys. A Mater. Sci. Process. 124 (2018) 1–12. https://doi.org/10.1007/s00339-017-1482-4.

[11] R.E. Adam, G. Pozina, M. Willander, O. Nur, Synthesis of ZnO nanoparticles by co-precipitation method for solar driven photodegradation of Congo red dye at different pH, Photonics Nanostructures - Fundam. Appl. 32 (2018) 11–18. https://doi.org/10.1016/j.photonics.2018.08.005.

Accepted Manuscript

[12] K. Prabakar, H. Kim, Growth control of ZnO nanorod density by sol-gel method, Thin Solid Films. 518 (2010) e136–e138. https://doi.org/10.1016/j.tsf.2010.03.103.
[13] A.S. Ibraheam, J.M. Rzaij, M.A. Fakhri, A.W. Abdulwahhab, Structural, optical and electrical investigations of Al:ZnO nanostructures as UV photodetector synthesized by spray pyrolysis technique, Mater. Res. Express. 6 (2019). https://doi.org/10.1088/2053-1591/ab06d4.
[14] J. Wojnarowicz, T. Chudoba, W. Lojkowski, A review of microwave synthesis of zinc oxide nanomaterials: Reactants, process parameters and morphoslogies, Nanomaterials. 10 (2020). https://doi.org/10.3390/nano10061086.
[15] J. Singh, T. Dutta, K.H. Kim, M. Rawat, P. Samddar, P. Kumar, "Green" synthesis of metals and their oxide nanoparticles: Applications for environmental remediation, J. Nanobiotechnology. 16 (2018) 1–24. https://doi.org/10.1186/s12951-018-0408-4.
[16] S.K. Panda, C. Jacob, Preparation of transparent ZnO thin films and their application in UV sensor devices, Solid. State. Electron. 73 (2012) 44–50. https://doi.org/10.1016/j.sse.2012.03.004.
[17] R.N. Aljawfi, M.J. Alam, F. Rahman, S. Ahmad, A. Shahee, S. Kumar, Impact of annealing on the structural and optical properties of ZnO nanoparticles and tracing the formation of clusters via DFT calculation, Arab. J. Chem. 13 (2020) 2207–2218. https://doi.org/10.1016/j.arabjc.2018.04.006.
[18] H. Agarwal, S. Venkat Kumar, S. Rajeshkumar, A review on green synthesis of zinc oxide nanoparticles – An eco-friendly approach, Resour. Technol. 3 (2017) 406–413. https://doi.org/10.1016/j.reffit.2017.03.002.
[19] F.T. Thema, E. Manikandan, M.S. Dhlamini, M. Maaza, Green synthesis of ZnO nanoparticles via Agathosma betulina natural extract, Mater. Lett. 161 (2015) 124–127. https://doi.org/10.1016/j.matlet.2015.08.052.
[20] N. Mayedwa, A.T. Khalil, N. Mongwaketsi, N. Matinise, Z.K. Shinwari, M. Maaza, The Study of Structural, Physical and Electrochemical Activity of Zno Nanoparticles Synthesized by Green Natural Extracts of Sageretia Thea, Nano Res. Appl. 03 (2017). https://doi.org/10.21767/2471-9838.100026.
[21] V.A. Soares, M.J.S. Xavier, E.S. Rodrigues, C.A. de Oliveira, P.M.A. Farias, A. Stingl, N.S. Ferreira, M.S. Silva, Green synthesis of ZnO nanoparticles using whey as an effective chelating agent, Mater. Lett. 259 (2020) 126853. https://doi.org/10.1016/j.matlet.2019.126853.
[22] M. Valian, A. Khoobi, M. Salavati-Niasari, Green synthesis and characterization of DyMnO3-ZnO ceramic nanocomposites for the electrochemical ultratrace detection of atenolol, Mater. Sci. Eng. C. 111 (2020) 110854. https://doi.org/10.1016/j.msec.2020.110854.
[23] B. Müller, Citric acid as corrosion inhibitor for aluminium pigment, Corros. Sci. 46 (2004) 159–167. https://doi.org/10.1016/S0010-938X(03)00191-4.
[24] Y. Miyake, M. Hiramitsu, Isolation and extraction of antimicrobial substances against oral bacteria from lemon peel, J. Food Sci. Technol. 48 (2011) 635–639. https://doi.org/10.1007/s13197-011-0330-3.
[25] A. Priya, S. Hait, Extraction of metals from high grade waste printed circuit board by conventional and hybrid bioleaching using Acidithiobacillus ferrooxidans, Hydrometallurgy. 177 (2018) 132–139. https://doi.org/10.1016/j.hydromet.2018.03.005.
[26] P. Singh, A. Kumar, Deepak, D. Kaur, Growth and characterization of ZnO nanocrystalline thin films and nanopowder via low-cost ultrasonic spray pyrolysis, J. Cryst. Growth. 306 (2007) 303–310. https://doi.org/10.1016/j.jcrysgro.2007.05.023.
[27] M.A. Gondal, Q.A. Drmosh, Z.H. Yamani, T.A. Saleh, Synthesis of ZnO 2

Accepted Manuscript

nanoparticles by laser ablation in liquid and their annealing transformation into ZnO nanoparticles, Appl. Surf. Sci. 256 (2009) 298–304. https://doi.org/10.1016/j.apsusc.2009.08.019.

[28] A. Khataee, A. Karimi, M. Zarei, S.W. Joo, Eu-doped ZnO nanoparticles: Sonochemical synthesis, characterization, and sonocatalytic application, Ultrason. Sonochem. 67 (2020) 102822. https://doi.org/10.1016/j.ultsonch.2015.03.016.

[29] N. Salah, W.M. AL-Shawafi, A. Alshahrie, N. Baghdadi, Y.M. Soliman, A. Memic, Size controlled, antimicrobial ZnO nanostructures produced by the microwave assisted route, Mater. Sci. Eng. C. 99 (2019) 1164–1173. https://doi.org/10.1016/j.msec.2019.02.077.

[30] A.M. Pineda-Reyes, M. de la L. Olvera, Synthesis of ZnO nanoparticles from water-in-oil (w/o) microemulsions, Mater. Chem. Phys. 203 (2018) 141–147. https://doi.org/10.1016/j.matchemphys.2017.09.054.

[31] M. Zare, K. Namratha, K. Byrappa, D.M. Surendra, S. Yallappa, B. Hungund, Surfactant assisted solvothermal synthesis of ZnO nanoparticles and study of their antimicrobial and antioxidant properties, J. Mater. Sci. Technol. 34 (2018) 1035–1043. https://doi.org/10.1016/j.jmst.2017.09.014.

[32] K. V. Chandekar, M. Shkir, A. Khan, B.M. Al-Shehri, M.S. Hamdy, S. AlFaify, M.A. El-Toni, A. Aldalbahi, A.A. Ansari, H. Ghaithan, A facile one-pot flash combustion synthesis of La@ZnO nanoparticles and their characterizations for optoelectronic and photocatalysis applications, J. Photochem. Photobiol. A Chem. 395 (2020) 112465. https://doi.org/10.1016/j.jphotochem.2020.112465.

[33] V. Ganesan, M. Hariram, S. Vivekanandhan, S. Muthuramkumar, Periconium sp. (endophytic fungi) extract mediated sol-gel synthesis of ZnO nanoparticles for antimicrobial and antioxidant applications, Mater. Sci. Semicond. Process. 105 (2020) 104739. https://doi.org/10.1016/j.mssp.2019.104739.

[34] G. Vijayaprasath, R. Murugan, Y. Hayakawa, G. Ravi, Optical and magnetic studies on Gd doped ZnO nanoparticles synthesized by co-precipitation method, J. Lumin. 178 (2016) 375–383. https://doi.org/10.1016/j.jlumin.2016.06.004.

[35] A. Khorsand Zak, W.H. Abd. Majid, M.E. Abrishami, R. Yousefi, X-ray analysis of ZnO nanoparticles by Williamson-Hall and size-strain plot methods, Solid State Sci. 13 (2011) 251–256. https://doi.org/10.1016/j.solidstatesciences.2010.11.024.

[36] P. Bindu, S. Thomas, Estimation of lattice strain in ZnO nanoparticles: X-ray peak profile analysis, J. Theor. Appl. Phys. 8 (2014) 123–134. https://doi.org/10.1007/s40094-014-0141-9.

[37] G. Srinivasan, R.T. Rajendra Kumar, J. Kumar, Li doped and undoped ZnO nanocrystalline thin films: A comparative study of structural and optical properties, J. Sol-Gel Sci. Technol. 43 (2007) 171–177. https://doi.org/10.1007/s10971-007-1574-2.

[38] G.B. Harris, X. Quantitative measurement of preferred orientation in rolled uranium bars, London, Edinburgh, Dublin Philos. Mag. J. Sci. 43 (1952) 113–123. https://doi.org/10.1080/14786440108520972.

[39] J. Vishnu Chandar, S. Shanmugan, P. Murugan, D. Mutharasu, K. Sudesh, Structural Analysis of ZnO Nanoparticles Reinforced P(3HB-co-15 mol% 3HHx) Bioplastic Composite, J. Polym. Environ. 25 (2017) 1251–1261. https://doi.org/10.1007/s10924-016-0893-1.

[40] U.G. Deekshitha, K. Upadhya, A. Antony, A. Ani, M. Nowak, I. V. Kityk, J. Jedryka, P. Poornesh, K.B. Manjunatha, S.D. kulkarni, Effect of Na doping on photoluminescence and laser stimulated nonlinear optical features of ZnO nanostructures, Mater. Sci. Semicond. Process. 101 (2019) 139–148. https://doi.org/10.1016/j.mssp.2019.06.001.

Accepted Manuscript

[41] Y. Caglar, S. Ilican, M. Caglar, F. Yakuphanoglu, J. Wu, K. Gao, P. Lu, D. Xue, Influence of heat treatment on the nanocrystalline structure of ZnO film deposited on p-Si, J. Alloys Compd. 481 (2009) 885–889. https://doi.org/10.1016/j.jallcom.2009.03.140.

[42] M.L. Cui, X.M. Wu, L.J. Zhuge, Y.D. Meng, Effects of annealing temperature on the structure and photoluminescence properties of ZnO films, Vacuum. 81 (2007) 899–903. https://doi.org/10.1016/j.vacuum.2006.10.011.

[43] H. Sarma, K.C. Sarma, B.N. College, A. Dhubri, X-ray Peak Broadening Analysis of ZnO Nanoparticles Derived by Precipitation method, Int. J. Sci. Res. Publ. 4 (2014). www.ijsrp.org.

[44] D. Nath, F. Singh, R. Das, X-ray diffraction analysis by Williamson-Hall, Halder-Wagner and size-strain plot methods of CdSe nanoparticles- a comparative study, Mater. Chem. Phys. 239 (2020) 122021. https://doi.org/10.1016/j.matchemphys.2019.122021.

[45] Y.T. Prabhu, K.V. Rao, V.S.S. Kumar, B.S. Kumari, X-Ray Analysis by Williamson-Hall and Size-Strain Plot Methods of ZnO Nanoparticles with Fuel Variation, World J. Nano Sci. Eng. 04 (2014) 21–28. https://doi.org/10.4236/wjnse.2014.41004.

[46] R. Sivakami, S. Dhanuskodi, R. Karvembu, Estimation of lattice strain in nanocrystalline RuO2 by Williamson-Hall and size-strain plot methods, Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 152 (2016) 43–50. https://doi.org/10.1016/j.saa.2015.07.008.

[47] J.F. Nye, Physical properties of crystals: their representation by tensors and matrices, Oxford university press, 1985.

[48] G.C. Nicholson, Grain Growth in Zinc Oxide 552 (1965) 214–215.https://doi.org/10.1111/j.1151-2916.1965.tb14716.x

[49] R. Bhargava, P.K. Sharma, S. Kumar, A.C. Pandey, N. Kumar, Consequence of doping mediated strain and the activation energy on the structural and optical properties of ZnO:Cr nanoparticles, J. Solid State Chem. 183 (2010) 1400–1408. https://doi.org/10.1016/j.jssc.2010.04.014.

[50] Y.L. Zhang, Y. Yang, J.H. Zhao, R.Q. Tan, P. Cui, W.J. Song, Preparation of ZnO nanoparticles by a surfactant-assisted complex sol-gel method using zinc nitrate, J. Sol-Gel Sci. Technol. 51 (2009) 198–203. https://doi.org/10.1007/s10971-009-1959-5.

[51] D. Jesuvathy Sornalatha, S. Bhuvaneswari, S. Murugesan, P. Murugakoothan, Solochemical synthesis and characterization of ZnO nanostructureswith different morphologies and their antibacterial activity, Optik (Stuttg). 126 (2015) 63–67. https://doi.org/10.1016/j.ijleo.2014.07.138.

[52] I.M. Alibe, K.A. Matori, E. Saion, A.M. Ali, M.H.M. Zaid, The influence of calcination temperature on structural and optical properties of ZnO nanoparticles via simple polymer synthesis route, Sci. Sinter. 49 (2017) 263–275. https://doi.org/10.2298/SOS1703263A.

[53] K. Davis, R. Yarbrough, M. Froeschle, J. White, H. Rathnayake, Band gap engineered zinc oxide nanostructures: Via a sol-gel synthesis of solvent driven shape-controlled crystal growth, RSC Adv. 9 (2019) 14638–14648. https://doi.org/10.1039/c9ra02091h.

[54] B.T. Sone, E. Makamu, H.E.A. Mohamed, O. Oputu, V. Fester, Green-synthesized ZnO via Hyphaene thebaica fruit extracts: Structure & catalytic effect on the ozonation of Coralene Rubine-S2G azo disperse dye, Environ. Nanotechnology, Monit. Manag. 16 (2021) 100515. https://doi.org/10.1016/j.enmm.2021.100515.

[55] R.N. Ali, H. Naz, J. Li, X. Zhu, P. Liu, B. Xiang, Band gap engineering of transition metal (Ni/Co) codoped in zinc oxide (ZnO) nanoparticles, J. Alloys Compd. 744 (2018) 90–95. https://doi.org/10.1016/j.jallcom.2018.02.072.

Accepted Manuscript

[56] N.M. Al-Hada, E.B. Saion, A.H. Shaari, M.A. Kamarudin, M.H. Flaifel, S.H. Ahmad, S.A. Gene, A facile thermal-treatment route to synthesize ZnO nanosheets and effect of calcination temperature, PLoS One. 9 (2014) 2–10. https://doi.org/10.1371/journal.pone.0103134.

[57] K. Sowri Babu, A. Ramachandra Reddy, C. Sujatha, K. Venugopal Reddy, A.N. Mallika, Synthesis and optical characterization of porous ZnO, J. Adv. Ceram. 2 (2013) 260–265. https://doi.org/10.1007/s40145-013-0069-6.

[58] M. Goswami, N.C. Adhikary, S. Bhattacharjee, Effect of annealing temperatures on the structural and optical properties of zinc oxide nanoparticles prepared by chemical precipitation method, Optik (Stuttg). 158 (2018) 1006–1015. https://doi.org/10.1016/j.ijleo.2017.12.174.

[59] M. Abdelfatah, H.Y. Salah, M.I. EL-Henawey, A.H. Oraby, A. El-Shaer, W. Ismail, Insight into Co concentrations effect on the structural, optical, and photoelectrochemical properties of ZnO rod arrays for optoelectronic applications, J. Alloys Compd. 873 (2021) 159875. https://doi.org/10.1016/j.jallcom.2021.159875.

[60] A. George, S.K. Sharma, S. Chawla, M.M. Malik, M.S. Qureshi, Detailed of X-ray diffraction and photoluminescence studies of Ce doped ZnO nanocrystals, J. Alloys Compd. 509 (2011) 5942–5946. https://doi.org/10.1016/j.jallcom.2011.03.017.

[61] P.K. Samanta, T. Kamilya, Optical Properties of Surface Modified ZnO Nanorods, J. Nanoeng. Nanomanufacturing. 4 (2014) 173–176. https://doi.org/10.1166/jnan.2014.1176.

[62] P.K. Samanta, S.K. Patra, A. Ghosh, P.R. Chaudhuri, Visible Emission from ZnO Nanorods Synthesized by a Simple Wet Chemical Method, Int. J. Nanosci. Nanotechnol. 1 (2009) 81–90. http://www.irphouse.com/sci/ijnn.htm.

[63] C. Abinaya, T. Prasankumar, S.P. Jose, K. Anitha, C. Ekstrum, J.M. Pearce, J. Mayandi, Synthetic Method Dependent Physicochemical Properties and Electrochemical Performance of Ni-Doped ZnO, ChemistrySelect. 2 (2017) 9014–9023. https://doi.org/10.1002/slct.201701584.

[64] M. Ramesh, M. Anbuvannan, G. Viruthagiri, Green synthesis of ZnO nanoparticles using Solanum nigrum leaf extract and their antibacterial activity, Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 136 (2015) 864–870. https://doi.org/10.1016/j.saa.2014.09.105.

[65] J.A. Maldonado-Arriola, R. Sánchez-Zeferino, M.E. Álvarez-Ramos, Photoluminescent properties of ZnO nanorods films used to detect methanol contamination in tequila, Sensors Actuators, A Phys. 312 (2020). https://doi.org/10.1016/j.sna.2020.112142.

[66] R. Kumari, A. Sahai, N. Goswami, Effect of nitrogen doping on structural and optical properties of ZnO nanoparticles Effect of nitrogen doping on structural and optical properties of ZnO nanoparticles, Prog. Nat. Sci. Mater. Int. 25 (2015) 300–309. https://doi.org/10.1016/j.pnsc.2015.08.003.

[67] G.S. Thirumoorthy, O. Balasubramaniam, P. Kumaresan, Tetraselmis indica Mediated Green Synthesis of Zinc Oxide ( ZnO ) Nanoparticles and Evaluating Its Antibacterial , Antioxidant , and Hemolytic Activity, (2021) 172–181. htts://doi.org/10.1007/s12668-020-00817-y

[68] A. Abdelkhalek, A.A. Al-askar, Green Synthesized ZnO Nanoparticles Mediated by Mentha Spicata Extract Induce Plant Systemic Resistance against Tobacco Mosaic Virus, (2020).https://doi.org/10.3390/app10155054

[69] I. Journal, Tunable of Optical and Electrical Properties for Pure and Mn, (2016) 1–6.

[70] S. Alamdari, M.S. Ghamsari, C. Lee, W. Han, H.H. Park, M.J. Tafreshi, H. Afarideh, M.H.M. Ara, Preparation and characterization of zinc oxide nanoparticles using leaf

Accepted Manuscript

extract of sambucus ebulus, Appl. Sci. 10 (2020) 1–19. https://doi.org/10.3390/app10103620.

[71] S. Sahoo, G.L. Sharma, R.S. Katiyar, Raman spectroscopy to probe residual stress in ZnO nanowire, J. Raman Spectrosc. 43 (2012) 72–75. https://doi.org/10.1002/jrs.3004.

[72] D. Kumar, M.K. Singh, M.S. Mehata, Exploration of grown cobalt-doped zinc oxide nanoparticles and photodegradation of industrial dye, Mater. Res. Bull. 150 (2022) 111795. https://doi.org/10.1016/j.materresbull.2022.111795.

[73] J.M. Calleja, M. Cardona, Resonant Raman scattering in ZnO, Phys. Rev. B. 16 (1977) 3753–3761. https://doi.org/10.1103/PhysRevB.16.3753.

[74] G.H. Mhlongo, D.E. Motaung, S.S. Nkosi, H.C. Swart, G.F. Malgas, K.T. Hillie, B.W. Mwakikunga, Temperature-dependence on the structural, optical, and paramagnetic properties of ZnO nanostructures, Appl. Surf. Sci. 293 (2014) 62–70. https://doi.org/10.1016/j.apsusc.2013.12.076.

Accepted Manuscript

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/360959600

# Effect of Interference Between Twin Tall Buildings with Recessed Corners on Wind-Induced Pressure

**Article** *in* International Review of Civil Engineering (IRECE) · March 2022

DOI: 10.15866/irece.v13i2.21249

CITATIONS
0

READS
20

**3 authors**, including:

Ritu Raj Ahirwar
Delhi Technological University
**35** PUBLICATIONS **99** CITATIONS

SEE PROFILE

**Some of the authors of this publication are also working on these related projects:**



Wind effects on tall buildings View project



Wind effects on tall buildings having the shape other than the regular building cross sectional shape View project

All content following this page was uploaded by Ritu Raj Ahirwar on 14 June 2022.

The user has requested enhancement of the downloaded file.

Praise Worthy Prize

*International Review of Civil Engineering (I.RE.C.E.), Vol. 13, N. 2*
*ISSN 2036 - 9913* *March 2022*

# Effect of Interference Between Twin Tall Buildings with Recessed Corners on Wind-Induced Pressure

Suresh Kumar Nagar[1], Ritu Raj[2], Nirendra Dev[2]

**Abstract** – *This paper investigates the results of wind tunnel model testing on pressure distribution and proximity effects between twin tall buildings with large recessed corners in an open circuit boundary layer wind tunnel. Experiments were carried out for isolated and for full, half and no blockage interference conditions. Three interference conditions were created by placing the building of identical shape and size at different locations on the windward side of the principal building under consideration. Based on the experimental results, mean and Root Mean Square (RMS) wind pressure coefficients and interference factors are presented and discussed for all cases. Mean wind pressure coefficients and interference factors at all faces of the building model are plotted in the form of contours and also plotted along the height of the building at all faces. Pressure coefficients and interference factors for three interference conditions are also compared with the results of isolated condition to bestow comprehensive evaluation of the effects of the interference on changing of wind loads. Results show that pressure coefficients variation greatly depends on the position of the interfering building. Oblique configuration produces more severe interference effects on windward surfaces. The re-entrant corners present in the model also affect the pressure distribution because of flow stagnation in these corners. The interference effects in no blockage condition are significant and unfavorable at windward faces near recessed corners towards the interfering building, where the mean interference factor of more than 13 is observed.* ***Copyright © 2022 Praise Worthy Prize S.r.l. - All rights reserved.***

***Keywords***: *Tall Buildings, Boundary Layer Wind Tunnel, Interference Effects, Wind Loads, Pressure Coefficient*

## I. Introduction

Due to the shortage of land and availability of advanced materials and technologies, a large number of tall buildings are being constructed in close proximity.

Numerous buildings being constructed are of similar shape and height. Due to the interaction effect, a complex wind flow pattern exists around these buildings. Due to this, the characteristics of wind load are very different from those of isolated buildings. Wind load on buildings may be enhanced or reduced depending on the shape of the building; the external shape of tall buildings plays an influential role in the generation of wind load on high-rise buildings. Guidelines to estimate wind effects on the tall building are available only for isolated buildings only or these standards offer little guidance regarding the proximity effect for unconventional plan-shaped buildings. This entails a detailed study of these buildings with interfering building, experimentally or numerically for wind load. Since a large number of variables that include the shape and size of buildings, their relative positions, wind direction, etc., are involved [1], it is challenging to provide the compendious and generalized set of guidelines for modification in wind load due to the presence of adjacent buildings. Therefore, the proximity effect of adjacent buildings should be appropriately evaluated for the realistic wind load design of structures. Proximity effects have been studied by many researchers over the recent years [2]-[22]. Blessmann [2] studied the influence of the neighboring buildings on the force and torsion of Citibank and Rio building models in a large city center. Khanduri et al. [3] studied the wind flow mechanisms and measured the extent of modifications in wind loads on a square building due to the presence of interfering square building of varying size and height.

Xie and Gu [4] studied the behavior of mean interference effects between two and three tall buildings present in a group. A series of wind tunnel tests were conducted by authors, and shielding and channeling effects were discussed to understand the convolution of the proximity effects. Zhao and Lam [5] investigated the interference effects on mean and fluctuating components of force and moment between five square tall buildings in the wind tunnel, arranging them in L- and T-shaped patterns. Kim et al. [6] studied the interference effects on local peak pressures between two buildings for different wind directions with different locations and height ratios of interfering building. Amin and Ahuja [7] investigated the mean interference effects between two closely spaced tall buildings of rectangular cross-sections arranged in the ‘L’ and ‘T’ plan shapes geometrical configuration.

Hui et al. [8] investigated the mutual interference

*Copyright © 2022 Praise Worthy Prize S.r.l. - All rights reserved* https://doi.org/10.15866/irece.v13i2.21249

effects on local peak pressure coefficients between two high-rise building models of different shapes. Li et al. [9] investigated the distribution of wind pressure and interference effects on a tall building surrounded by tall office and hotel buildings. Hui et al. [10] investigated the interference effects on local peak pressure coefficients between two rectangular-section high-rise buildings of the same shape, arranged in parallel and perpendicular configurations. You et al. [11] investigated the interference effects on mean and Root Mean Square (RMS) pressure coefficients between two square buildings with varying interference distances from 0.5B to 2B at an interval of 0.5B. Kheyari and Dalui [12] carried out an analytical study to investigate the interference effects on a rectangular building due to the presence of a rectangular building of a similar cross-section and height of 1/5th of the principal building. Yu et al. [13] studied the interference effects between two high-rise buildings arranged in tandem, oblique, and parallel configurations, considering various breadth and height ratios of interfering and principal building. Kar and Dalui [15] presented the interference effects on an octagonal plan-shaped tall building due to three square buildings of the same height. Mukherjee and Bairagi [16] present interference effects on the square tall building due to the presence of a setback building. Yu et al. [17] reported the modification in acceleration response of principal building due to the presence of interfering building of different section sizes based on a series of wind tunnel experiments and response studies. Zu and Lam [18] reported the shielding effects on a square tall building provided by a row of medium-rise buildings with various arrangements and variable separation distances. Bailey and Kwok [23] studied the change in the dynamic response of a tall square building due to interference excitation from neighboring tall buildings using wind tunnel model testing. Yan and Li [24] investigated the interference effects between aerodynamically modified twin super tall buildings with sculptured top, cut corners, and tapering. Dongmei et al. [25] studied interference effects on wind load and local wind pressure acting on a square building when an identical interfering building with slightly shorter height was placed at different relative locations around the principal building. Zu and Lam [26] performed wind tunnel tests to investigate the flow excitation mechanism for the interference effects on dynamic wind loads on twin tall buildings arranged in tandem and side by side arrangements at various relative positions. Most of the past studies presented the interference effects between square plan tall buildings [4]-[6], [11], [13], [14], [17], [18], [23], [25], [26], interference effects between unconventional plan shaped tall buildings have rarely been studied [24], [27], [28] in the past. Hence this study focused on the wind-induced interference effects on mean and RMS pressure coefficients between twin plus plan-shaped tall buildings with three different configurations in tandem and oblique arrangements.

Interference factors for mean and RMS pressure coefficients on surfaces of the principal building are presented and discussed for different interference conditions.

The remainder of the article is organized as follows. Section II presents the experimental setup, including characteristics of wind and details of the model. Section III details typical results obtained from the experimental study in isolated and various interference conditions, where the mean and RMS pressure coefficients at different surfaces followed by effects of interference as Interference Factors (IF) are included. The main conclusions of this study are summarized in Section IV.

## II. Experimental Details

### *II.1. Approach Wind Characteristics*

Wind tunnel tests on a tall building model were carried out in an open circuit boundary layer wind tunnel (BLWT) at the Civil Department of IIT Roorkee, Roorkee. The dimensions of the working section of the tunnel were 15 m long with 2 m × 2 m cross-section. The approach wind flow inside the wind tunnel represents terrain category II as per IS 875:part 3 [29]. Fig. 1 shows the velocity profile and the variation of the turbulence intensity of flow measured at the downstream end of the wind tunnel on the top of the turntable. The mean wind speed profile inside the tunnel has a power low index $\alpha$=0.22 and followed well the power law for the current study. The mean wind speed and turbulence intensity at the top of the model was 9.87 m/s and 12%, respectively.

### *II.2. Details of Models*

Two types of building models were considered in this study, one is used as a pressure model of the principal building under consideration, and the second is the interfering building model. The principal building model was made of a 4 mm thick transparent Perspex sheet, whereas the interfering building model was made of wood with the same shape and size without pressure taping. Models were scaled at a geometric scale of 1:300 to study in the wind tunnel with a Plan area 40,000 $mm^2$ and 600 mm in height. The maximum blockage ratio of the building models in the test section is approximately 7.5%, which is acceptable in wind tunnel tests [30]. The geometric scale is decided according to the cross-section of the wind tunnel and the prototype building's height so that the building model represents the prototype building of 180 m height. The plan showing different faces and detailed dimensions is shown in Fig. 2(c). A total number of 196 pressure taps were installed on the walls of the pressure model, which were located at seven different heights levels of 10, 60, 180, 300, 420,540, and 590 mm (Level G-A) from the bottom, as shown in Fig. 2(a) to obtain a proper distribution of wind pressure on all the faces of models. Total 28 pressure taps were at each level (Fig. 2(b)). Each tap is sampled for 10 seconds at a sampling frequency of 200 $H_z$.

Copyright © 2022 Praise Worthy Prize S.r.l. - All rights reserved

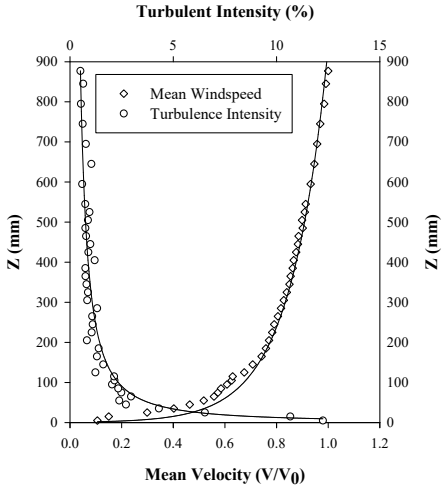
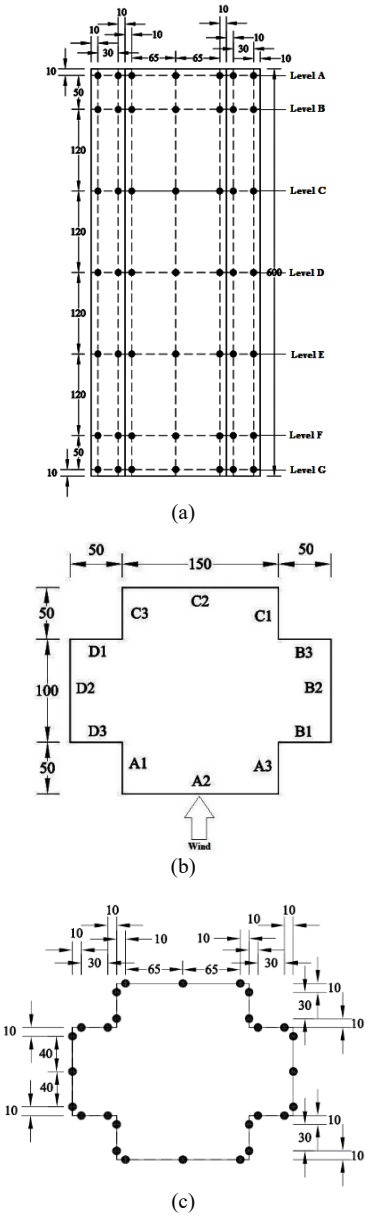
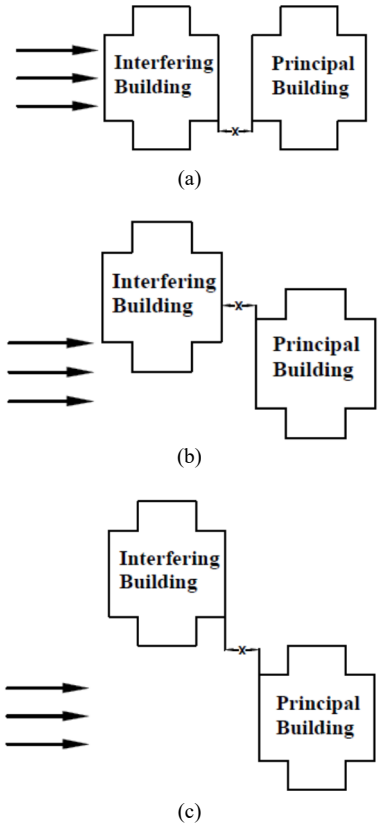



Fig. 1. Mean wind speed and Turbulence Intensity Profile




Figs. 2. (a) Pressure Tap Levels (unit: mm) (b) Plan Dimension and Building Faces (c) Plan of Pressure Tapping




Figs. 3. Interference Blockage Conditions
(a) Full (b) Half (c) No Blockage

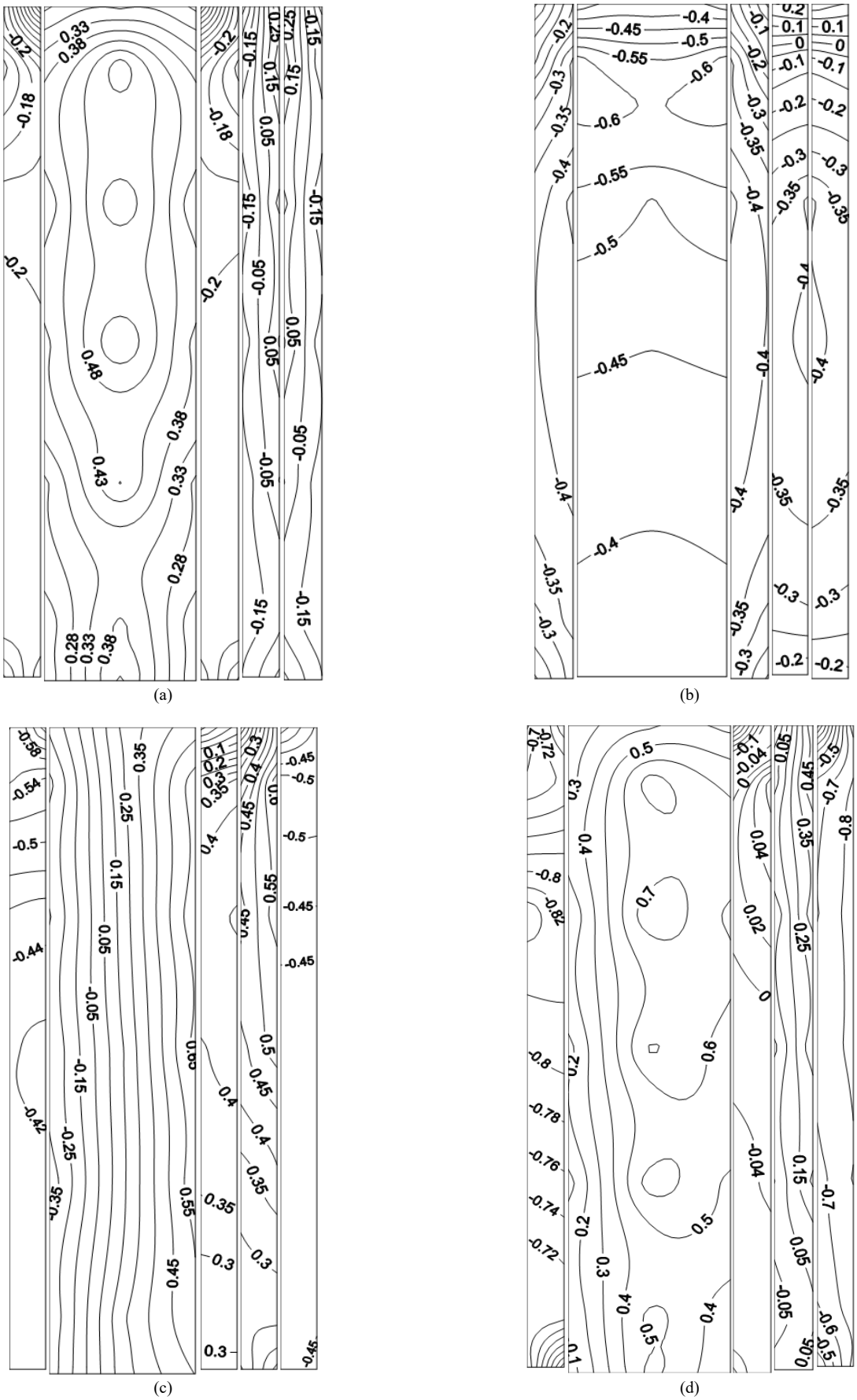
### *II.3. Test Arrangement*

The principal building was tested in isolated and with an interfering building placed at various locations. Face A was a windward face for the isolated building. Figs. 3 show the different configurations of the interfering building. Three interference conditions were created by placing interfering building at various positions. In the first arrangement interfering building was placed in line with the principal building with full blockage interference condition. Distance (x) between the principal and the interfering building was kept equal to 60 mm (1/10th of height of the principal building model) for all cases. In the second and third arrangement, the interfering building was placed in oblique configuration, creating a half blockage and no blockage conditions for the principal building, respectively.

## III. Results and Discussion

It is assumed that at normal flow, wind separated at the edges of the upwind interfering building, and the location of the principal building lies inside the wake of the interfering building. Pressure distribution due to wind on the surfaces of a bluff body is very convoluted for the condition of vortex shedding, flow separation, and re-attached. Characteristics of wind pressure distribution may be described by contour map in the simplest way.

Copyright © 2022 Praise Worthy Prize S.r.l. - All rights reserved




Figs. 4. Distribution of $C_{p,mean}$ at windward surfaces (A1, A2, A3, B1, D3) (a) Isolated (b) Full (c) Half (d) No Blockage

Copyright © 2022 Praise Worthy Prize S.r.l. - All rights reserved

Wind pressure at various points on the surface of the principal building depends upon the position inside the wake; thus, interference effects on different surfaces of the principal building are not similar for similar wind flow conditions.

### *III.1. Mean Wind Pressure Coefficients*

Contours of mean pressure coefficients ($C_{p,mean}$) on the surfaces of principal building for isolated and in the presence of interfering building at different locations are shown in Figs. 4. The pressure distribution on windward surfaces is significantly affected by the presence of the interfering building; hence contour plots of windward surfaces are only presented. From the figure, $C_{p,mean}$ in isolated principal building condition at wind-ward face A2 at the front is positive 0.34 and negative at all other surfaces. Minimum negative $C_{p,mean}$ -0.69, is at side faces B2 and D2. In the presence of the interfering building in full blockage tandem arrangement condition front face A2 of principal building immersed in the wake of interfering building and thus originally positive pressure in isolated condition changes to negative pressure and the negative pressure recover to be positive as the blockage to the principal building due to interfering building reduces to half, but the magnitude of this positive pressure is half of that in isolated condition and recover to be higher than original when interfering building put at oblique no blockage condition as shown in Fig. 4(c).

Nature of $C_{p,mean}$ does not change at all other faces for all interference conditions, except at small side face A3 in half blockage condition, although magnitude has increased significantly at face A1 for all interference conditions and the highest $C_{p,mean}$ -0.76 observed for no blockage. At side front face D3 towards interfering building, the magnitude of suction coefficient $C_{p,mean}$ increased from -0.05 in isolated to -0.65 in no blockage.

The magnitude of pressure coefficient also increased for side front face B1 in full and half blockage. For all other surfaces of the principal building, interference effects are favorable. Variation of mean pressure coefficients are plotted along with the height of the building and compared for isolated and three interference conditions.

Fig. 5 presents the comparative distribution for all interference conditions at all faces. Interference effects at windward faces can be seen from the figure. Parabolic variation is observed at all windward faces for full blockage condition. Face A1 experiences suction for all conditions, and the largest interference effect in the magnitude of mean pressure coefficients between different wind flow conditions is observed at faces A1 and D3. At face A2 distribution is almost coincides with each other for isolated and no blockage conditions.

Distribution trend along height for different conditions are for face A3 and B1. Side faces and leeward faces from B2 to D2 have almost straight variation along the height and do not experience much variation from isolated condition except full blockage condition.

### *III.2. RMS Pressure Coefficients*

The unstable, turbulent and fluctuating nature of wind flow around bluff bodies results in flow separation and sometimes re-attachment and yields fluctuating pressure and forces on bodies in the normal wind. This fluctuating nature of pressure and force may cause the excitation of resonant dynamic response for structure.

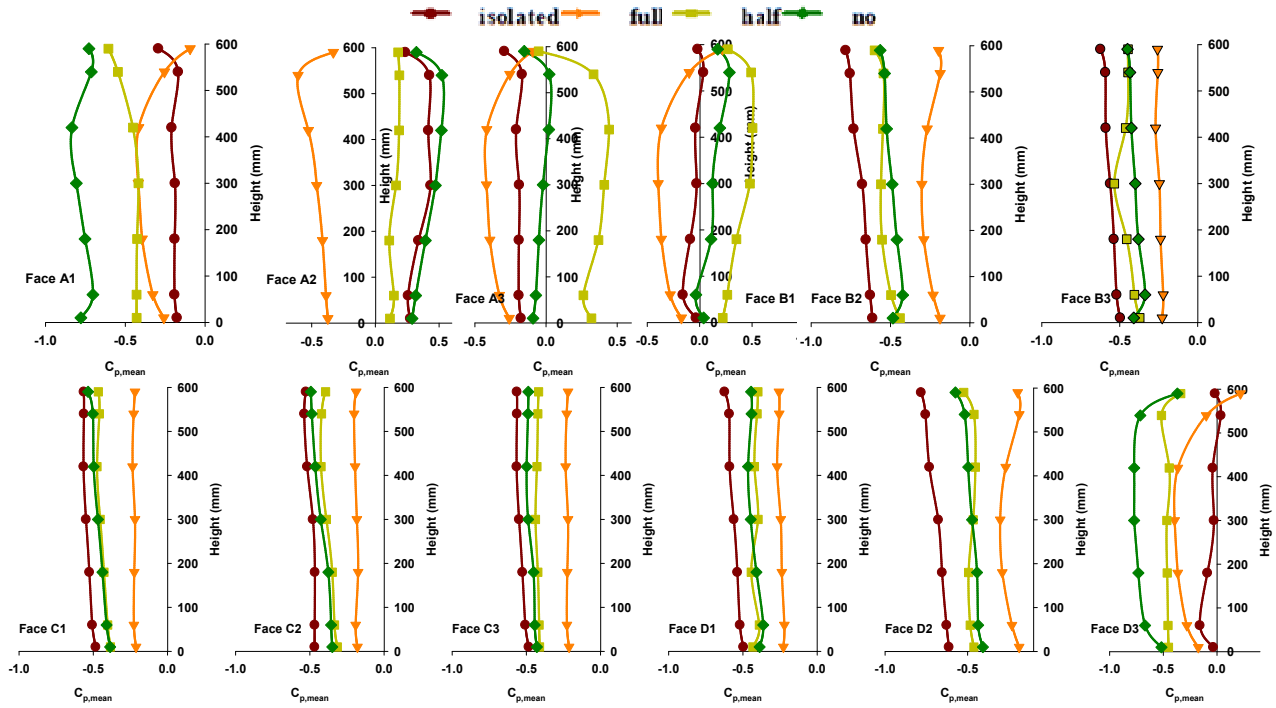
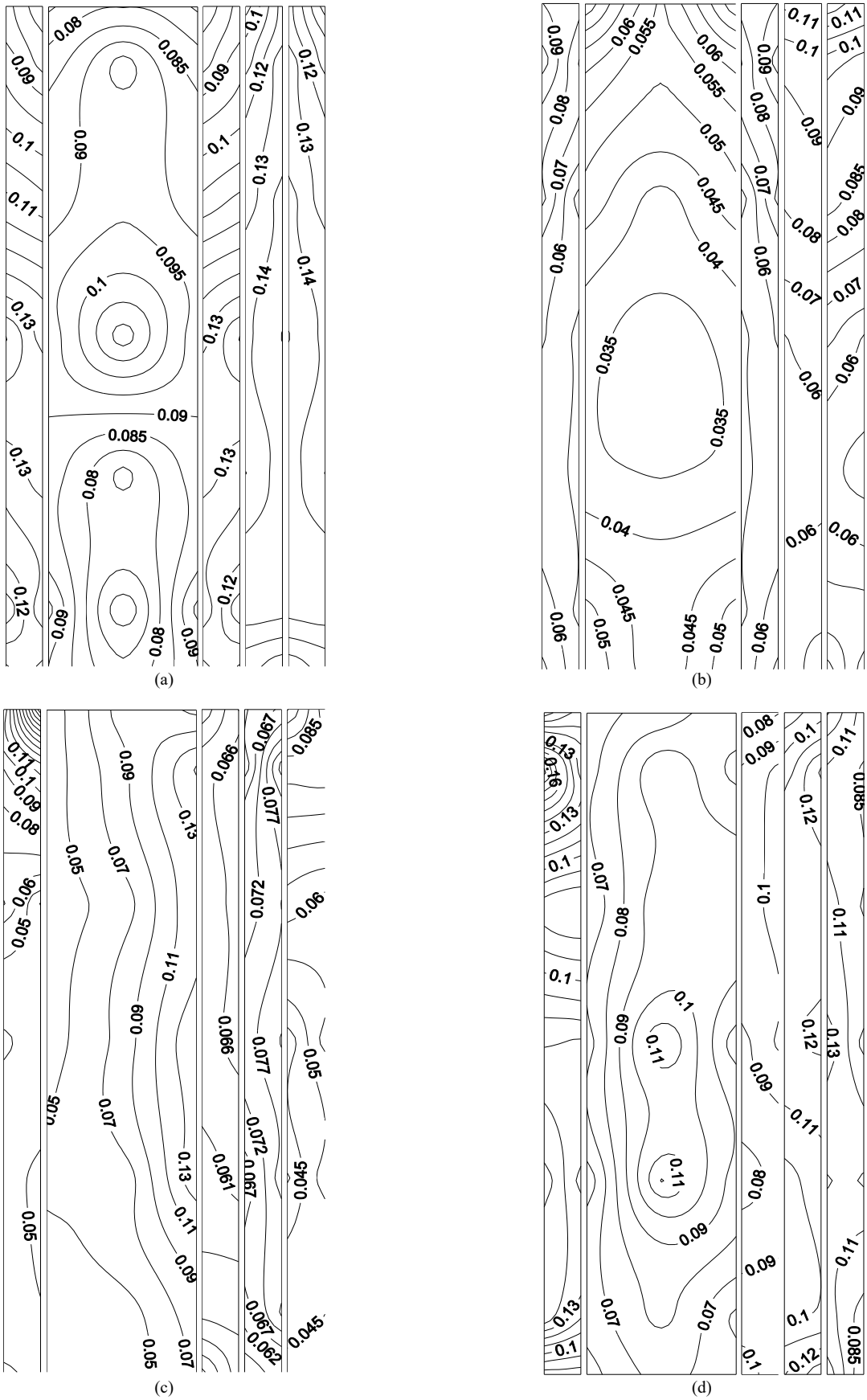



Fig. 5. $C_{p,\,mean}$ at the surfaces along the height of the building

Copyright © 2022 Praise Worthy Prize S.r.l. - All rights reserved




Figs. 6. Distribution of $C_{p,rms}$ at windward surfaces (A1, A2, A3, B1, D3) (a) Isolated (b) Full (c) Half (d) No Blockage

Copyright © 2022 Praise Worthy Prize S.r.l. - All rights reserved

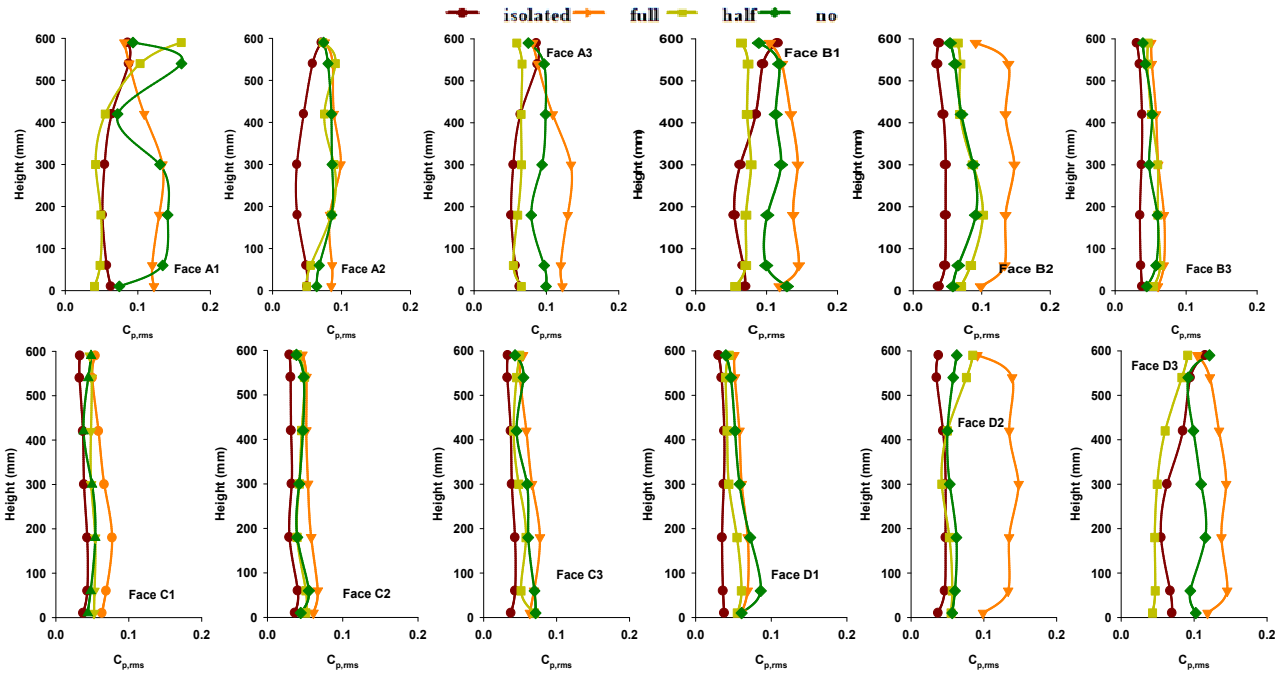



Fig. 7. $C_{p,rms}$ at the surfaces along the height of building

Contour plots of RMS pressure coefficients ($C_{p,rms}$) on surfaces of the principal building are presented in Figs. 6 for various wind flow conditions around the building due to interfering building. The magnitude of fluctuating component is significantly low compared to the mean pressure component at all surfaces, especially with the presence of the interfering building. When the principal building is immersed in the wake of upstream interfering building, reduced RMS pressure coefficients are observed. Maximum $C_{p,rms}$ 0.13 in isolated building case is observed at surfaces connected to the outer front edge where flow gets separated on both sides of the building.

The magnitude of $C_{p,rms}$ is reduced with the presence of the upwind interfering building in full blockage tandem arrangement at all faces compared to that of isolated building. In the half blockage arrangement, RMS pressure coefficients are reduced on all faces from the isolated case but enhanced at the top of side face A1, which may be due to the across wind flow separation at the top edge caused by the presence of an interfering building. $C_{p,rms}$ on this face varies between 0.04 to 0.22.

The fluctuating component of the pressure coefficient is higher for no blockage oblique arrangement compared to previous blockage conditions. Similar to that of mean pressure coefficients, RMS pressure coefficients are plotted along the height of the building and compared for isolated and three interference conditions as shown in Fig. 7. Effect of interference is significant on windward faces, and variation is more along height in RMS pressure coefficient. RMS pressure distribution is arbitrary and does not follow any specific trend along the height. Side and leeward faces do not experience much more variation along the height and also from each other except at side faces B2 and D2 in case of full blockage condition, which may be due to strong shielding created by upwind interfering building.

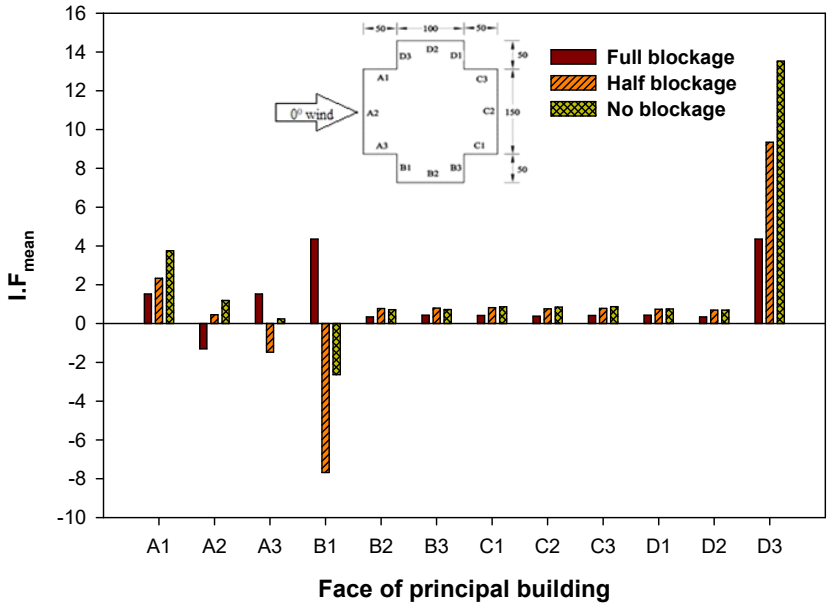
### *III.3. Interference Factor*

This section presents the interference effects on the mean and RMS pressure coefficients of the principal tall building. Interference factors for mean and RMS pressure coefficients are calculated as follows:

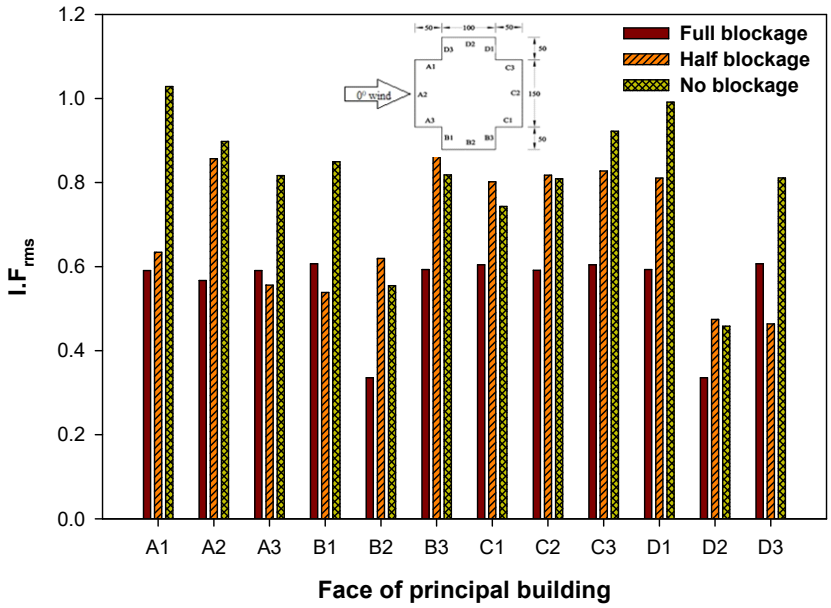
$$\text{Interference factor (I.F)} = \frac{\left(\begin{array}{c}\text{pressure coefficient}\\ \text{under interference condition}\end{array}\right)}{\left(\begin{array}{c}\text{pressure coefficient}\\ \text{in isolated condition}\end{array}\right)} \quad (1)$$

In this section, interference factors are presented and discussed as a bar chart with respect to mean and RMS pressure coefficients on the surfaces of principal building for all three interference conditions. Mean and RMS Interference factors are presented in Figs. 8(a) and 8(b), respectively. From the figure, mean interference effects are critical at windward faces from face D3 to face B1 anticlockwise for all interference conditions. Face A1 experiences a higher interference effect for no blockage condition among all three conditions with a mean interference factor of more than 3. At face A2, interference factor of mean pressure coefficients for full blockage condition is negative with a magnitude of 1.3, shows the change in the nature of pressure from positive to negative due to the presence of the interfering building. Face A3 also has a negative interference factor for half blockage condition, which may be explained by the fact that due to the presence of upwind interfering building flow becomes normal to the side face and creates positive pressure, which was suction in the normal isolated flow.

Copyright © 2022 Praise Worthy Prize S.r.l. - All rights reserved




(a)




(b)

Figs. 8. Interference Factors (a) Mean pressure (b) RMS Pressure

At face B1, interference effects are more severe than previously discussed surfaces. Interference factor is negative with magnitude greater than 7 for half blockage condition. At face D3 largest interference effects are observed among all faces; the interference factor is more than 4 for full blockage condition, which tends to increase up to 13 when the blockage is reduced from full to zero. Interference effects at all side face and leeward faces are favorable with interference factors less than 1.

## IV. Conclusion

Remarkable interference effects may occur between twin tall buildings situated in proximity. Considering the continuous dense arrangement of tall buildings in cities, this problem may be likely more significant in the future.

Interference effects have been analyzed in many studies; however, they usually focus on the regular tall buildings. In this study, wind-induced interference effects between closely spaced twin tall buildings with recessed corners on mean and RMS pressure coefficients have been investigated for various positions of the interfering building present at windward. Local wind pressure in terms of mean and RMS pressure coefficients at various height levels of the tall building has been evaluated.

The results revealed that Pressure distribution is symmetric at symmetrical faces for isolated building condition and full blockage interference condition but does not remain symmetrical for half blockage and no

Copyright © 2022 Praise Worthy Prize S.r.l. - All rights reserved

blockage interference conditions. Interference effects are significant and unfavorable at windward faces. Interference effects are favorable at face anticlockwise from B2 to D2, and pressure reduced up to 65%. In full blockage condition, the nature of pressure changes from positive to negative at face A2 with a slight increase in magnitude. Suction at faces near re-entrant corners has increased significantly. The maximum mean interference factor is more than 4 observed at faces B1 and D3. In half blockage condition, the nature of pressure changes from negative to positive with increased magnitude at face A3 and B1. Suction at faces near re-entrant corners again increased significantly even higher than full blockage condition at faces A1 and D3, which are on the same side of the interfering building. The maximum interference factor is more than 9 at face D3. In no blockage condition, suction at faces D3 increased severely and was highest among all three interference conditions. A mean interference factor of more than 13 is observed. The channeling effect can cause amplification at the D3 face. Interference effects are favorable on fluctuating pressure components. RMS Interference factors for all interference conditions are less than one at all faces. The proximity effects between twin tall buildings with large recessed corners still need more in-depth study and investigations. This study has only considered one direction of the wind. The interference effects with the change of wind direction are required.

Moreover, further research with more interference buildings is needed. In addition, further investigation of the interference effects on aerodynamic forces and wind-induced response, which are important in safety and comfort, also proposed.

## Acknowledgements

The authors would like to express their gratitude to the Department of Civil Engineering, Indian Institute of Technology, Roorkee, for providing the experimental setup.

## References

[1] A.C. Khanduri, T. Stathopoulos and C. Bedard, Wind-induced interference effects on buildings – a review of the state-of-the-art, *Engineering Structures,* vol. 20: 617-630, 1998.

[2] J. Blessmann, Neighbouring wind effects on two tall buildings, *Journal of Wind Engineering and Industrial aerodynamics,* vol. 41-44:1041-1052, 1992.

[3] A.C. Khanduri, T. Stathopoulos and C. Bedard, Generalization of wind-induced interference effects for two buildings, *Wind and Structures,* vol. 3:255-266, 2000.

[4] Z.N. Xie and M. Gu, Mean interference effects among tall buildings, *Engineering Structures,* vol. 26:1173-1183, March 2004.

[5] J.G. Zhao and K.M. Lam, Interference effects in a group of tall buildings closely arranged in an L- or T-shaped pattern, *Wind and Structures,* vol. 11:1-18, January 2008.

[6] W. Kim, Y. Tamura and A. Yoshida, Interference effects on local peak pressures between two buildings, *Journal of Wind Engineering and Industrial aerodynamics,* vol. 99: 584–600, 2011.

[7] J.A. Amin and A.K. Ahuja, Wind-induced mean interference effects between two closed spaced buildings, *KSCE Journal of Civil Engineering,* vol. 16:119-131, May 2012.

[8] Y. Hui, Y. Tamura and A. Yoshida, Mutual interference effects between two high-rise building models with different shapes on local peak pressure coefficients, *Journal of Wind Engineering and Industrial aerodynamics,* vol. 104–106:98–108, 2012.

[9] Y. Li., Q.S. Li and K.L Ju, Experimental Investigation of the Wind Pressure Distribution and Wind Interference Effects on a Typical Tall Building, *Advanced Materials Research,* Vol. 639-640: 444-451, 2013.

[10] Y. Hui, A. Yoshida and Y. Tamura, Interference effects between two rectangular-section high-rise buildings on local peak pressure coefficients, *Journal of fluids and structures,* vol. 37: 120-133, January 2013.

[11] K.P. You, Y.M. Kim and J.Y. You, Interference effect tall building to fluctuations wind load, *Advanced Materials Research,* Vol. 871: 9-14, 2014.

[12] P. Kheyari and S.K. Dalui, Estimation of wind load on a tall building under interference effects: a case study, *Jordan Journal of Civil Engineering,* Vol. 9: 84-101, 2015.

[13] X.F. Yu, Z.N. Xie, J.B. Zhu and M. Gu, Interference effects on wind pressure distribution between two high-rise buildings, *Journal of Wind Engineering and Industrial aerodynamics,* vol. 142: 188-197, 2015.

[14] W Kim, Y. Tamura and A. Yoshida, Interference effects on aerodynamic wind forces between two buildings, *Journal of Wind Engineering and Industrial aerodynamics,* vol. 147:186–201, 2015.

[15] R. Kar and S.K. Dalui, Wind interference effect on an octagonal plan shaped tall building due to square plan shaped tall buildings, *Int J Adv Struct Eng,* vol. 8:73–86, February 2016.

[16] S. Mukherjee and A.K. Bairagi, interference effect on principal building due to setback tall building under wind excitation, *Proceedings of the 11th Structural Engineering Convention,* Kolkata: 13-18, 2018.

[17] X. Yu, Z. Xie and M. Gu, Interference effects between two tall buildings with different section sizes on wind-induced acceleration, *Journal of Wind Engineering and Industrial aerodynamics,* vol. 182: 16-26, 2018.

[18] G.B. Zu and K.M. Lam, Shielding effects on a tall building from a row of and medium rise buildings, *Wind and Structures,* vol. 6:439-449, September 2018.

[19] Nagar S.K., Raj R. and Dev N., Experimental study of wind-induced pressures on tall buildings of different shapes, *Wind and Structures,* Vol. 31 (Issue 5): 431-443, 2020.

[20] Quan Y., Jialu C., and Gu M., Aerodynamic interference effects of a proposed taller high-rise building on wind pressures on existing tall buildings, *Struct Design Tall Spec Build,* 29(4):e1703, 2020.

[21] M.S. Germi and H.E. Kalehsar, Numerical investigation of interference effects on the critical wind velocity of tall buildings. *Structures,* Vol 30: 239–252, 2021.

[22] B.S. Chauhan, A. Chakrabarti, and A.K. Ahuja, Investigation of wind load alteration on rectangular cross-section tall building due to change in relative orientation of interfering building, *Structures,* Vol. 31: 970–981, 2021.

[23] P.A. Baily and K.C.S. Kwok, Interference excitation of twin tall buildings, *Journal of Wind Engineering and Industrial aerodynamics,* Vol. 21:323-338, July 1985.

[24] B. Yan and Q.S. Li, Wind tunnel study of interference effects between twin super-tall buildings with aerodynamic modifications, *Journal of Wind Engineering and Industrial aerodynamics,* vol. 156: 129-145, 2016.

[25] H. Dongmei, Z. Xue, H. Shiqing, H. Xuhui and H. Hua, Characteristics of the aerodynamic interference between two high-rise buildings of different height and identical square cross-section, *Wind and Structures,* vol. 24: 501-528, May 2017.

[26] G.B. Zu and K.M. Lam, Across-wind excitation mechanism for interference of twin tall buildings in tandem arrangement, *Wind and Structures,* vol. 6: 397-413, March 2018.

[27] Pal, S., Raj, R. & Anbukumar, S., Comparative study of wind induced mutual interference effects on square and fish-plan shape tall buildings, *Sādhanā,* Vol 46: 86, 2021.

[28] Pal, S. and Raj, R., Evaluation of Wind Induced Interference

*Copyright © 2022 Praise Worthy Prize S.r.l. - All rights reserved*

Effects on Shape Remodeled Tall Buildings, *Arabian Journal for Science and Engineering*, 1-21, July 2, 2021.
[29] I.S.:875, Part 3, *Indian standard code of practice for design loads (other than earthquake) for buildings and structures*, bureau of Indian standard, New Delhi, 2015.
[30] Choi C.K. and Kwon D.K., Wind tunnel blockage effects on aerodynamic behavior of bluff body, *Wind and Structures,* vol. 1(4): 351-364, 1998.

## Authors' information

[1]Department of Civil Engineering, Rajasthan Technical University, Kota, Rajasthan, India.

[2]Department of Civil Engineering, Delhi Technological University, Shahabad Dolatpur, Delhi, India.

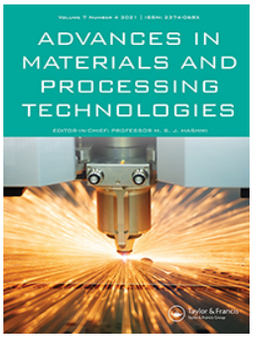
**Suresh Kumar Nagar** (Corresponding Author) is currently working as Assistant Professor in the Department of Civil Engineering, Rajasthan Technical University, Kota and pursuing his Ph.D. from Delhi Technological University, Delhi. He received his B.E from the M.B.M Engineering College, Jai Narain Vyas University, Jodhpur, India in the field of Civil Engineering in 2009 and his M.E in the field of Structural Engineering from Delhi College of Engineering, Delhi University, Delhi, India in 2011. His research interests include Concrete technology, structural analysis and Wind effects on tall buildings.
E-mail: suresh44kota.mbm@gmail.com

Dr. **Ritu Raj** is Assistant Professor in Civil Engineering Department, DTU. He has published more than 50 technical papers in reputed journal and International conferences. He has supervised no. of Students at the level of graduates, postgraduates, and doctorates. His research interest includes behavior of wind load on the response of tall building, Aerodynamics Study, aero-elastic model study inside BLWT, Behavior of high strength concrete with and without fiber, earthquake engineering. He has more than 7.5 years of graduate and post-graduate teaching experience. Teaching Areas: Structure Engineering, Strength of Materials, Steel Structures, Design of Tall Buildings.

Prof. **Nirendra Dev** is Professor in Civil Engineering Department, DTU. He has more than 35 years of academic and educational administration experience. He has published numerous research paper in reputed journal and international conferences. His research interest includes Behavior of high strength concrete with and without fiber, earthquake engineering and concrete technology. He has supervised more than 45 students in gradates, post-graduates and doctorates level.

*Copyright © 2022 Praise Worthy Prize S.r.l. - All rights reserved* 

View publication stats



Advances in Materials and Processing Technologies

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tmpt20

# Effect of preheating on the mechanical and high temperature tribological behaviour of reusing marble dust with aluminium composite

Sumit Chaudhary, Sonia Dangi, Vikas Yadav, Ravinderjit Singh Walia, Narendra Mohan Suri & Mohit Tyagi

**To cite this article:** Sumit Chaudhary, Sonia Dangi, Vikas Yadav, Ravinderjit Singh Walia, Narendra Mohan Suri & Mohit Tyagi (2022): Effect of preheating on the mechanical and high temperature tribological behaviour of reusing marble dust with aluminium composite, Advances in Materials and Processing Technologies, DOI: 10.1080/2374068X.2022.2091091

**To link to this article:** https://doi.org/10.1080/2374068X.2022.2091091

Published online: 06 Jul 2022.

Submit your article to this journal

Article views: 13

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tmpt20

ADVANCES IN MATERIALS AND PROCESSING TECHNOLOGIES
https://doi.org/10.1080/2374068X.2022.2091091

Taylor & Francis
Taylor & Francis Group

Check for updates

# Effect of preheating on the mechanical and high temperature tribological behaviour of reusing marble dust with aluminium composite

Sumit Chaudhary[a], Sonia Dangi[b], Vikas Yadav[c], Ravinderjit Singh Walia[b], Narendra Mohan Suri[b] and Mohit Tyagi[d]

[a]Department of Mechanical Engineering, Thapar Institute of Engineering & Technology, Patiala, India; [b]Production and industrial engineering department, PEC, Chandigarh, India; [c]Department of Mechanical Engineering, Delhi Technological University, Delhi, India; [d]Department of Industrial and Production Engineering, NIT Jalandhar, India

**ABSTRACT**

In this paper the effect on the mechanical and tribological behaviour of aluminium alloy (AA 7175) reinforced with marble dust has been studied. The waste powder of marble developed during its processing, which is a huge environmental waste as its disposal is very difficult. The remarkable wear and friction resistant properties of marble dust make to promising material to be used in MMC. Marble dust used untreated may result in defective MMC due to the presence of moisture and chemical contamination. In this article the marble dust was preheated at a temperature of 150°C and 300°C to remove the volatile chemical contamination and moisture. The composite of preheated marble dust and AA 7175 was formed by stir casing. The optical microscopy and XRD analysis was carried for examining the microstructure and dispersion of marble dust particles in the AA 7175 matrix. The microhardness of the MMC sample improved but the tensile strength slightly decreased. The high-temperature tribological study of the developed composite was performed using the Pin-on-Disc tribometer. The scanning electron microscopy was done to study the worn out surfaces. There was considerable improvement in the tribological behaviour of composite prepared by pre-heating the marble dust.

**ARTICLE HISTORY**
Accepted 12 June 2022

**KEYWORDS**
Aluminium alloy AA 7175; Marble dust; Stir Casting; Metal Matrix Composite; High-temperature Tribology

## 1. Introduction

The developing world has been facing a challenge for the disposal of environmental waste produces due to the overuse of resources. The waste left behind if not degradable than the problem becomes more critical. These environmental wastes sometimes become dangerous for human health as they may pollute the environment. Marble dust has been such a waste that has been produced by the residue left over after cutting of marble. Due to its small size particles then remain suspended in the air for a long time and cause respiratory issues. The utilisation of such environmental waste in some applications would not only reduce the waste accumulation but reduce environmental pollution [1–5]. The

**CONTACT** Sonia Dangi soniadangi21@gmail.com Production and industrial engineering department, PEC, Chandigarh, India

© 2022 Informa UK Limited, trading as Taylor & Francis Group

tribological properties of marble dust can be utilised by producing the composite for various engineering application. The recent research have shown that marble dust can be used as reinforcement with variety of materials like Al 6061 [6], ZA-27 [7], Al-Cu-Ni [8] to enhance the mechanical and tribological performance. The marble dust cannot be alloyed with the metallic material. The marble dust has been included in the matrix material as reinforcement to make the metal matrix composites.

The Metal matrix composites (MMC) has been developed by the blending of two or more materials which can't be alloyed together. The MMC has superior mechanical, thermal, chemical, tribological and electrical properties compared to the base material. Such improved properties can be utilised for various engineering applications [9–11]. There have been various methods of producing MMC like stir casting, ultrasonic casting, FSP etc. FSP has been a very green method for producing MMC as it has negligible material wastage and it doesn't leave any carbon footprints [12].

Aluminium nitride as reinforcement on the copper surface with the friction stir processing to make metal matrix composite to study the effect of AIN particles for different volume % and concluded that microhardness increased and there was a decrease in ductility, wear rate and COF [13]. Titanium oxide as reinforcement in pure copper using FSP improved the anti-corrosive, mechanical and tribological properties by increasing the number of passes [14]. Al 5083 was reinforced by $B_4C$ using FSP resulted in improvement in tensile strength, microhardness and tribological performance when tested on the pin on dice tribometer [15]. The effect of Cr embedded in pure Cu using FSP has been investigated and found enhanced when investigated [16]. FSP was used to enhance the yield strength of Al by adding Ni to the matrix [17]. Similarly, the grain refinement of Al was done using FSP [18]. FSP has been utilised for Al based MCC having both metallic and non-metallic reinforcement [19,20]. The grain refinement of Cu has also been done to achieve the required mechanical properties using multipass FSP [21]. Cu having $B_4C$ as a matrix component resulted in smaller grain size, improved harness and tribological behaviour [22,23,24].

The preheating of base material or the reinforcement has affected the tribo-mechanical behaviour of the product. Kumar and Menghani [25] observed homogeneous mixing and better wetting of reinforcement particles with the base metal when preheated. The preheating of base material surface has also affected the mechanical behaviours of the material [26]. The authors have emphasised on the preheating of reinforcement as it eliminates the moisture and gases produced due to decomposition of material that may result in high porosity of the product [27,28]. The recent studies show that AA has been one of the material which can be processed by stir casting [29,30], FSP/FSW [31] to achieve the enhanced mechanical and tribological performance. This can be further improved by controlled heat treatment [32].

AA has been used in various engineering applications. The material required to have superior tribo-mechanical behaviour so that it could be used for advanced applications. The AA has superior characteristics to produce the MMC. The environmental waste like marble dust can be used in MMC due to its tribological behaviour. The researchers used marble dust to produce MMC but due to the presence of moisture and chemical contamination the MMC produced showed some defects. In this article the moisture

and volatile chemical contamination has been removed from the marble dust by pre-heating it at different temperature. The tribo-mechanical performance of pre-heating and non pre-heated MMC has been studied and compared.

## 2. Materials and methods

### 2.1. Material selection

The Al alloy 7175 was used as matrix material for the study. The material was procured from MNV metals Pvt. Ltd. Delhi. The chemical spectroscopy of the Al alloy 7175 has been carried out to check the elemental composition as shown in Table 1. The major alloying element in the Al matrix was Zinc (≈ 6%), followed by copper and manganese with few traces of Cr.

Marble dust which is an environmental waste produced during the cutting of marble. Marble dust has been an odour less powder having the particle size in microns. The main chemical components of marble dust are oxides of Si, Ca and Mg. There has been an accumulation of marble dust near the marble mining and processing area. This marble dust gets suspended in the air to affect the living organisms in that area, the inhaled marble dust could be very fatal for the human health. The management of such dangerous waste is quite necessary. The marble dust was collected from the marble processing site and ball milled to reduce the particle size. The ball milled marble dust was sieved to obtain the particle size of 600 mesh.

### 2.2. Stir casting

The AA7175 was crushed to reduce its particle size and melted in the induction furnace at 900°C. The furnace was well insulated to avoid the atmospheric gases. The marble dust was preheated at 150°C and 300°C to remove the chemical contamination and moisture present. The MMC was prepared with preheated and non-preheated marble dust with w/w% of 5%. The calculated amount of marble dust was added to the molten AA 7175 and stirred using graphite impeller at 400 rpm for 15 minutes as shown in Figure 1. The molten material was poured in a mould of 150 mm × 80 mm × 8 mm and allowed to solidify.

### 2.3. Material characterisation

The MMC samples were investigated for the microstructural examination of 200× magnification using (RXLr-4 M, Redical, India) optical microscope. The MMC samples were super finished using SiC paper of 220, 440, 600, 1000 and 1500 grit size and then wet diamond polished before the microstructural examination. The surface roughness of the sample was 10 μm after diamond polishing. The samples were washed in ethanol and etched in 50% $HNO_3$ + 50% $H_2O$ for contamination removal and clearer grain structure.

Table 1. Composition of AL ALLOY 7175 obtained by chemical spectroscopy.

| Element | Al | Zn | Mg | Cu | Cr | Others |
|---|---|---|---|---|---|---|
| %wt | 89.6815 | 5.7149 | 2.5221 | 1.6281 | 0.2143 | Remaining |

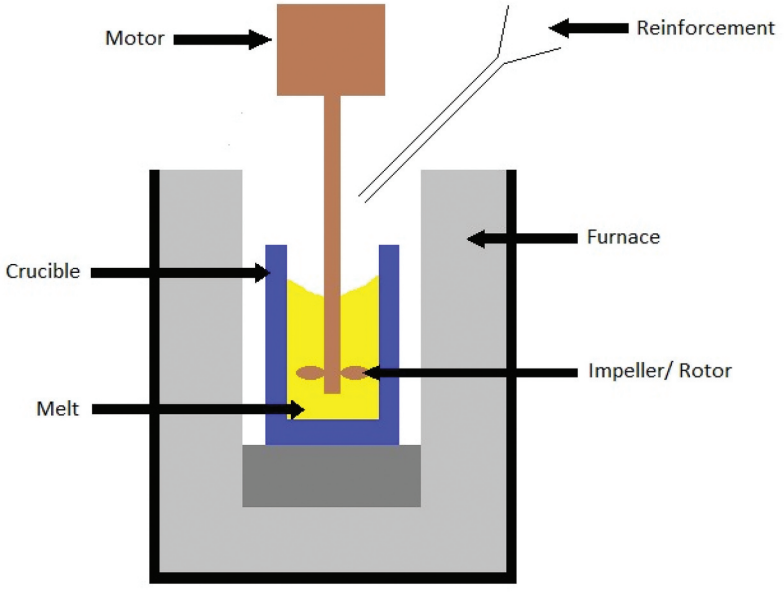



Figure 1. Stir casting schematic.

The phase of MMC material was investigated by the X-Ray diffraction (XRD) (D8, Bruker, USA). The MMC sample was placed under the XRD investigator to a 2θ angle range between 10° to 80°. The principle of operation of XRD is constructive interference of monochromatic X-ray fulfilling Bragg's law. The focused X-rays have been get diffracted and studied to find the crystallographic phases in the material. These crystallographic phases help in detection of composition of material.

The tensile strength of the sample was measured using the UTM. The samples were prepared according to ASTM-E8 using wire cut for the tensile test. The loaded samples showed elongation under tensile load and when the load was increased further it caused the fracture of the material. The fracture was observed in the gauge length of the material as it was having minimum area.

The microhardness of the diamond polished samples was measured on Vikers scale under 300gf load with a dwell time of 20s at 5 different points using the microhardness tester (HM 2000S, Fischer, Germany). The observations were taken 3 times at the same point and average value of microhardness was calculate for accurate results. The morphological analysis of the MMC samples was done using a scanning electron microscope (SEM) (JSM-IT500, Jeol, Japan). The SEM analysis was also conducted after the tribological investigation to predict the nature of wear.

### 2.4. Pin-on-Disc tribometer

The tribological test was conducted on the high temperature pin-on-disc tribometer (TR-20 L-PHM800_DHM850, DUCOM, India) as shown in Figure 2. The tribopair used for the pin-on-disc tribometer has the cylindrical pin that remains stationary and the load is applied

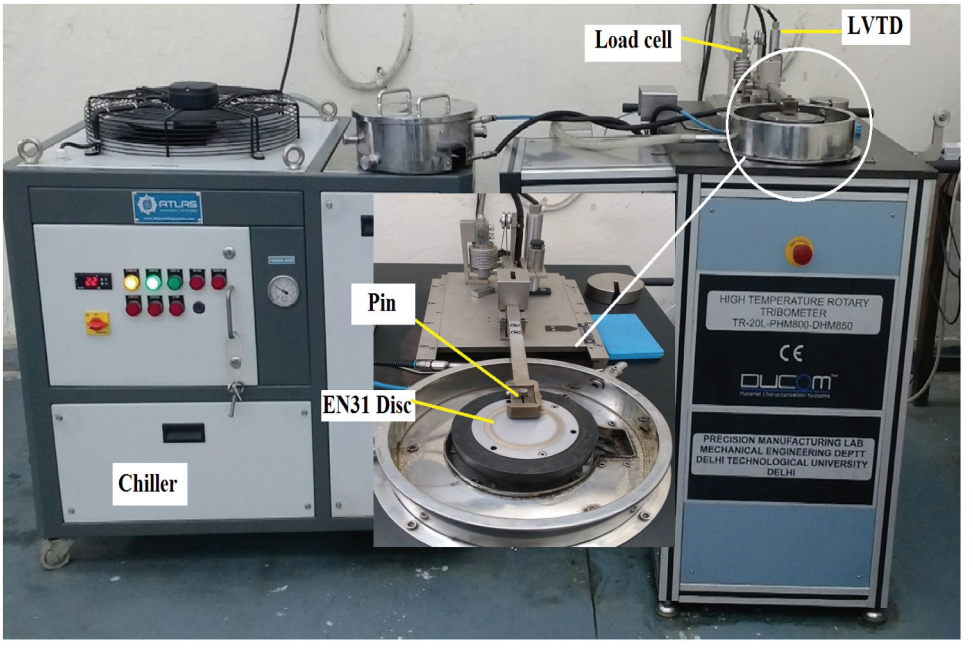



**Figure 2.** High temperature tribometer.

on it. The MMC samples (of circular cross-section) were used as pin material and the EN31 steel disc (microhardness ≈750 HV) was used for the tribological investigation. The disc can be rotated at the desired speed ranging form 300 rpm to 3000 rpm. The load of 40 N was applied on the pin through 1:1 leverage by hanging dead weight. The pin and the disc were super finished to a surface roughness of 1 μm. The test was conducted three times and the average value and COF and wear rate were recorded. The disc had been clamped on the spindle with the help of screw which had been driven by an AC motor through timer belt. The temperature of disc was varied using the induction heating. The holding lever assembly made of a single bar with specimen holder and dead weights. The dead weights had been suspended by wire rope to apply normal load on pin specimen. The fixed pivot point of the lever at centre ensured that load applied at one end of pin get equally distributed. The frictional force produced between pin and disc was measured by the load cell. The LVDT (Linear variable differential transducer) had been used to measure the wear between pin and disc. To achieve 1:1 leverage ratio this sensor is mounted at an exactly the same distance of specimen from the pivot point. The detailed parameters for tribological investigation have been given in Table 2.

**Table 2.** Parameters for tribological investigation.

| S Number | Parameters | Description |
|---|---|---|
| 1 | Disc material | EN 31 |
| 2 | Pin material | AL ALLOY 7175 and MMC (Φ 8 mm) |
| 3 | Load | 40 N |
| 4 | Sliding speed (m/s) | 1, 2 |
| 5 | Temperature (°C) | 100, 200 |
| 6 | Lubricating condition | Dry |

## 3. Results and discussions

### 3.1. Optical microscopy study

The microstructure study of the MMC sample was conducted using the optical microscope. The presence of marble dust in the AA 7175 matrix can be observed as shown in Figure 3. The marble dust was observed to have accumulated at a specific location in the non-preheated sample having larger size particles as shown in Figure 3(a). The agglomeration marble dust particles were due to the enhanced adhesion between the particles in presence of moisture. The escape of moisture left the micro voids on the surface of non-preheated sample. In case of pre-heated sample at 150°C and 300°C samples the marble dust was uniformly distribute throughout the AA 7175 matrix shown in Figure 3(b,c). The preheated samples had lesser voids compared to the non-preheated sample This due to the absence of moisture and chemical contamination in the reinforcement material [33].

### 3.2. Tensile test

The tensile test was conducted according to ASTM-E8. The MMC samples were tested in the perpendicular to the processing. Total length of sample was 70 mm and the gauge length was 25 mm. The width of sample in the gauge length was 6 mm and on the dumbbell part it was 12 mm. The chamfer radius was maintained 12.5 mm as shown in Figure 4(a).

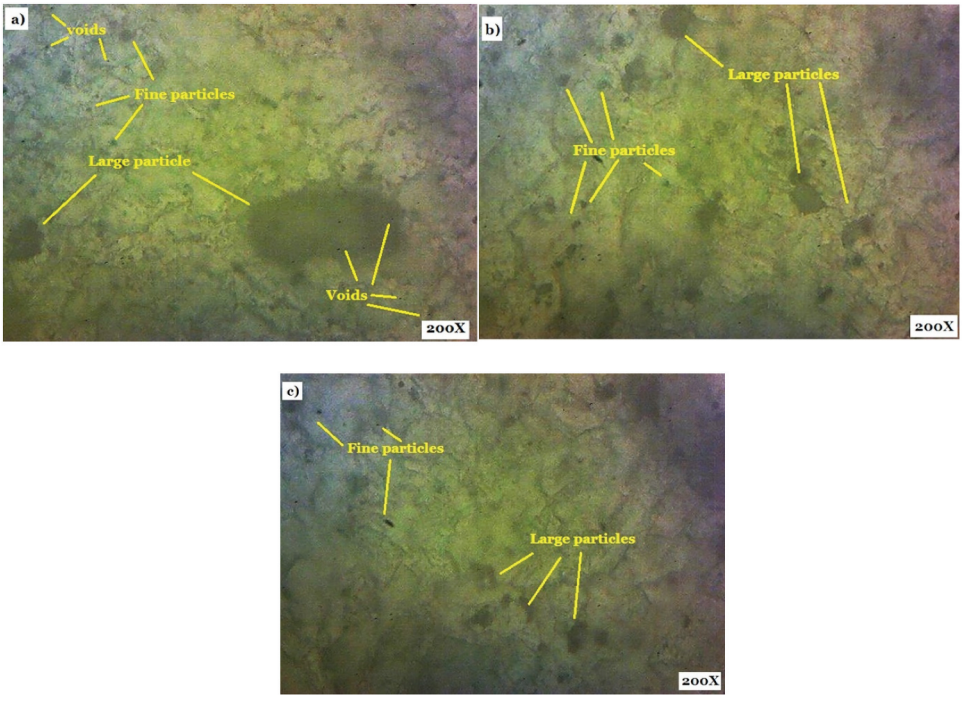



Figure 3. Microstructure study; a) MMC non preheated b) MMC 150°C preheated c) MMC 300°C preheated.

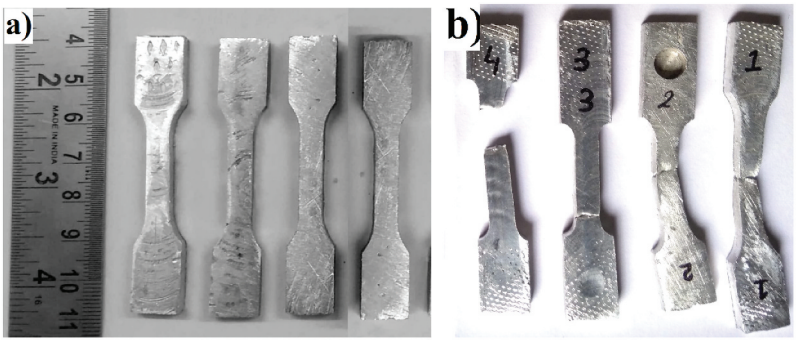



Figure 4. Tensile Test a) Prepared samples for test b) Failed sample after tensile failure.

The tensile strength of base material was comparatively better than the MMC samples as shown in Figure 5. The marble dust is brittle in nature having very low tensile strength might be responsible for the decreased tensile strength of MMC samples. The results reveals that the composite produced have tensile strength 81.1%, 89.35% and 98.17% for non-preheated, pre-heated sample at 150°C and 300°C respectively, with respect to base material. The non-preheated sample had minimum tensile strength due to the presence of marble dust agglomeration, the voids present in the non-pretreated MMC sample were also the culprit for reduction in tensile strength. There has been a considerable improvement in the tensile strength of preheated MMC samples. The preheating enhanced the wettability and bonding between the base material and the marble dust. The decreased amount of voids has also been the reason for increased tensile strength [6].

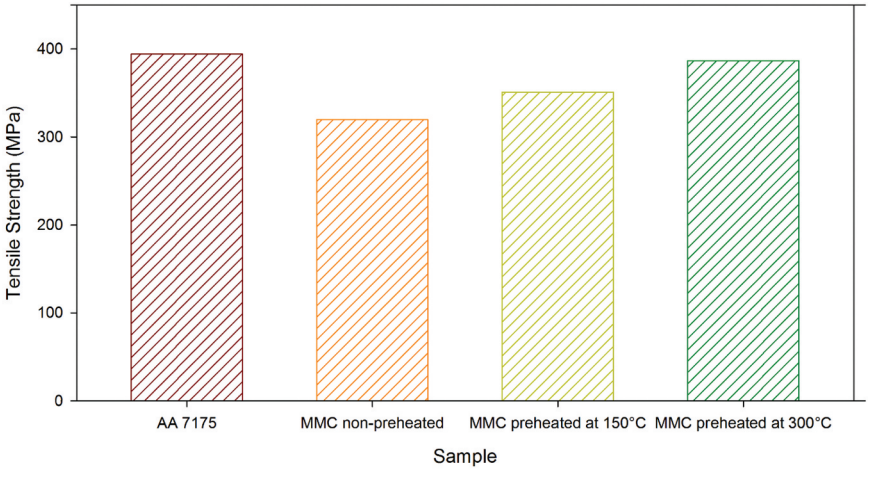



Figure 5. Tensile strength of various samples.

### 3.3. Micro-hardness test

The micro-hardness of the samples was tested on Vickers hardness scale at ten different locations and average value was computed as shown in Figure 6. The marble dust as reinforcement increased the micro hardness. The preheated samples had more uniform microhardness values tested at distinct locations due to the homogeneous distribution of marble dust in the AA 7175 matrix. It provides a positive suggestion that particle dispersion increased the hardness value. The microhardness in case of the MMC pre-heated at 300°C was around 20% more than the AA7175.

### 3.4. SEM and EDS analysis of MMC

The mechanical and tribological characteristics of the material depends on the phase generated during the stir casting. The shape of particles observed in the SEM morphology was irregular shown in Figure 7. The Al was the primary element shown in the EDS. The EDS also shows Ca, O, Si and Mg and their oxides which confirms the presence of marble dust in the matrix. The loose marble dust particles can be observed along with some interfacial bonding with the matrix material. The loose particles were observed with bigger particle size in case of non-preheated samples this is due to the cohesion of particles due to moisture content. As the reinforcement was preheated the wettability increased hence increasing the binding of reinforcement and matrix material.

### 3.5. XRD Test

The XRD pattern shows clear peaks of base material Al as shown in Figure 8. The Al combined with calcium present in the marble dust and formed calcium aluminium silicate which can be observed in XRD of all the MMC prepared. The marble dust resulted in the formation of calcium magnesium silicate, $Ca_2Si$ due to the presence of Silica in the marble

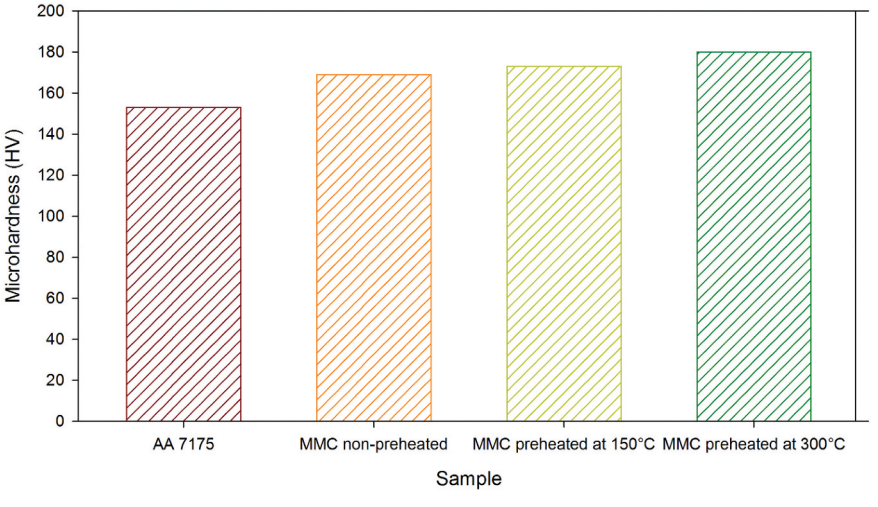



Figure 6. Average microhardness of samples at different locations.

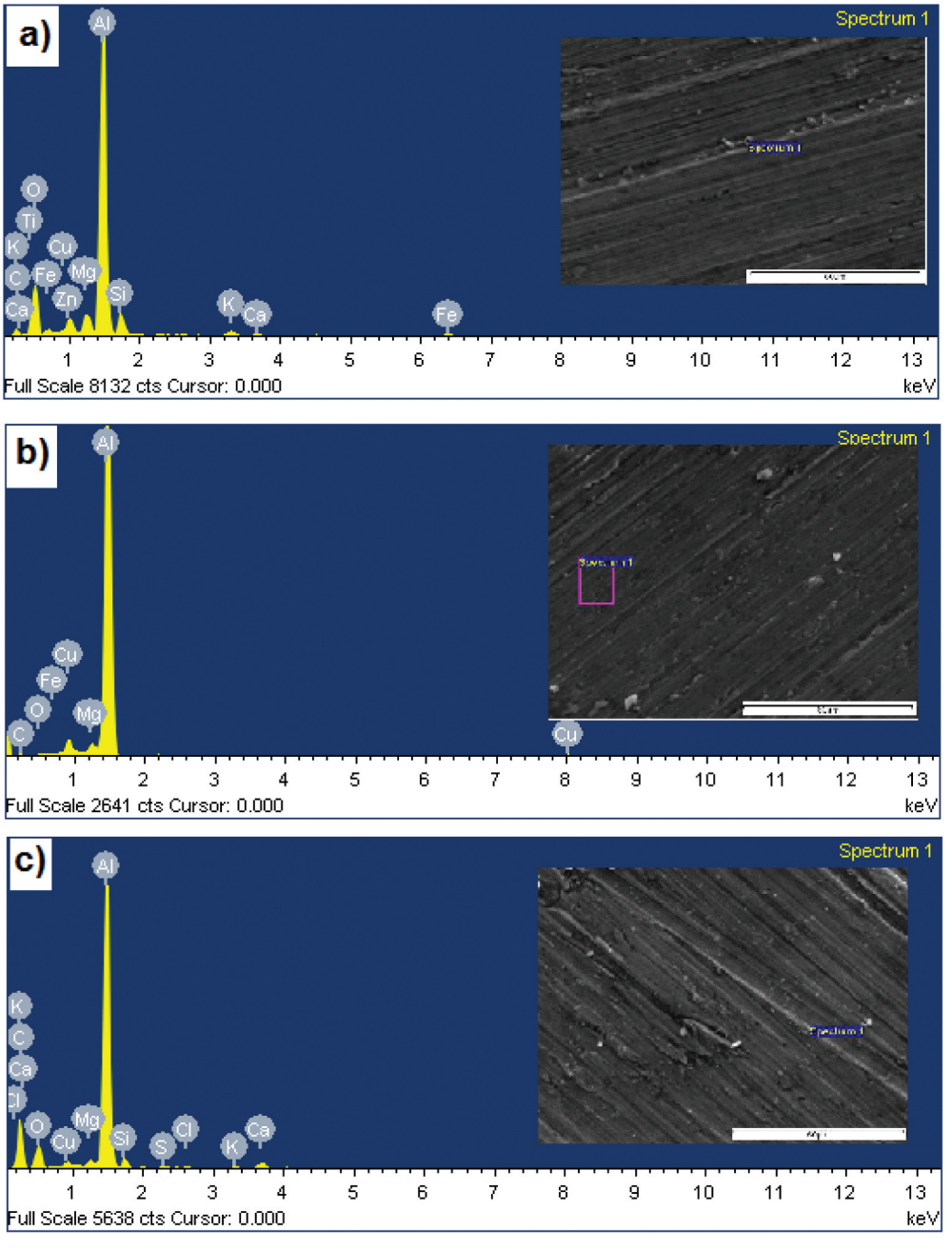



**Figure 7.** SEM-EDS images of sample (a) non-preheated (b) preheated at 150°C (c) preheated at 300°C.

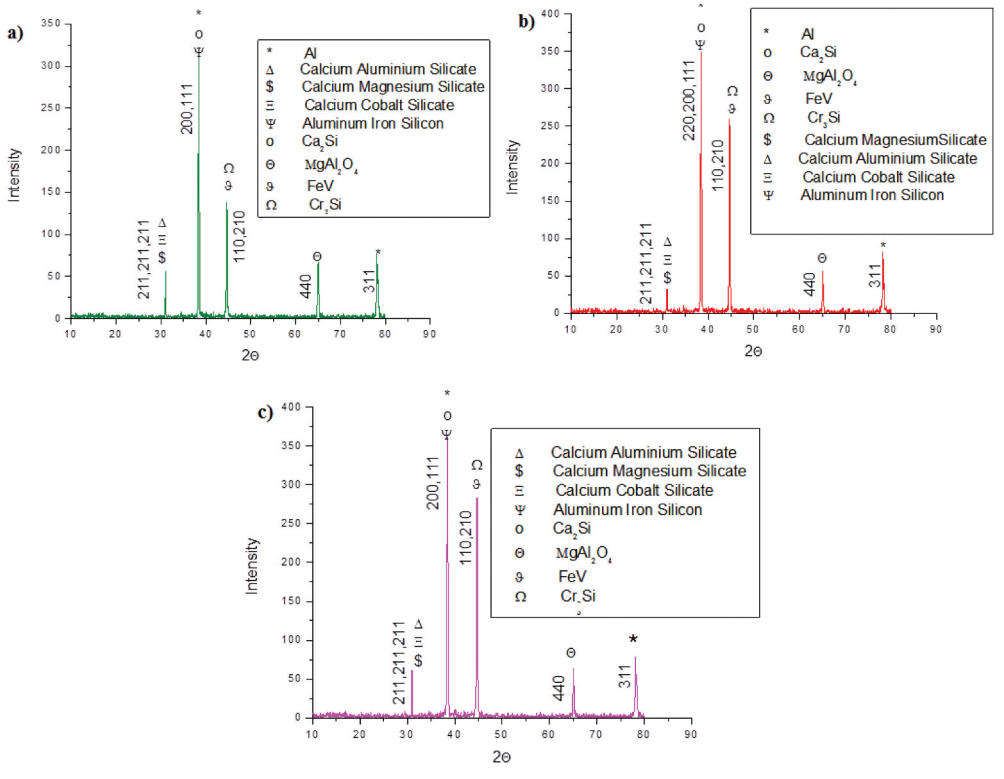



Figure 8. XRD of sample (a) non-preheated (b) preheated at 150°C (c) preheated at 300°C.

dust.There are strong peaks of Aluminium iron silicate and $Cr_3Si$ where observed. The preheated samples had sharper peaks as compared to the non-preheated samples. Moreover, the intensity of stable phases was higher in case of per-heated samples.

### 3.6. High temperature tribological investigation

The high temperature tribological investigation was performed on the Pin-on-disc tribometer. The prepared MMC was used as pin having surface roughness of 10 μm. The temperature of the EN31 plate was increased by induction heating. The coefficient of friction and specific wear rate was calculated for the pin samples. The SEM morphological study of the wear scar on the pin gave an idea about the mechanism of wear.

#### 3.6.1. Analysis of coefficient of friction (COF)

The coefficient of friction for the tribological investigation have been studied as shown in Figure 9. The MMC samples showed better performance as they recorded lesser value of COF.

At 100°C the COF was calculated at 0.5 m/s, 1 m/s and 1.5 m/s. For the base material the COF increased with the increase in sliding speed. This may be due to the more heat generation at the pin surface resulting more stick slip type of effect [15]. The similar increment in COF with sliding speed has been observed for non-preheated and preheated

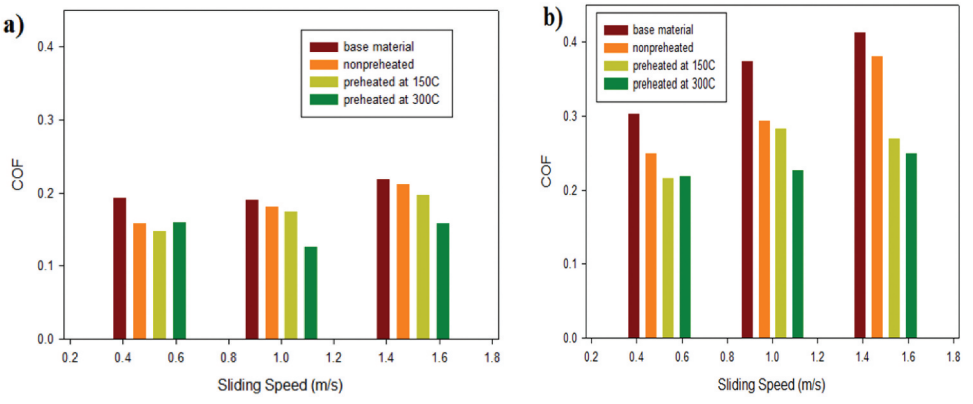



Figure 9. COF vs Sliding Speed a) at 130°C b) at 300°C.

at 150°C as shown in Figure 9 (a). In case of 300°C preheated sample, there has been a reduction in COF at 1 m/s as compared to 0.5 m/s; this decrease in COF has been observed due to the easy sliding assisted by the uniformly distributed marble dust particles in the AL ALLOY 7175 matrix.

At 200°C the COF was calculated at 0.5 m/s, 1 m/s and 1.5 m/s. The COF have higher values at 200°C as compared to 100°C due more localised welding assisted by the temperature rise [6,16]. For the base material the COF increased with the increase in sliding speed. The similar increment in COF with sliding speed has been observed for non-preheated and preheated as shown in Figure 9(b). In case of 150°C preheated sample the COF reduced remarkably. The COF values rise till sliding speed of 1 m/s and then decrease at 1.5 m/s. In case of 300°C preheated sample the COF was observed minimum due to the uniform distribution of marble dust in AL ALLOY 7175 matrix.

### 3.6.2. Analysis of specific wear rate (SWR) and wear track

The specific wear rate for the tribological investigation have been studied as shown in Figure 10. The MMC samples showed better performance as they recorded lesser value of SWR.

At 100°C the SWR was calculated at 0.5 m/s, 1 m/s and 1.5 m/s. For the base material the SWR increased with the increase in sliding speed this may be due to the more heat generation at the pin surface resulting more stick slip type of effect. The similar increment in SWR with sliding speed has been observed for MMC samples Figure 10(a). At 200°C the SWR was calculated at 0.5 m/s, 1 m/s and 1.5 m/s. The SWR have higher values at 200°C as compared to 100°C due to more localised welding assisted by the temperature rise. For the base material the SWR increased with the increase in sliding speed. This may be due to the more heat generation at the pin surface resulting more stick slip type of effect. The similar increment in SWR with sliding speed has been observed MMC samples.

The specific wear has been reduced for the MMC reduced as compared to base material because of its high hardness values. Moreover, the thermal resistance of the material has increased and the thermal conductivity decreased. The base material showed wear wedges and plastic yield (Figure 11(a)) with wear debris scattered all over the wear track which has been the major reason for the increased SWR. The wear

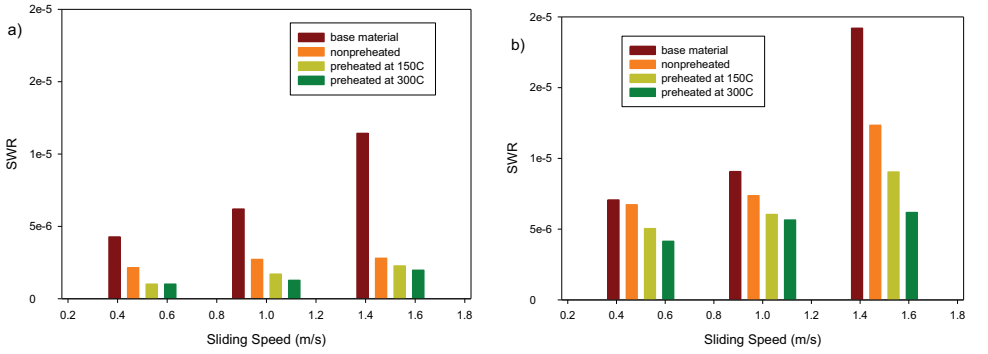



Figure 10. SWR vs Sliding speed a) at 150°C; b) at 300°C.

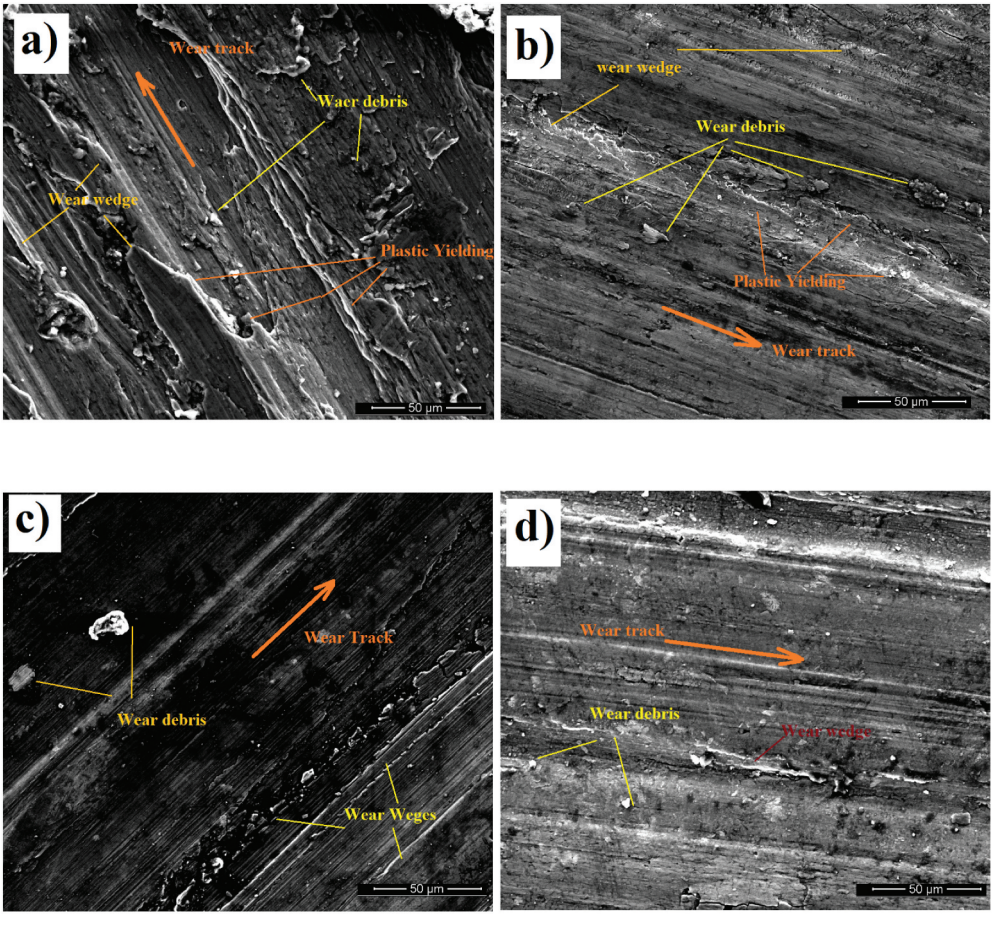



Figure 11. SEM of the wear track on sample after tribological investigation (a) base material (Al ALLOY 7175), (b) non-preheated (c) preheated at 150°C, (d) preheated at 300°C.

wedge and plastic yielding has also been observed at some points on the non-preheated MMC but there are only few sites for these on the wear track (Figure 11(b)) [34–37]. The wear debris have been seen on the wear track. The plastic yielding and wedges were absent in case of preheated MMC. This was due to the formation of hard phase, high hardness and reduced thermal conductivity (Figure 11(c,d)) [6,16]. Few wear debris were observed on the wear track.

## 4. Conclusion

The Al alloy 7175 alloy based metal matrix composite was fabricated using marble dust as reinforcement. The marble dust was pre-heated to eliminate moisture and volatile chemical contamination. The tribological, mechanical and morphological behaviour of prepared specimens was evaluated and then compared with properties of base material.

The MMC showed decreased tensile strength due to the brittleness of marble dust and porosity but this drop was very small in case of pre-heated marble dust as the porosity was less compared to non pre-heated sample. The hardness of sample increased by including marble dust as reinforcement.

The XRD pattern showed clear peaks of base material Al, calcium aluminium silicate, calcium magnesium silicate, $Ca_2Si$, aluminium iron silicate and $Cr_3Si$ that are the evidence of marble dust combined in the aluminium matrix.

The high temperature tribological investigation on Pin-on-Disc tribometer resulted in COF reduction up to 37% and SWR reduction up to 83%. The SEM images shows plastic yielding and wear wedges in case of base material alone with the debris. In case of MMC only wear debris were observed over the wear track.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Data availability statement

The authors confirm that the data supporting the findings of this study are available within the article [and/or] its supplementary materials.

## References

[1] Dinaharan AE. Low cost metal matrix composites based on aluminum, magnesium and copper reinforced with fly ash prepared using friction stir processing. Compos Commun. 2018;9:22–26.
[2] Kumar H, Prasad R, Srivastava A, et al. Utilisation of industrial waste (Fly ash) in synthesis of copper-based surface composite through friction stir processing route for wear applications. J Clean Prod. 2018;196:460–468.
[3] Apasi YDS, Abdulkwereem S, Kolawole MY. Improving mechanical properties of aluminium alloy through addition of coconut shell-ash. J Sci Technol. 2016;36:34–43.
[4] Aruntaş HY, Gürü M, Dayi M, et al. Utilization of waste marble dust as an additive in cement production. Mater Des. 2010;31:4039–4042.

[5] Bilgin N, Yeprem HA, Arslan S, et al. Use of waste marble powder in brick industry. Constr Build Mater. 2012;29:449–457.
[6] Sharma VK, Chaudhary S, Sonia V, et al. Reusing marble dust as reinforcement material for better mechanical performance: studies on compositing aluminum matrix. Mater Res Express. 2020;6:1265f6.
[7] Gangwar S, Patnaik A, Yadav PC, et al. Development and properties evaluation of marble dust reinforced ZA-27 alloy composites for ball bearing application. Mater Res Express. 2019;6:10.1088/2053–1591/ab1362.
[8] Kumar RN, Saravanan R, Sellamuthu R. Effect of marble dust on microstructure and mechanical properties of Al-Cu-Ni/marble dust particles composite. Int J Pure Appl Math. 2018;118:1–12.
[9] Craddock PT. The composition of the copper alloys used by the Greek, Etruscan and roman civilizations: 3. The origins and early use of brass. J Archaeol Sci. 1978;5:1–16.
[10] Zronik J. Meta Shaping our World 10 (Crabtree Publishing Company) . 2005;32.
[11] Heidarzadeh PH, Mahdavi S, Jandaghi M. Ceramic nanoparticles addition in pure copper plate: FSP approach, microstructure evolution and texture study using EBSD. Ceram Int. 2018;44:3128–3133.
[12] Singh L, Singh B, Saxena KK. Manufacturing techniques for metal matrix composites (MMC): an overview. Adv Mater Process Technol. 2020;6(2):441–457.
[13] Thankachan PK. Microstructural, mechanical and tribological behavior of aluminum nitride reinforced copper surface composites fabricated through friction stir processing route. Mater Sci Eng A. 2017;688:301–308.
[14] Heidarpour MY, Roknian M, Ghasemi S. Development of Cu-$TiO_2$ surface nanocomposite by friction stir processing: effect of pass number on microstructure, mechanical properties, tribological and corrosion behavior. J Alloys Compd. 2019;783:886–897.
[15] Yuvaraj N, Aravindan S. Fabrication of Al5083/ B4C surface composite by friction stir processing and its tribological characterization. J Mater Res Technol. 2015;4:398–410.
[16] Khiyavi BA, Aghchai AJ, Arbabtafti M, et al. Effect of friction stir processing on mechanical properties of surface composite of Cu reinforced with Cr particles. Adv Mater Res. 2014;829:851–856.
[17] Ke L, Huang C, Xing L, et al. Al-Ni intermetallic composites produced in situ by friction stir processing. J Alloys Compd. 2010;503:494–499.
[18] Shigematsn KI, Saito N. Mechanical properties of the fine-grained aluminium alloy produced by friction stir process. Scripta Materials. 2003;49:783–789.
[19] Qian J, Li J, Xiong J, et al. In situ synthesizing $Ai_3Ni$ for fabrication of intermetallic-reinforced aluminium alloy composites by friction stir processing. Mater Sci Eng A. 2012;550:279–285.
[20] Shahi A, Sohi MH, Ahmadkhaniha D, et al. In situ formation of Al-$Ai_3Ni$ composites on commercially pure aluminium by friction stir processing. Int J Manuf Technol. 2014;75:1331–1337.
[21] Xue P, Xiao BL, Ma ZY. Achieving large-area bulk ultrafine grained Cu via submerged multiple-pass friction stir processing. J Mater Sci Technol. 2013;29:1111–1115.
[22] Sathiskumar R, Murugan N, Dinaharan I, et al. Characterization of boron carbide particulate reinforced in situ copper surface composites synthesized using friction stir processing. Mater Charact. 2013;84:6–27.
[23] Ghazi JH. Production and properties of Silicon Carbide Particles Reinforcement Aluminium Alloy Composites. Int J Min, Metall Mech Eng. 2013;1:2320–4052.
[24] Ali P, Walia RS, Murtaza Q, et al. Modeling and analysis of developed thermal additive centrifugal, abrasive flow machining process. Surf Topogr Metrol and Prop. 2020;8:035013.
[25] Kumar B, Menghani JV. Aluminium-based metal matrix composites by stir casting: a literature review. Int J Mater Eng Innov. 2016;7:1–14.

[26] Garg A, Bhattacharya A. Friction stir spot welding of AA6061-T6 and Cu with preheating: strength and failure behavior at different test temperatures. Int J Adv Manuf Technol. 2020;108:1613–1629.
[27] Suresha S, Sridhara BK. Effect of silicon carbide particulates on wear resistance of graphitic aluminium matrix composites. Mater Des. 2010;31:4470–4477.
[28] Elsayed A, Nofal A, Attia G. Optimization of stir casting process parameters for producing MMC Aluminum sacrificial anode incorporated with manganese dioxide concentrate powder Int. J of Engg Res Technol. 2020;13:2651–2659.
[29] Karloopia J, Mozammil S, Jha PK, et al. The abrasive wear behavior of in situ processed aluminum alloy metal-matrix composites. In: Srivatsan TS, Rohatgi PK, Hunyadi Murph S, editors. Metal-matrix composites. the minerals, metals & materials series. 2022.
[30] Karloopia J, Mozammil S, Jha PK. Influence of in situ titanium diboride particulate reinforcement on mechanical properties of aluminum–silicon based metal matrix composites. JOM. 2020;72;2927–2936.
[31] Mozammil S, Karloopia J, Verma R, et al. Mechanical response of friction stir butt weld Al-4.5%Cu/TiB2/2.5p in situ composite: statistical modelling and optimization. J Alloys Compd. 2020;826:154184.
[32] Mozammil S, Karloopia J, Jha PK, et al. Effect of heat treatment on mechanical properties of an aluminum alloy and aluminum alloy composite: a comparative study. In: Srivatsan TS, Harrigan WC Jr., Hunyadi Murph S, editors. Metal-matrix composites. the minerals, metals & materials series. Springer; 2021.
[33] Dinaharan SS, Kalaiselvan K, Gopalakrishnan S. Microstructure and sliding wear characterization of Cu/TiB2 copper matrix composites fabricated via friction stir processing. J Asian Ceram Soc. 2017;5:295–303.
[34] Tan YB, Wang XM, Ma M, et al. A study on microstructure and mechanical properties of AA 3003 aluminum alloy joints by underwater friction stir welding. Mater Charact. 2017;127:41–52.
[35] Tailor S, Vashishtha N, Modi A, et al. High-Performance $Al_2O_3$ Coating by Hybrid-LVOF (Low-Velocity Oxyfuel). Process, J Therm Spray Tech. 2020;29:1134–1143.
[36] Khodaverdizadeh H, Heidarzadeh A, Saeid T. Effect of tool pin profile on microstructure and mechanical properties of friction stir processing (FSP). Wear. 2013;304:1–12.
[37] Sahlot P, Jha K, Dey G, et al. Quantitative wear analysis of H13 steel tool during friction stir welding of Cu-0.8%Cr-0.1%Zr alloy. Wear. 2017;378-379:82–89.

*Sādhanā* (2022) 47:126
https://doi.org/10.1007/s12046-022-01895-2
© Indian Academy of Sciences

# Effect of wind load on irregular shape tall buildings having different corner configuration

RAHUL KUMAR MEENA, RITU RAJ* and S ANBUKUMAR

Department of Civil Engineering, Delhi Technological University, New Delhi 110042, India
e-mail: rahulkumar_phd2k18@dtu.ac.in; rituraj@dtu.ac.in; sanbukumar@dce.ac.in

MS received 21 December 2021; revised 10 March 2022; accepted 25 April 2022

**Abstract.** Tall buildings are mainly designed to resist wind loads because wind-induced vibration is more critical for occupant's comfort level on top of the building. Therefore, evaluation of wind-generated pressure is must for such type of tall buildings. Investigation of response due to wind loads on tall buildings is possible by employing numerical modelling such as computational fluid dynamic tools. Regular plan shape structure can be designed for wind load using various international standards and available literatures. In this research comparison is made between four models in which model-A and B is having the regular shape while model-C and D are of irregular shape. These study is useful for the designer to evaluate the response of structures under strong wind conditions. The grid sensitivity and validation study are performed for rectangular shape tall buildings. Validation of the study shows that numerical simulation results are more or less identical with experimental results. In present research work selected building models have equal cross-sectional area and height. Numerical study performed by employing the k-$\varepsilon$ turbulence model and the results are compared with the experimental findings. Out of all the four models, the Y-shape model with round in corners has a minimum base moment and the least coefficient of drag.

**Keywords.** Tall buildings; Irregular shapes; horizontal pressure distribution; Base shear and base moment; CFD; Corner configuration.

## 1. Introduction

Numerous tall buildings are already being constructed or developing all around the world. Peoples are migrating to urban areas now-a-days. In order to accommodate such big population in cities, many lands are occupied by the building authorities for the construction of high-rise building projects. Construction of these high-rise structures becomes difficult when available land have many irregularities in terms of size and shape. Therefore, the construction of tall buildings having irregular cross-sectional shapes is the better idea to overcome these irregular land problems. Wind effects on such irregular shape high-rise structures are the major concern for the structural designer due to unavailability of wind data. Instead of designing the irregular shape building, the design of regular shape tall buildings is possible through different international standards [1–7]. The available international standards are discussing the wind effects on regular shape structure and silent in case of irregular shape structure. The wind induced effects can be obtained on such irregular structures by two major available options like one is wind tunnel testing [8, 9] and other is computational fluid dynamics methods [10–12]. In this research evaluation of wind effects is done using numerical simulation technique i.e. ANSYS CFX.

Wind effects on high-rise structures have been investigated by many researchers using wind tunnel test like Nagar *et al* [13] executed an experimental study on "H" shaped high rise buildings. Kwok [14] investigated the wind induced response on the various shape of tall buildings. Blessmann and Riera [15] performed the test in wind tunnel and obtained the wind induced response on the interfering tall buildings. Pal *et al* [16] performed the experiment in boundary layer wind tunnel on square plane shape model and remodel triangle shape model of high rise structure. Sharma *et al* [17] presented the techniques of reducing wind load by modifying the corner configuration on tall buildings. Yi and Li [18] studied the wind effects on a super tall building using wind tunnel techniques and compared the result on full scale building. Bairagi and Dalui [19] compared the aerodynamic coefficients of set-back tall building for wind load. Bhattacharyya and Dalui [20] performed experiment and simulation on "E" shape tall building to investigate the mean wind pressure. Bearman and Morel [21] studied the effect of free stream turbulence effect of flow around bluff bodies. Raj and Ahuja [22] performed the experiment on a cross shape tall

*For correspondence
Published online: 28 June 2022

building. Raj *et al* [23] studied the response of square and plus shape building by varying wind load. Li *et al* [24] investigated the wind induced torque on "L" shape tall building. Blackmore [25] explored the design of building structures using wind tunnel method and concluded from experimental research that the design data available in various standards are in adequate. During design of tall building, the evaluation of wind load is necessary for such structures whose design data are absent, building having both the side setback is affected more with respect to single side setback. Limited research work done for interference of the building such as Hui *et al* [26] done the experiment in wind tunnel for two rectangular building model. Pal and Raj [27] evaluated the wind induced interference effect on remodeled shape tall building. Zu and Lam [28] investigated the across wind response of interference for twin tall buildings. Nagar *et al* [29] on two plus plan shape tall building. Gaur *et al* [30] investigated the interference effect on corner configured structure using CFD. Important outcomes from these findings such as pressure distribution for isolated building and as well as in full blockage condition. The interference effect is more on the principal building if the interfering building is located on the upstream of wind. Corner cut model configuration helps to reduce the drag force and flow separation at the principal model.

Investigation of wind effects on irregular shape can be possible by using various available tools such as experimental and numerical methods. However, experimental methods are time consuming and having some types of limitations. Computational fluid dynamics techniques are employed to evaluate the wind effect on such irregular shape high rise structure in order to overcome such limitations. Rajasekarababu and Vinayagamurthy [31] investigated wind effects on setback building. Taniguchi and Akamine [32] analyzed the wind pressure coefficient on detached houses in a dense residential block. Zheng *et al* [33] predicted the improvement of inflow boundary conditions in the numerical simulation to investigate the wind flow around the tall building. Haque *et al* [34] estimated the flow fields around rectangular cylinder under turbulent flow by large eddy simulation turbulence model. Gaur and Raj [35] numerically investigated the wind loads effects on a square shape building model. Mou *et al* [36] performed a numerical simulation to observe the variation in pressure distribution on various building models. Hu *et al* [37] numerically observed the mean wind speed and turbulence intensity effect on tall building using shear stress transport and k-ω turbulence model. Blocken *et al* [38] provided recommendation for accurate CFD simulation for atmospheric boundary layer flow. Bairagi and Dalui [39] evaluated wind generated effect on stepped tall building. Wind induced effect are investigated by various researcher like Kumar and Raj [40] on octagonal plan oval shape building, Okajima [41] on rectangular cylinder, Mallick *et al* [42] on "C" shape, Stathopoulos and Baskaran [43] discussed wind environmental conditions around tall buildings, Cheng *et al* [44] evaluated the wind pressure effect and aerodynamics forces on tall building of square shape, Bhattacharjee *et al* [45] on butterfly plan shape tall buildings, Tang *et al* [46] numerically simulated on polygonal buildings, Zheng and Zhang [47] investigated the drag effects on the controlled suction on high rise building, Bairagi and Dalui [48] on setback building, Raj [49] analyzed the response of plus shaped tall building with different bracing system under wind load, the notable outcomes from these numerical studies are as mean pressure on wind ward face will have positive values while other faces will be under the impact of negative pressure. The height and distance of neighboring buildings contribute significantly to external wind flow patterns. Wake length is the governing factor that controls the drag force on high-rise structures. Numerical computations time can be reduced by employing shear stress transport and k-ω turbulence model and Reynold's number is strongly dependent on the breadth to width ratio of the building model. Pressure distribution patterns are mainly influenced by wind flow at different angle and available terrain surrounding the tall structures. Wind fluctuation increases with the increment in the opening in the high-rise structure and generated drag force decreases by twisting the building model shape. The influence of wind can be reduced on high-rise building by controlling the suction. Building with double side setback is efficient to maintain the velocity at the pedestrian level.

In this research study, two corners modified rectangular regular building model and two corners modified Y-shape irregular building models are considered having equal cross-sectional area and height. Corners having the modification such as corner cut, chamfer and fillet etc. helps to reduce the wind impact on high rise structure because of that in this research a comparative study on chamfer and fillet corners are selected. The numerical analysis on such shapes of high-rise building is performed using ANSYS CFX. A very few amount research is done on such type of comparison and this comparison will help a structural designer to choose the building shape having equal cross-sectional area.

## 2. Methodology

The numerical investigation is performed on the regular and irregular shape building model using ANSYS CFX, before the starting of the work validation study was carried out and the results are found similar to the experimental results and various international standards.

### 2.1 *Numerical simulation*

Numerical simulation is performed as per the guidelines available in ANSYS CFX Solver modeling guide [50].

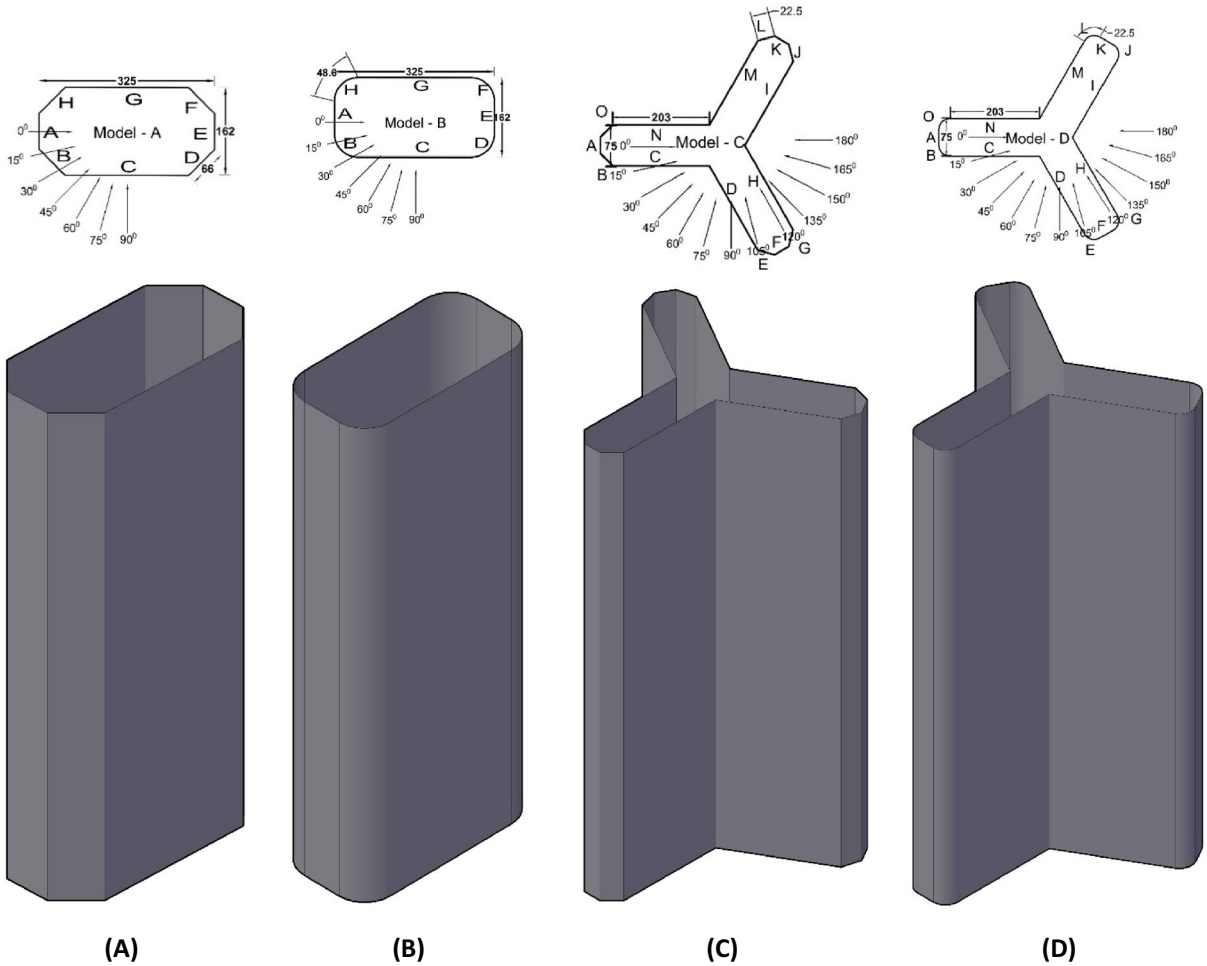



**Figure 1.** Building model in plan and isometric view.

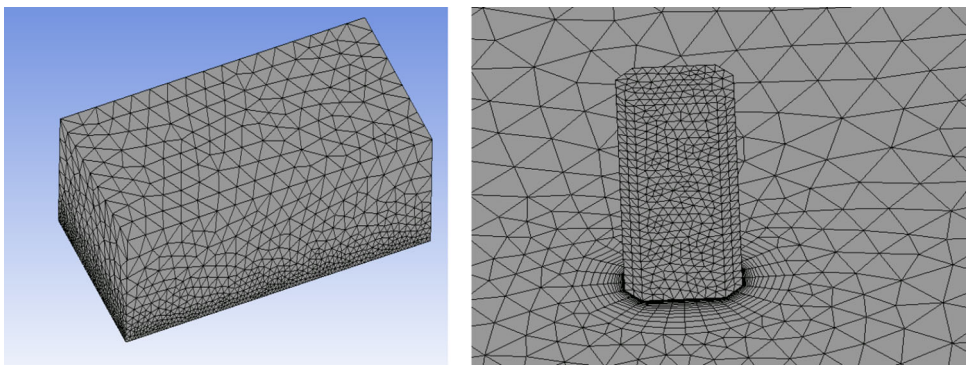


**Figure 2.** Description of meshing.

Some basic equations are utilized in numerical simulation, such as the Navier Stock equation and continuity equation.Navier Stokes equation

$$\frac{\partial(\rho u_i)}{\partial t} = -\frac{\partial(\rho u_i u_j)}{\partial x_j} - \frac{\partial P}{\partial x_j} + \frac{\partial}{\partial_{x_j}}\left[\mu\left(\frac{\partial_{u_i}}{\partial_{x_j}} + \frac{\partial_{u_j}}{\partial_{x_i}}\right)\right] + F \quad (1)$$

Acceleration term = −Convection term
− Pressure gradient
+ Effects of viscosity + Body force

Continuity equation

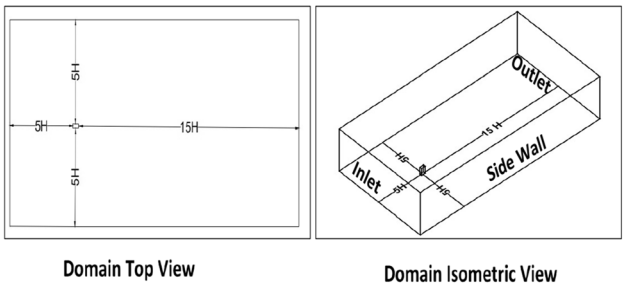



**Figure 3.** Domain.

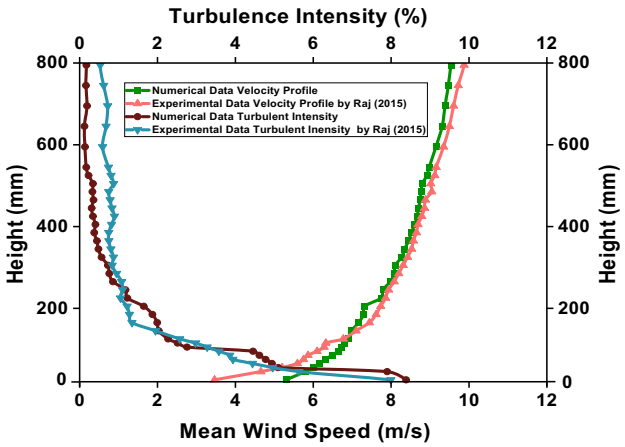



**Figure 4.** Mean wind speed and turbulent intensity profile numerical simulation and experimental.

$$\frac{\partial_\rho}{\partial_t} + \frac{\partial_{\rho_i}}{\partial_{x_i}} = 0 \tag{2}$$

Eddy viscosity

$$\mu_{t=}\rho c_\mu \frac{k^2}{\in} \tag{3}$$

where $c_\mu$ is the dimensionless constant;

This research study is performed by utilizing the k-ε turbulence model, which is two-equation model.

The standard $k$-$\varepsilon$ model uses the following transport equations for the turbulence kinetic energy and turbulence dissipation rate:

$$\frac{\partial(\rho k)}{\partial t} + \frac{\partial}{\partial x_j}(\rho U_j k) = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_k}\right)\frac{\partial k}{\partial x_j}\right] + P_k - \rho\varepsilon + P_{kb} \tag{4}$$

$$\frac{\partial(\rho\varepsilon)}{\partial t} + \frac{\partial}{\partial x_j}(\rho U_j \varepsilon) = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_k}\right)\frac{\partial \varepsilon}{\partial x_j}\right] + \frac{\varepsilon}{k}(C_{\varepsilon 1}P_k - C_{\varepsilon 2\rho\varepsilon} + C_{\varepsilon 1}P_{\varepsilon b}) \tag{5}$$

where $C_{\varepsilon 1}, C_{\varepsilon 2}, \sigma_k$ *and* $\sigma_\varepsilon$ are constant; $P_{kb}$ and $P_{\varepsilon b}$ Represent the influence of the buoyancy forces; $P_k$ Turbulence production due to viscous forces, which is modeled using.

$$P_k = \mu\left(\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i}\right)\frac{\partial U_i}{\partial x_j} - \frac{2}{3}\frac{\partial U_k}{\partial x_k}\left(3\mu_t\frac{\partial U_k}{\partial x_k} + \rho k\right) \tag{6}$$

## 2.2 *Model*

The irregular shape building models are depicted in figure 1 with dimensions and wind incidence angles. The aim of this study is to investigate the wind effects on equal area model by keeping the same ratio of modifications. For these purpose the dimension of such corners and the height of buildings model are same. Models are prepared in ANSYS, CFX design modular by designing the geometry and then extrude of 750 mm height is applied for each case of the tall building model.

## 2.3 *Meshing*

Meshing plays a very important role in numerical simulation and done as per the available guidelines in ANSYS Meshing User's Guide [51], meshing is of two types, structured that follows a fix pattern of mesh and this is generally hex or quad meshing. Unstructured mesh that does not follow the uniform pattern and sometimes it may produce the irrelevance results. Poor meshing will give error in numerical simulation or will require more computational resources also there may be a chance of bad solution thus for this study the meshing adopted in this numerical simulation is demonstrated in figure 2. The

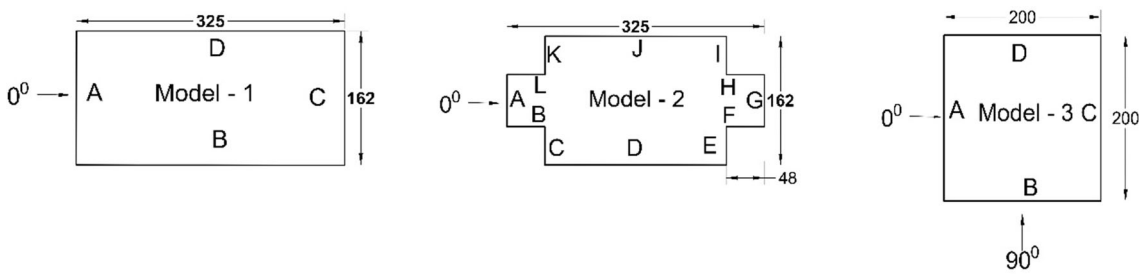



**Figure 5.** Validation model with dinemsions.

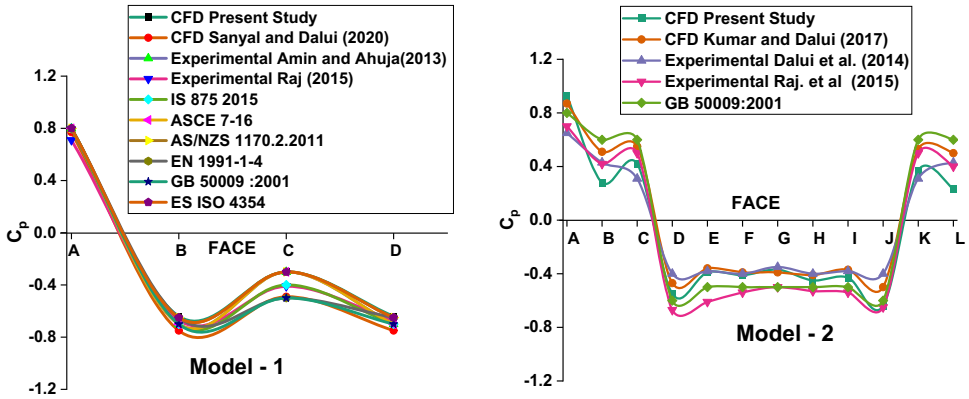



**Figure 6.** Results validation with experimental and different international standards.

**Table 1.** Comparison of face average pressure coefficient ($C_p$) on the model-3 of tall building.

| International code | Wind angle | Windward side | Side wall | Leeward side |
|---|---|---|---|---|
| CFD Results | 0° | 0.78 | −0.64 | −0.30 |
| | 90° | 0.7 | −0.61 | −0.45 |
| Experimental Result (Raj. 2015, Pal *et al*, and Nagar *et al*) | 0° | 0.71 | −0.67 | −0.41 |
| | 90° | 0.73 | −0.66 | −0.42 |
| IS 875 (Part 3) | 0° | 0.7 | −0.7 | −0.4 |
| | 90° | 0.8 | −0.5 | −0.1 |
| ASCE/SEI 7-16 | 0° | 0.8 | −0.7 | −0.5 |
| | 90° | 0.8 | −0.7 | −0.5 |
| AS/NZS 1170.2.2011 | 0° | 0.8 | −0.65 | −0.5 |
| | 90° | 0.8 | −0.65 | −0.5 |
| EN 1991-1-4 | 0° | 0.8 | −0.5 | −0.7 |
| | 90° | 0.8 | −0.5 | −0.7 |
| BS 6399-2 | 0° | 0.8 | −0.5 | −0.7 |
| | 90° | 0.8 | −0.5 | −0.7 |
| GB 50009-2001 | 0° | 0.8 | −0.5 | −0.7 |
| | 90° | 0.8 | −0.5 | −0.7 |
| NSCP 2015 | 0° | 0.8 | −0.5 | −0.7 |
| | 90° | 0.8 | −0.5 | −0.7 |
| ES/ISO 4354: 2012 | 0° | 0.8 | −0.65 | −0.7 |
| | 90° | 0.8 | −0.65 | −0.7 |

**Table 2.** Grid convergence test result for model-1.

| | | | Mean external pressure | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Face | | | | % Error | | | |
| Name | Type of meshing | No of elements | A | B | C | D | A | B | C | D |
| GC-1 | Coarse | 957324 | 0.59 | −0.49 | −0.24 | −0.49 | 36% | 32% | 26% | 32% |
| GC-2 | Medium-1 | 1284687 | 0.73 | −0.56 | −0.27 | −0.56 | 10% | 16% | 12% | 16% |
| GC-3 | Medium-2 | 1439589 | 0.78 | −0.64 | −0.30 | −0.64 | 3% | 2% | 0% | 2% |
| GC-4 | Medium-3 | 1561140 | 0.81 | −0.69 | −0.32 | −0.69 | 1% | 5% | 6% | 5% |
| GC-5 | Fine | 2497236 | 1.02 | −0.98 | −0.43 | −0.98 | 33% | 33% | 30% | 33% |

meshing on building is finer than the ground meshing and domain meshing while inflation is also provided to capture the flow behaviour more accurately. Very fine meshing will need more time and high-end computational resources to perform the simulation and it's not mandatory that very fine meshing will generate the accurate solution while this may produce the irrelevant results.

## 2.4 *Domain*

The domain is the environment where the numerical simulation is performed. Various recommendations are available for constructing the domain such as Zidan *et al* [52] optimize the fluid domain in CFD simulation for tall buildings, domain used in this study is presented with dimensions in figure 3, this domain is constructed on the recommendation provided by Revuz *et al* [53] also the most of the studies in the past is already produces the numerical simulation results on the basis of domain presented in figure 3.

The sidewall and inlet of the domain are kept at 5H. The outlet is at 15H and height of the domain is at 6H, where H is the height of the building model, such large domain is constructed in the design modular for proper wake generation behind the building model. The sidewall and top are assigned as free slip. Ground of the domain where building model is placed and the wall of the building models are assigned as no-slip wall, free slip wall means that velocity is same as that of the inlet and no-slip means that the velocity distributed as per the power low defined at the inlet.

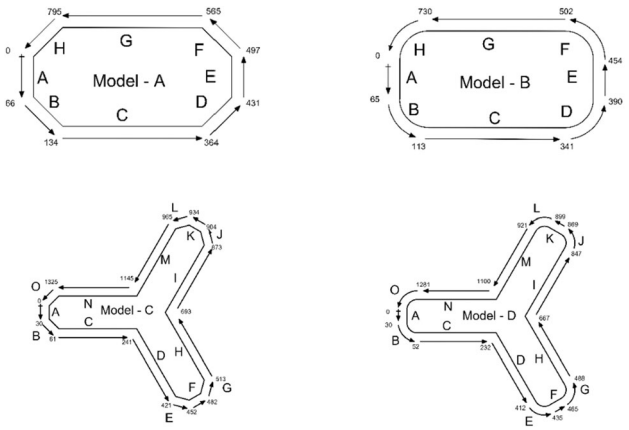



**Figure 7.** Peripheral distance along the building model-A, model-B, model-C and model-D.

## 2.5 *Turbulent intensity and mean wind speed profile*

Wind incident on tall buildings is followed by power-law index of wind distribution available for all types of terrain. Generally, wind speed is increases with respect to height and wind speed on the ground is almost zero due the obstruction faces at ground level. Wind velocity become gradient height becomes constant and this is known as free stream wind or gradient wind. Mean wind speed and turbulent intensity are presented in figure 4, the boundary conditions are kept similar with those experimental performed by Raj [54] in boundary layer wind tunnel and comparative graph is also presented to validate the numerical study with the experimental study.

Mean wind speed and turbulent intensity profile should be defined at the inlet to performed the numerical simulation. The boundary layer mean wind speed profile is governed by the power-law equation (eq.7) and the same is applied in the numerical simulation.

$$U = U_h\left(\frac{Z}{Z_H}\right)^{\alpha} \tag{7}$$

where $U_h$ is the boundary layer velocity, which is 10 m/s for this study; $Z$ is the reference height; $U$ is the mean wind speed at a reference height $Z$; $\alpha$ is a parameter that varies with ground roughness. (Known as power law index); $Z_H$ is the boundary layer depth;

$$I_z = \frac{\sigma_z}{U(z)} \tag{8}$$

where $I_z$ is the turbulence intensity at height z; $\sigma_z$ is the standard deviation of the wind speed at height z; $U(z)$ is the mean wind speed at height z.

## 2.6 *Validation*

The solution obtained with the help of numerical techniques are validated with three standards models represented in figure 5. Numerical findings are depicted in the graphical form in figure 6 and table 1. Results presented in paper are identical with experimental and different international standards.

The results of mean $C_p$ are calculated and compared with experimental study performed by Amin and Ahuja [55], Raj [54], Raj *et al* [23] as well as with the numerical finding of Sanyal and Dalui [56], Kumar and Dalui [57] and the result on model-1 is compared with different international standards [1–3, 5, 6, 58].

The external pressure coefficient " $C_p$" is calculated using the equation (9).

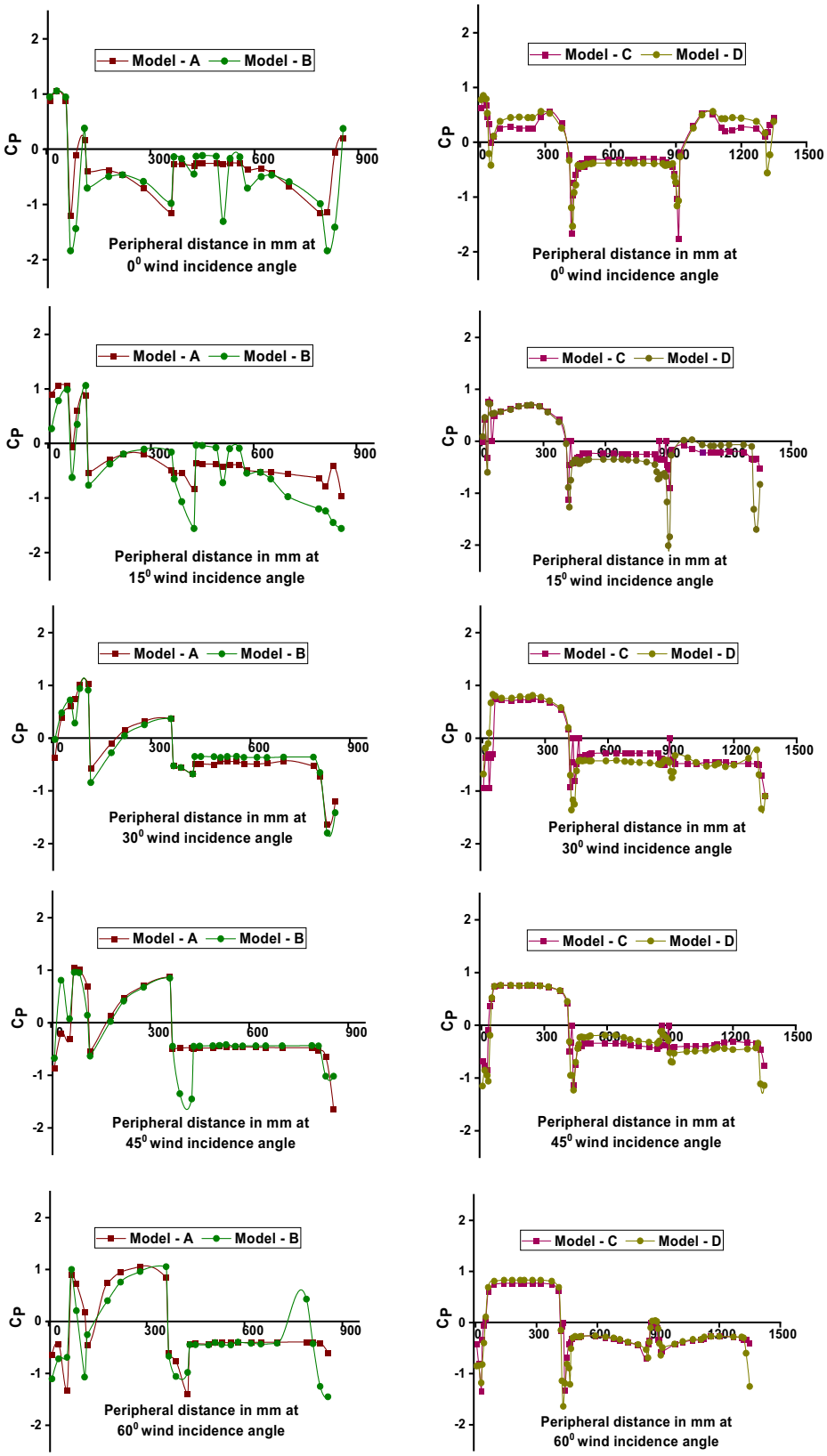



**Figure 8.** Pressure distribution along with the peripheral distance for model-A, model-B, model-C and model-D.

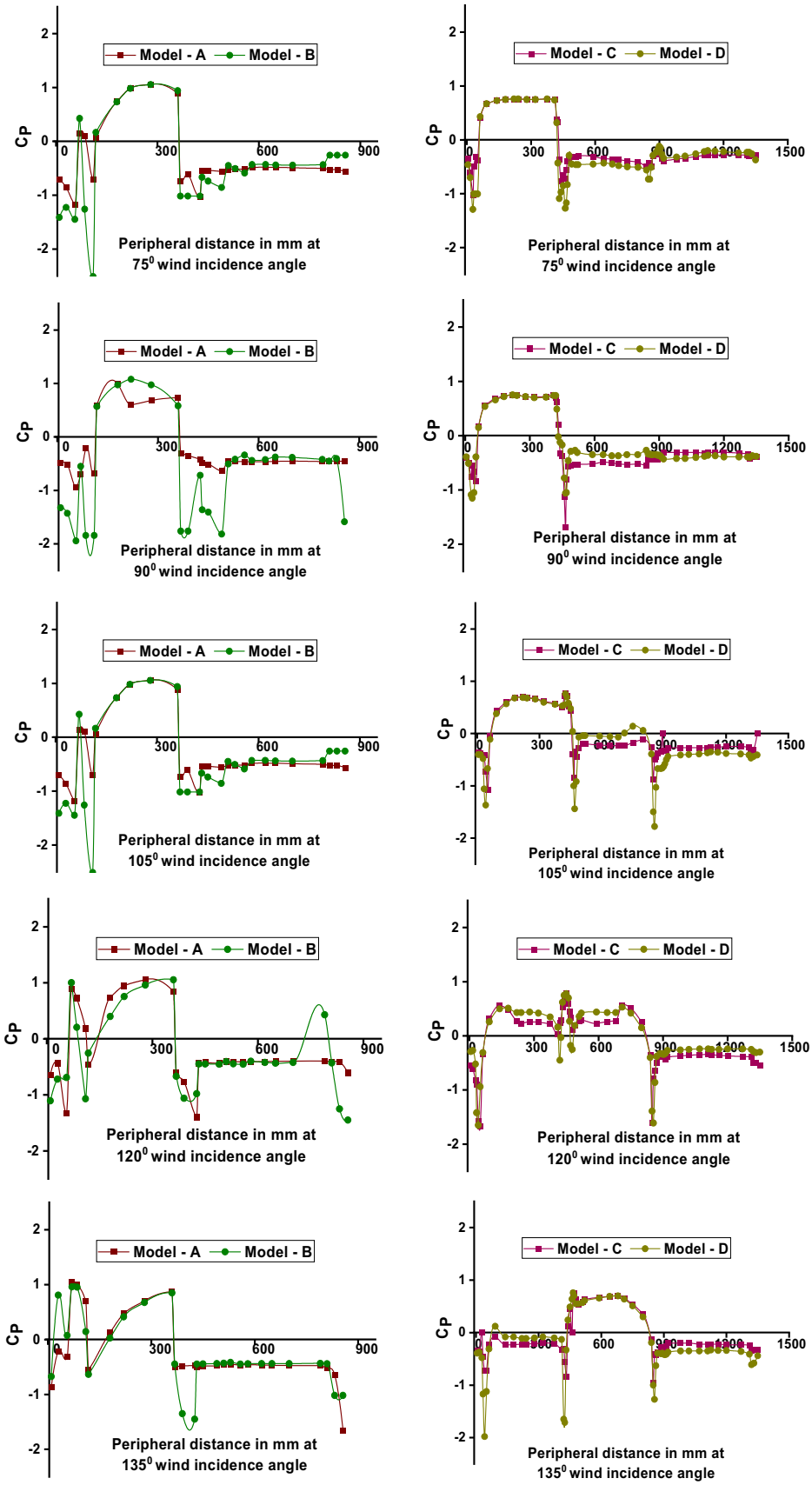



**Figure 8.** continued

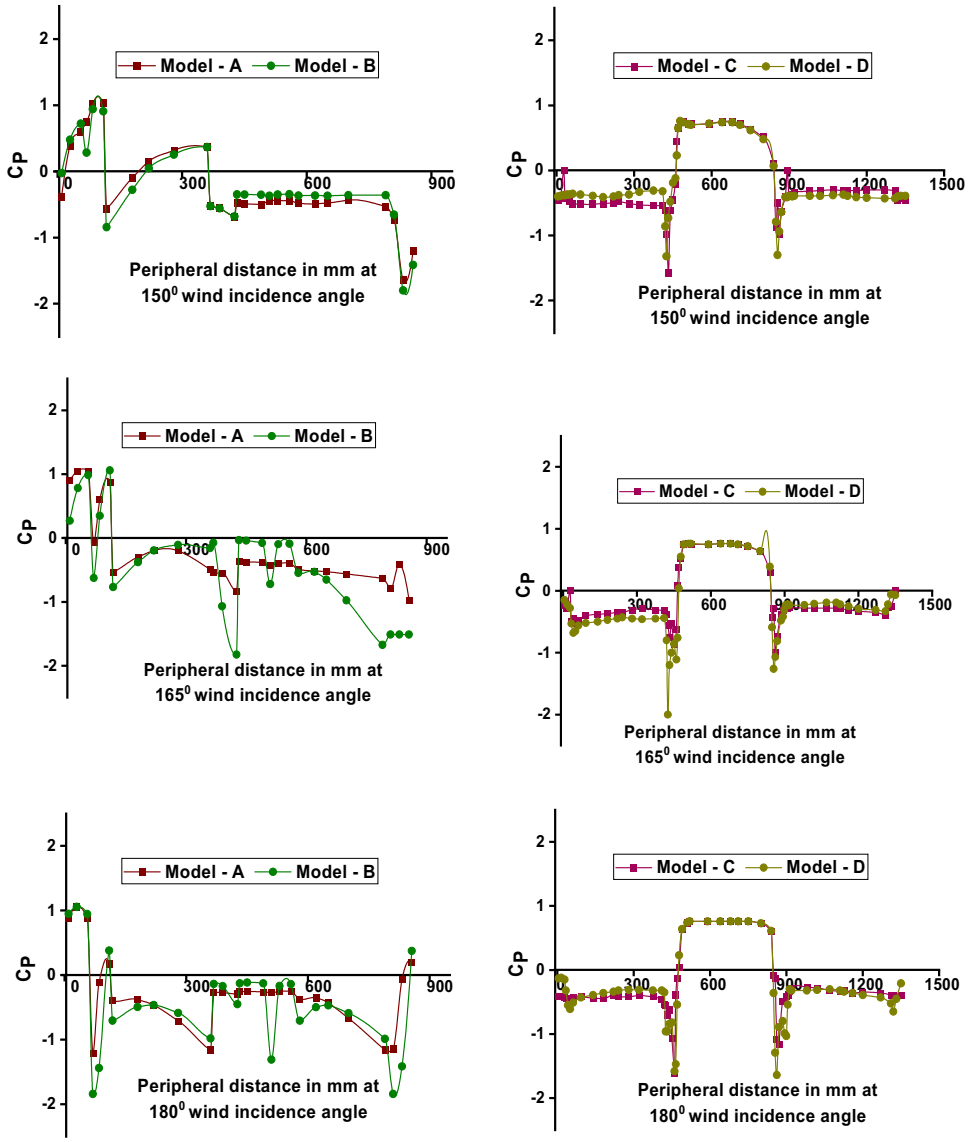



**Figure 8.** continued

$$C_p = \frac{p - p_o}{\frac{1}{2}\rho U_H{}^2} \tag{9}$$

## 2.7 *Grid convergence*

In this study a grid convergence study was performed on model-1. Grid convergence study is necessary for a CFD programming as it helps to know the accuracy of CFD solver settings. For the present study GC-3 is adopted for all the wind incidence angle varies from 0° to 180° at the interval of 15° each. In this study the grid convergence study is performed on the basis of procedure provided by Celik *et al* [59], Derakhshandeh and Alam [60]. The percentage error is reported in the table 2 and GC-3 is adopted because of the less percentage error reported in the mean Cp compared with the IS: 875 (part-3): 2015 [5] for 0° wind incidence angle. The grid convergence study is performed on five different cases by varying the meshing type namely coarse, medium and fine. Number of elements for coarse, medium and fine meshing are 957324, 1439589 and 2497236 respectively. Grid convergence study is performed on model-1 while for validation purposes the three model are taken into consideration and validation study is performed on model-1, 2 and 3.

## 3. Result and discussion

Results of pressure distribution and velocity streamlines are discussed and presented in a graphical and external Cp is tabulated for the building model- A, model- B, model- C and model- D.

### 3.1 *Horizontal pressure distribution at mid height of building models*

The pressure distribution along the peripheral distance for various building models are presented in figure 7. The pressure distribution along peripheral length are presented at mid height of building in form of graph at 0° to 180° at an interval of 15°. The pressure distribution for building model-A and model-B is of same nature for 90° winds. Figure 8 presented the pressure distribution for model-A, model-B, model-C and model-D. Face-A under the direct exposure to wind depicted positive pressure distribution throughout the face and it is seen that the maximum positive pressure for model-A is 1.05 on face-A at 0° and 15° wind incidences while for face-B at 45° and face-C at 60° & 75° wind incidence angles. The minimum pressure of −1.64 is noted at 832 mm on face-H at 30° wind, which is 156 % less from the maximum positive pressure.

The maximum positive pressure for model-B, is 1.07 at 220 mm which is generated at face-C in case of 90° wind incidence angle and the minimum negative pressure is −1.94 noted on face-A in the case of 90° wind incidence angle at 56 mm of peripheral distance, maximum negative pressure observed on model-B that is almost 180 % of maximum positive pressure. It can be concluded from the results that model-A is more efficient in terms of pressure distribution with respect to model-B having fillet corners. Y- shape building model with chamfer corners i.e. model-C having the maximum pressure of 0.79 at 15 mm and 450 mm peripheral distance on face-A in case of 90° & 120° wind incidence angle respectively.

The minimum pressure of −1.76 which is more than maximum positive pressure on model-C, and similar observation found at 912 mm peripheral distance in case of model-D at 0° wind angle. Y- shape with round corner in each limb having the maximum and minimum pressure of 0.85 at 15 mm and −2.01 detect at 905 mm on face-A at 0° wind angle respectively. Among the model-C and model-D the maximum efficient model is model-C which is having the chamfer shape in each limb of irregular Y- shape.

### 3.2 *Velocity streamlines*

The Velocity streamlines are presented pictorially for model-A (Rectangular Chamfer) in figure 9 from 0° to 90° wind incidence angle at 15° intervals. Streamlines are obtained from ANSYS CFX Post processing. The wind flow pattern is clearly presenting the size of wake in downstream of the wind, maximum wake is observed for 90° winds as the model is obstructing more wind flow with respect to 0° wind angle and thus the wake is increasing. The pattern of vortex formed in the downstream is changing at each wind angle also, the reattachment of the flow is more in the case of 0° wind.

### 3.3 *Numerical simulation ISO surface of pressure*

The ISO-surface of pressure is a three-dimensional visualization which present the physical shape of pressure distribution around the building model and is obtained through ANSYS CFX Post by applying Q-Criterion [61, 62]. The ISO-surface of pressure is generated and presented in figure 10 for building model- A, model- B, model- C and model- D.

### 3.4 *External pressure coefficient*

The external pressure coefficient on the high-rise structure helps designer to calculate the wind effects on such tall buildings. External pressure coefficient for building model-A i.e., rectangular chamfer shown in figure 11 and numerical output are presented in table 3. After the observation face-B and face-H have slightly similar pressure coefficients. The highest pressure observed at 0° wind angle and it is found (0.91) on the face-A, smallest pressure (−1.11) is spot on face-A in case of 75° wind incidence angle. The external pressure coefficient is more or less equivalent with each other in the case of wind ward faces (the face which is in the directly exposure to wind). It is discovered that identical face having uniform pressure coefficient for particular wind incidence angle.

The External pressure coefficient evaluated for building model-B presented in figure 11 (Rectangular fillet) is tabulated in table 4, at 0° to 90° wind incidence angles at the interval of 15°. Maximum positive pressure (0.71) is detected on face-A at 0° wind incidence angle while the least pressure (-1.49) is noticed on face- B and face- D in the case of 90° wind incidence angle.

The external pressure coefficient calculated and tabulated in table 5 for building model-C (Y-shape Chamfer) presented in figure 12. Research study is done on Y- shape for 0° to 180° wind incidence angle while for rectangular model 0° to 90° wind incidence angle. The external pressure coefficient is presented in a tabular form in table 3, the highest pressure (0.69) is observed at face-C in the case of 45° wind incidence and minimum pressure (−1.14) is noticed on face-E, at 150° wind. This is concluded that symmetric face is having identical nature of pressure distribution are observed for 0°, 60°, 120° and 180° wind. for building model-D (Y-shape Fillet) represented in figure 12, the external pressure coefficients are tabulated in table 6,

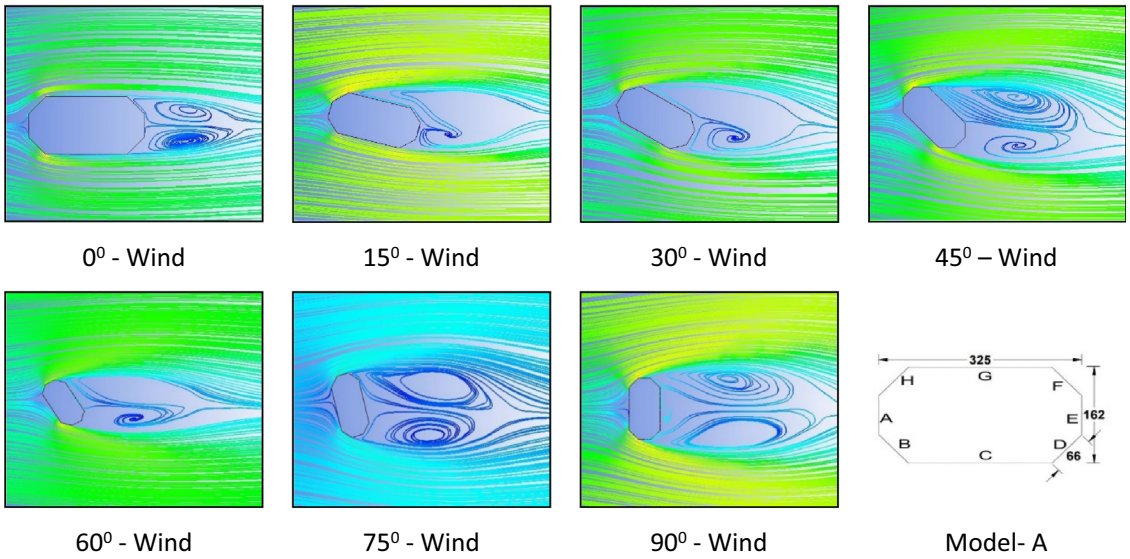



**Figure 9.** Velocity Streamlines for model-A at 0° to 90° wind incidence angles at an interval of 15°

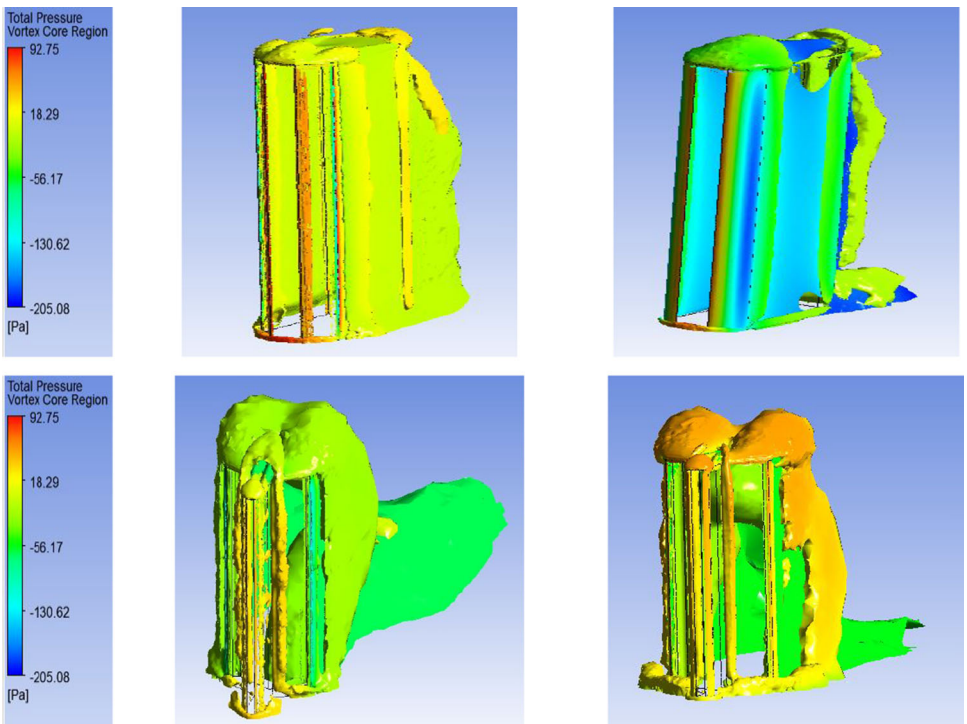



**Figure 10.** ISO Surface of Pressure around building model-A, model-B, model-C and model-D.

**Figure 11.** Building models (A) Rectangular Chamfer and (B) Rectangular Fillet.

the maximum pressure (0.69) is noticed on the face-A, face-C, face-D, face-F and face-H at 0°, 45°, 75°, 120° and 165° respectively and minimum pressure (−1.4) on face-L in the case of 15° wind incidence angle. The maximum positive pressure for building models- B and C is nearly of same magnitude on the wind ward face while in the case of negative pressure it is 22 % more for model- D with respect to model- C.

**Table 3.** External pressure coefficient for building model-A (Rectangular chamfer).

| Model-A (Rectangular chamfer) | | | | | | | |
|---|---|---|---|---|---|---|---|
| Face | 0° | 15° | 30° | 45° | 60° | 75° | 90° |
| **A** | **0.91** | **0.83** | 0.21 | −0.36 | −0.78 | **−1.11** | **−0.50** |
| **B** | −0.30 | 0.46 | **0.89** | **0.87** | **0.57** | −0.12 | −0.45 |
| **C** | **−0.59** | −0.32 | 0.02 | 0.29 | 0.19 | **0.71** | **0.82** |
| **D** | −0.32 | −0.46 | −0.52 | **−0.69** | **−1.12** | −0.76 | −0.45 |
| **E** | −0.28 | −0.37 | −0.45 | −0.46 | −0.46 | −0.51 | **−0.50** |
| **F** | −0.32 | −0.42 | −0.44 | −0.45 | −0.4 | −0.48 | −0.41 |
| **G** | **−0.59** | −0.54 | −0.47 | −0.43 | −0.42 | −0.44 | −0.42 |
| **H** | −0.30 | **−0.65** | **−0.99** | −0.83 | −0.50 | −0.45 | −0.41 |

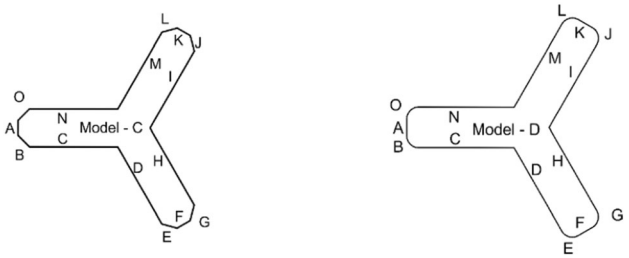



**Figure 12.** Building models (C) Y-shape Chamfer and (D) Y-shape Fillet.

## 3.5 *Base shear*

Base shear is derived from the results obtained through the computational fluid dynamic tool ANSYS CFX. The force is calculated on the base and presented in graphical form in figure 13 for the building models considered in this study. The largest base shear along x-direction that is a drag for model- A is 1.84 for at 45° and 135° while for model- B the maximum drag is 1.68 at 30° and 135° in the case of the model- C the drag of 0.65 is maximum at 60° and 180°. The

**Table 4.** External pressure coefficient for building model-B (Rectangular fillet).

| Model-B (Rectangular fillet) | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Face** | **0°** | **15°** | **30°** | **45°** | **60°** | **75°** | **90°** |
| **A** | **0.71** | **0.51** | 0.28 | 0.05 | −0.59 | −0.92 | −0.99 |
| **B** | −0.76 | 0.18 | **0.50** | **0.49** | **0.04** | −0.74 | **−1.49** |
| **C** | −0.45 | −0.24 | −0.08 | 0.19 | **−1.48** | **0.56** | **0.61** |
| **D** | −0.40 | −0.78 | −0.39 | **−1.24** | −0.32 | **−1.48** | **−1.49** |
| **E** | −0.11 | −0.09 | −0.29 | −0.31 | −0.35 | −0.41 | −0.99 |
| **F** | −0.40 | −0.23 | −0.32 | −0.32 | −0.35 | −0.41 | −0.60 |
| **G** | −0.45 | −0.59 | −0.32 | −0.32 | −0.32 | −0.34 | −0.32 |
| **H** | **−0.76** | **−1.94** | **−1.22** | **−1.24** | −1.06 | −0.83 | −0.60 |

**Table 5.** External pressure coefficient for building model-C (Y-shape Chamfer).

| Model-C (Y-shape Chamfer) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Face | 0° | 15° | 30° | 45° | 60° | 75° | 90° | 105° | 120° | 135° | 150° | 165° | 180° |
| **A** | **0.65** | −0.06 | −0.60 | **−0.66** | **−0.87** | −0.62 | −0.60 | −0.45 | −0.45 | −0.41 | −0.45 | −0.32 | −0.34 |
| **B** | 0.19 | 0.47 | −0.60 | 0.26 | −0.14 | −0.38 | −0.69 | **−0.77** | **−1.08** | **−0.71** | −0.60 | −0.46 | −0.39 |
| **C** | 0.20 | **0.54** | **0.66** | **0.69** | **0.67** | 0.62 | 0.53 | 0.43 | 0.20 | −0.12 | −0.46 | −0.39 | −0.35 |
| **D** | 0.23 | 0.43 | 0.54 | 0.62 | **0.67** | **0.69** | **0.66** | **0.53** | 0.21 | −0.25 | −0.59 | −0.35 | −0.30 |
| **E** | **−1.06** | −0.59 | −0.43 | −0.26 | −0.14 | 0.14 | 0.41 | 0.51 | 0.25 | −0.29 | **−1.14** | −0.58 | −0.15 |
| **F** | −0.49 | −0.46 | −0.53 | −0.65 | **−0.87** | **−0.63** | −0.65 | 0.17 | **0.66** | 0.34 | −0.42 | **−0.65** | **−0.75** |
| **G** | −0.40 | −0.36 | −0.36 | −0.29 | −0.32 | −0.44 | **−0.99** | −0.53 | 0.25 | **0.63** | 0.53 | 0.30 | −0.21 |
| **H** | −0.34 | −0.31 | −0.31 | −0.28 | −0.34 | −0.35 | −0.59 | −0.24 | 0.21 | 0.54 | **0.66** | **0.69** | **0.66** |
| **I** | −0.34 | −0.32 | −0.31 | −0.34 | −0.45 | −0.39 | −0.48 | −0.11 | 0.20 | 0.38 | 0.50 | 0.60 | **0.66** |
| **J** | −0.40 | −0.27 | −0.33 | −0.27 | −0.39 | −0.40 | −0.60 | −0.61 | **−1.08** | −0.60 | −0.74 | −0.58 | −0.21 |
| **K** | −0.49 | −0.41 | −0.45 | −0.32 | −0.29 | −0.31 | −0.47 | −0.41 | −0.45 | −0.41 | −0.47 | −0.45 | **−0.75** |
| **L** | **−1.06** | **−0.73** | −0.58 | −0.43 | −0.39 | −0.29 | −0.45 | −0.38 | −0.44 | −0.36 | −0.39 | −0.26 | −0.15 |
| **M** | 0.23 | −0.09 | −0.42 | −0.39 | −0.45 | −0.32 | −0.35 | −0.32 | −0.39 | −0.30 | −0.35 | −0.26 | −0.30 |
| **N** | 0.20 | −0.25 | −0.56 | −0.35 | −0.34 | −0.26 | −0.35 | −0.32 | −0.39 | −0.31 | −0.34 | −0.32 | −0.35 |
| **O** | 0.19 | −0.40 | **−0.84** | −0.52 | −0.32 | −0.26 | −0.31 | −0.25 | −0.44 | −0.28 | −0.35 | −0.25 | −0.39 |

**Table 6.** External pressure coefficient for building model-D (Y-shape fillet).

| Model-D (Y-shape fillet) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Face | 0° | 15° | 30° | 45° | 60° | 75° | 90° | 105° | 120° | 135° | 150° | 165° | 180° |
| **A** | **0.69** | 0.01 | −0.37 | **−0.94** | **−0.96** | −0.86 | −0.58 | −0.41 | −0.39 | −0.44 | −0.39 | −0.36 | −0.30 |
| **B** | −0.06 | **0.58** | 0.39 | −0.25 | −0.66 | **−1.24** | −0.76 | −0.95 | **−1.28** | **−1.18** | −0.67 | −0.51 | −0.46 |
| **C** | 0.31 | 0.54 | **0.65** | **0.69** | **0.67** | 0.62 | 0.53 | 0.41 | 0.24 | −0.06 | −0.26 | −0.49 | −0.42 |
| **D** | 0.26 | 0.41 | 0.53 | 0.62 | **0.67** | **0.69** | **0.65** | 0.53 | 0.32 | −0.11 | −0.43 | −0.48 | −0.36 |
| **E** | **−0.89** | −0.90 | −0.87 | −0.74 | −0.66 | −0.39 | 0.33 | **0.60** | 0.02 | −1.07 | **−0.98** | −1.20 | −0.77 |
| **F** | −0.39 | −0.39 | −0.57 | −0.84 | **−0.96** | −0.99 | −0.42 | 0.34 | **0.69** | 0.42 | −0.31 | −0.94 | **−0.96** |
| **G** | −0.36 | −0.33 | −0.32 | −0.38 | −0.79 | −0.79 | **−0.94** | **−1.01** | 0.02 | **0.59** | 0.47 | −0.07 | −0.78 |
| **H** | −0.29 | −0.3 | −0.32 | −0.3 | −0.36 | −0.47 | −0.4 | −0.06 | 0.32 | 0.54 | **0.65** | **0.69** | **0.68** |
| **I** | −0.29 | −0.32 | −0.35 | −0.37 | −0.43 | −0.49 | −0.25 | −0.01 | 0.24 | 0.37 | 0.50 | 0.60 | **0.68** |
| **J** | −0.36 | −0.48 | −0.33 | −0.34 | −0.44 | −0.52 | −0.63 | **−1.01** | **−1.28** | −0.92 | −0.9 | **−0.94** | −0.78 |
| **K** | −0.39 | −0.49 | −0.35 | −0.36 | −0.25 | −0.36 | −0.34 | −0.38 | −0.39 | −0.36 | −0.4 | −0.66 | **−0.96** |
| **L** | **−0.89** | −1.40 | −0.87 | −0.51 | −0.44 | −0.34 | −0.33 | −0.35 | −0.35 | −0.35 | −0.31 | −0.37 | −0.77 |
| **M** | 0.26 | −0.03 | −0.23 | −0.48 | −0.43 | −0.37 | −0.35 | −0.31 | −0.30 | −0.31 | −0.31 | −0.3 | −0.36 |
| **N** | 0.31 | −0.10 | −0.40 | −0.47 | −0.36 | −0.30 | −0.32 | −0.31 | −0.30 | −0.33 | −0.34 | −0.37 | −0.42 |
| **O** | −0.06 | **−1.23** | **−1.02** | −0.87 | −0.79 | −0.40 | −0.32 | −0.35 | −0.35 | −0.55 | −0.34 | −0.32 | −0.46 |

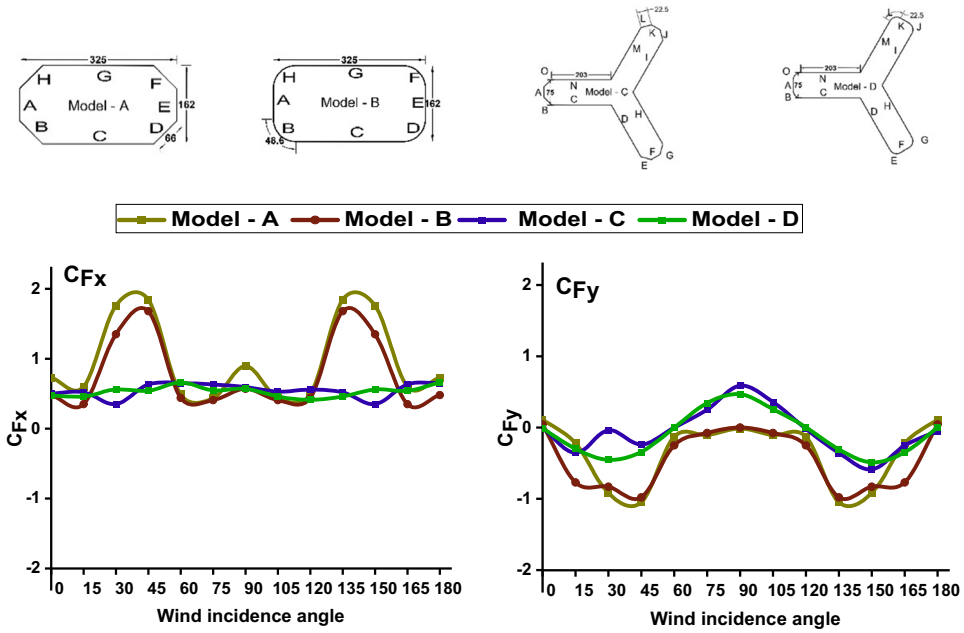



**Figure 13.** Drag and Lift Force coefficient for model-A, model-B, model-C and model-D at 0° to 180° wind incidence angle.

drag of 0.66 is found maximum for model-D at 60° and 180° wind incidence angles. Lift force is calculated and presented in figure 13 the minimum lift force is found −1.05 for model-A at 45° and 135° while on model-B the least lift force is −0.98 at 45° and 135° wind incidence angles.

Y-shape model-C having chamfer edges at the end of each limb is having the highest and lowest lift as 0.58 at 90° and −0.25 at 165° respectively, while the model- D having round in corner in each limb of Y-shape having the maximum lift of 0.46 and it is spotted at 90° and minimum of −0.45 in noted at 30° winds. Base shear in x and y direction are $C_{f_x}$ & $C_{f_y}$ and it is calculated using equation 10 and equation 11 where $F_x$ & $F_y$ is the base force obtained through numerical simulation, $U_h$ is the reference wind speed at height h and $A_p$ is the projected area.

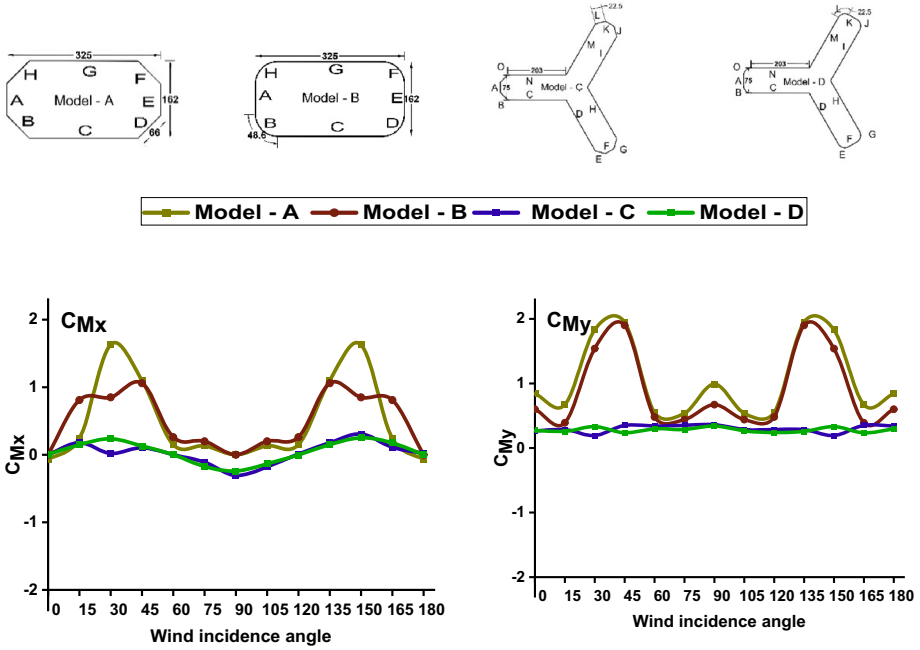



**Figure 14.** Base moment along X ($C_{Mx}$) and Y ($C_{My}$) for model-A, model-B, model-C and model-D at 0° to 180° at an interval of 15° wind incidence angle.

$$C_{f_x} = \frac{F_x}{(0.5\rho_{U_h{}^2}.A_p)} \tag{10}$$

$$C_{f_y} = \frac{F_y}{(0.5\rho_{U_h{}^2}.A_p)} \tag{11}$$

$$C_{M_x} = \frac{M_x}{(0.5\rho_{U_h{}^2}.A_p.H)} \tag{12}$$

$$C_{M_y} = \frac{M_y}{(0.5\rho_{U_h{}^2}.A_p.H)} \tag{13}$$

### 3.6 *Base moment*

The base moment for model-A, B, C & D are presented in figure 14, moments are calculated at the base of the building model. The maximum $M_x$ is 1.63 for model-A at 30° and 150° wind while for model-B is 1.06 at 45° and 135° wind incidence angle. The greatest $C_{Mx}$ for model- C is 0.30 at 90° and 150° wind while for model-D the maximum is 0.25 at 150° wind incidence angle. The base moment in the y-direction is found extreme of 1.95 for model-A at 45° and 135° wind angle while for model-C the maximum $C_{My}$ is 0.36 at 90° and for model having round corner in each limb of Y-shape is having the maximum $C_{My}$ of 0.34 at the 90°- wind incidence angle. Base moment along x and y direction are $C_{M_x}$ & $C_{M_y}$ where $M_x$ and $M_y$ is the base moment obtained through the numerical simulation, $U_h$ is the reference wind speed at height h, $A_p$ is the projected area and H is the height of the building model.

## 4. Conclusion

This research study is performed using ANSYS CFX on various shape building models and presents the comparison between regular and irregular shape building models having different corner configurations. The results obtained through numerical techniques is validated with the experimental results and different international standards. The significant outcomes from the present study are noted as follows.

- The validation study shows the very prominent results for model-1, model-2 and model-3, also the results obtained for $C_{Pmean}$ are found significantly in the closer range with experimental as well as with different international standards.
- Pressure distribution along the peripheral distance for building model having equal cross-sectional area are discussed and presented in the graphical form for

model-A, model-B, model-C and model-D at the interval of 15° for the wind incidence angle varies between 0° to 180°.

- While comparing the models- C & D, maximum pressure along the peripheral distance found in model-D and in the case of model-A & B, the maximum positive pressure is observed for model-B, out of all four models, the model- D is having lesser pressure than the model- B.
- Out of all models, the model having round in corner Y-shape has the minimum overall base moment which is 24 % less than the model-A rectangular chamfer. Model-A & B which is having equal number of faces, the lowest external pressure coefficient is observed for model-B in the case of 15° wind incidence angle. On comparing model-C & D, model-D is more aerodynamic with respect to model-C.
- The present study is comparing the result of equal cross sectional area buildings. The dimensions of the modifications are kept same for all model and it is found that the minimum drag force is obtained for model-D which is having round corner in each limb of Y-shape.

## Acknowledgement

Authors would like to express their sincere gratitude to Delhi Technological University, Delhi, India for providing funding to conduct the research work. Thanks are due to the DTU, Delhi for providing resources and research fellowship for the first author under Institutional Fellowship.

**Data availability statement** All data, models and code generated or used during the study appeared in the submitted article. Data are available in the form of tables and graphs.

## References

[1] ASCE: 7-16 2017 *Minimum Design Loads and Associated Criteria for Buildings and Other Structures.* Structural Engineering Institute of the American Society of Civil Engineering Reston

[2] AS/NZS:1170.2 2011: Structural Design Actions - Part 2: Wind actions Standards Australia/Standards New Zealand, Sydney

[3] GB 50009-2001: National Standard of The People's Republic of China

[4] NSCP C101-15: 2015 National Structural Code of the Philippines

[5] IS: 875- Part-3 2015 Code of Practice for Design Loads (other than earthquake loads) for buildings and structures-wind loads. India

[6] EN 1991-1-4 (2005)/AC: 2010 (E) European Standard Eurocode 1: Actions on structures - Part 1-4: General actions Wind actions European Committee for Standardization (CEN) Europe

[7] MNBC MYANMAR NATIONAL BUILDING CODE Part 3 and Part 4 2020 International relation and legal Section, Department of Building, Ministry of Construction

[8] ASCE: 49-12 2012 Wind Tunnel Testing for Buildings and Other Structures.Structural Engineering Institute of the American Society of Civil Engineering, Reston. *American Society of Civil Engineers* 1801 Alexander Bell Drive Reston Virginia

[9] ASCE Manuals and Reports on Engineering practice No. 67: 1999 *Wind Tunnel Studies of Buildings and Structures*

[10] Versteeg H K and Malalasekera W 1995 An introduction to Computational Fluid Dynamics the Finite Volume Method. Springer, England

[11] Ferziger J H and Peric M 2002 *Computational Methods for Fluid Dynamics,* Germany

[12] John D Anderson J 2010 *Computational Fluid Dynamic*, second edition. Singapore

[13] Nagar S K, Raj R and Dev N 2020 Experimental study of wind induced pressures on tall buildings of different shapes. *Wind Struct.* 31: 441–453. https://doi.org/10.12989/was.2020.31.5.441

[14] Kwok K C S 1988 Effect of building shape on wind-induced response of tall building. *J. Wind Eng. Ind. Aerodyn.* 28: 381–390. https://doi.org/10.1016/0167-6105(88)90134-1

[15] Blessmann J and Riera J D 1985 Wind excitation of neighbouring tall buildings. *J. Wind Eng. Ind. Aerodyn.* 18: 91–103. https://doi.org/10.1016/0167-6105(85)90076-5

[16] Pal S, Raj R and Bilateral Anbukumar S 2021 Interference of wind loads induced on duplicate building models of various shapes. *Latin Am. J. Solids Struct.*. https://doi.org/10.1590/1679-78256595

[17] Sharma A, Mittal H and Gairola A 2018 Mitigation of wind load on tall buildings through aerodynamic modifications. *J. Build. Eng.* 18: 180–194. https://doi.org/10.1016/j.jobe.2018.03.005

[18] Yi J and Li Q S 2018 Wind tunnel and full-scale study of wind effects on a super-tall building. *J. Fluids Struct.* 58: 236–253. https://doi.org/10.1016/j.jfluidstructs.2015.08.005

[19] Bairagi A K and Dalui S K 2018 Comparison of aerodynamic coefficients of setback tall buildings due to wind load. *Asian J. Civ. Eng.* 19: 205–221. https://doi.org/10.1007/s42107-018-0018-3

[20] Bhattacharyya B and Dalui S K 2018 Investigation of mean wind pressures on 'E' plan shaped tall building. *Wind Struct.* 26: 99–114. https://doi.org/10.12989/was.2018.26.2.099

[21] Bearman P W and Morel T 1983 Effect of free stream turbulence on the flow around bluff bodies. *Prog. Aerosp. Sci.* 20: 97–123. https://doi.org/10.1016/0376-0421(83)90002-7

[22] Raj R and Ahuja A K 2013 Wind loads on cross shape tall buildings. *J. Acad. Ind. Res.* 2: 111–113

[23] Raj R, Sharma A and Chauhan S 2018 Response of square and plus shaped buildings on varying wind loads. *Urban. Chall. Emerg. Econ.*. https://doi.org/10.1061/9780784482032.022

[24] Li Y, Li Q S and Chen F 2017 Wind tunnel study of wind-induced torques on L-shaped tall buildings. *J. Wind Eng. Ind. Aerodyn.* 167: 41–50. https://doi.org/10.1016/j.jweia.2017.04.013

[25] Blackmore P A 1997 The role of wind tunnel testing in the design of building structures. *Proc Inst. Civ. Eng. Struct. Build.* 122: 253–265. https://doi.org/10.1680/istbu.1997.29797

[26] Hui Y, Yoshida A and Tamura Y 2013 Interference effects between two rectangular-section high-rise buildings on local peak pressure coefficients. *J. Fluids Struct.* 37: 120–133. https://doi.org/10.1016/j.jfluidstructs.2012.11.007

[27] Pal S and Raj R 2021 Evaluation of wind induced interference effects on shape remodeled tall buildings. *Arab. J. Sci. Eng.* 46: 11425–11445. https://doi.org/10.1007/s13369-021-05923-x

[28] Zu G B and Lam K M 2018 Across wind excitation mechanism for interference of twin tall buildings in staggered arrangement. *J. Wind Eng. Ind. Aerodyn.* 177: 167–185. https://doi.org/10.1016/j.jweia.2018.04.019

[29] Nagar S K, Raj R and Dev N 2021 Proximity effects between two plus plan shaped high-rise buildings on mean and RMS pressure coefficients. *Sci. Iran.* https://doi.org/10.24200/sci.2021.55928.4484

[30] Gaur N, Raj R and Goyal P K 2021 Interference effect on corner configured structures with variable geometry and blockage configurations under wind loads using CFD. *Asian J. Civ. Eng.* 22: 1607–1623. https://doi.org/10.1007/s42107-021-00400-0

[31] Rajasekarababu K B, Vinayagamurthy G and Selvi Rajan S 2019 Experimental and computational investigation of outdoor wind flow around a setback building. *Build. Simul.* 12: 891–904. https://doi.org/10.1007/s12273-019-0514-8

[32] Taniguchi K and Akamine Y 2020 Wind pressure coefficient distribution of detached houses in a dense residential block. *Japan Archit. Rev.* 3: 629–645. https://doi.org/10.1002/2475-8876.12173

[33] Zheng D, Zhang A and Gu M 2012 Improvement of inflow boundary condition in large eddy simulation of flow around tall building. *Eng. Appl. Comput. Fluid Mech.* 6: 633–647. https://doi.org/10.1080/19942060.2012.11015448

[34] Haque M N, Katsuchi H, Yamada H and Nishio M 2014 Investigation of flow fields around rectangular cylinder under turbulent flow by LES. *Eng. Appl. Comput. Fluid Mech.* 8: 396–406. https://doi.org/10.1080/19942060.2014.11015524

[35] Gaur N and Raj R 2021 Aerodynamic mitigation by corner modification on square model under wind loads employing CFD and wind tunnel. *Ain Shams Eng. J.* 13: 101521. https://doi.org/10.1016/j.asej.2021.06.007

[36] Mou B, He B J, Zhao D X and Chau K W 2017 Numerical simulation of the effects of building dimensional variation on wind pressure distribution. *Eng. Appl. Comput. Fluid Mech.* 11: 293–309. https://doi.org/10.1080/19942060.2017.1281845

[37] Hu P, Li Y, Han Y, Cai S C S and Xu X 2016 Numerical simulations of the mean wind speeds and turbulence intensities over simplified gorges using the SST K-ω turbulence model. *Eng. Appl. Comput. Fluid Mech.* 10: 359–372. https://doi.org/10.1080/19942060.2016.1169947

[38] Blocken B, Stathopoulos T and Carmeliet J 2007 CFD simulation of the atmospheric boundary layer: wall function problems. *Atmos. Environ.* 41: 238–252. https://doi.org/10.1016/j.atmosenv.2006.08.019

[39] Bairagi A K and Dalui S K 2021 Estimation of Wind Load on Stepped Tall Building Using CFD Simulation. *Iran. J. Sci. Technol. Trans. Civ. Eng.* 45: 707–727. https://doi.org/10.1007/s40996-020-00535-1

[40] Kumar A and Raj R 2021 Study of pressure distribution on an irregular octagonal plan oval shape building using CFD. *Civ. Eng. J.* 7: 1787–1805. https://doi.org/10.28991/cej-2021-03091760

[41] Okajima A 1982 Strouhal number of rectangular cylinders. *J. Fluid Mech.* 123: 379–398. https://doi.org/10.1017/S0022112082003115

[42] Mallick M, Mohanta A, Kumar A and Raj V 2018 Modelling of wind pressure coefficients on C shaped building models. *Model. Simul. Eng.* 2018: 6524945. https://doi.org/10.1155/2018/6524945

[43] Stathopoulos T and Baskaran B A 1996 Computer simulation of wind environmental conditions around buildings. *Eng. Struct.* 18: 876–885. https://doi.org/10.1016/0141-0296(95)00155-7

[44] Cheng X, Huang G, Yang Q and Zhou X 2021 Influence of architectural facades on wind pressures and aerodynamic forces of tall buildings. *J. Struct. Eng.*. https://doi.org/10.1061/(asce)st.1943-541x.0002867

[45] Bhattacharjee S, Banerjee S, Majumdar S G, Dey A and Sanyal P 2021 Effects of irregularity on a butterfly plan shaped tall building under wind load. *J. Inst. Eng. (India) Ser. A* 102: 451–467. https://doi.org/10.1007/s40030-021-00511-6

[46] Tang J W, Xie Y M, Felicetti P, Tu J Y and Li J D 2010 Numerical simulations of wind drags on straight and twisted polygonal buildings. *Struct. Des. Tall Special Build.* 22: 62–73. https://doi.org/10.1002/tal.657

[47] Zheng C R and Zhang Y C 2010 Numerical investigation on the drag reduction properties of a suction controlled high-rise building. *J. Zhejiang Univ. Sci. A.* 11: 477–487. https://doi.org/10.1631/jzus.A0900593

[48] Bairagi A K and Dalui S K 2021 Wind environment around the setback building models. *Build. Simul.* 14: 1525–1541. https://doi.org/10.1007/s12273-020-0758-3

[49] Raj R, Jha S, Singh S and Choudhary S 2020 Response analysis of plus shaped tall building with different bracing systems under wind load. *Int. J. Adv. Res. Eng. Technol.* 11: 371–380. https://doi.org/10.34218/IJARET.11.3.2020.032

[50] ANSYS 2020 CFX Solver Modeling Guide Europe

[51] ANSYS 2010 Meshing User Guide Europe

[52] Abu Zidan Y, Mendis P and Gunawardena T 2021 Optimising the computational domain size in CFD simulations of tall buildings. *Heliyon* 7: e06723. https://doi.org/10.1016/j.heliyon.2021.e06723

[53] Revuz J, Hargreaves D M and Owen J S 2012 On the domain size for the steady-state CFD modelling of a tall building. *Wind Struct.* 15: 313–329. https://doi.org/10.12989/was.2012.15.4.313

[54] Raj R 2015 *Effects of cross-sectional shapes on response of tall buildings under wind loads*. Doctorate of Philosophy, Indian Institute of Technology Roorkee, Roorkee

[55] Amin J A and Ahuja A K 2013 Effects of side ratio on wind-induced pressure distribution on rectangular buildings. *J. Struct.*. https://doi.org/10.1155/2013/176739

[56] Sanyal P and Dalui S K 2018 Effects of courtyard and opening on a rectangular plan shaped tall building under wind load. *Int. J. Adv. Struct. Eng.* 10: 169–188. https://doi.org/10.1007/s40091-018-0190-4

[57] Kumar D and Dalui S K 2017 Effect of internal angles between limbs of cross plan shaped tall building under wind load. *Wind Struct.* 24: 95–118. https://doi.org/10.12989/was.2017.24.2.095

[58] ETHIOPIAN STANDARD: ES ISO 4354 2012 English Wind actions on structures

[59] Celik I B, Ghia U, Roache P J, Freitas C J, Coleman H and Raad P E 2008 Procedure for estimation and reporting of uncertainty due to discretization in CFD applications. *J. Fluids Eng. ASME* 130: 0780011–0780014. https://doi.org/10.1115/1.2960953

[60] Derakhshandeh J F and Alam M M 2018 Flow structures around rectangular cylinder in the vicinity of a wall. *Wind Struct.* 26: 293–304. https://doi.org/10.12989/was.2018.26.5.293

[61] Haller G 2005 An objective definition of a vortex. *J. Fluid Mech.* 525: 1–26. https://doi.org/10.1017/S0022112004002526

[62] Zheng X, Montazeri H and Blocken B 2021 CFD analysis of the impact of geometrical characteristics of building balconies on near-façade wind flow and surface pressure. *Build. Environ.* 200: 107904. https://doi.org/10.1016/j.buildenv.2021.107904

2022 IEEE 2nd International Conference on Computer Communication and Artificial Intelligence (CCAI) | 978-1-6654-9663-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/CCAI55564.2022.9807727

# Encryption and Decryption of Marker Based 3-Dimensional Augmented Reality Image Using Modified Hill Cipher Technique for Secure Transfer

Abhilasha Sharma
*Department of Software Engineering*
*Delhi Technological University*
Delhi, India
abhi16.sharma@gmail.com

Anshul Singh*
*Department of Software Engineering*
*Delhi Technological University*
Delhi, India
anshulsingh_2k18se029@dtu.ac.in

Appu Kumar
*Department of Software Engineering*
*Delhi Technological University*
Delhi, India
appukumar_2k18se035@dtu.ac.in

***Abstract*—The Mobility of digital data has increased over time, but securing them is a major task to do. Research has been conducted to help shield data from unauthorized intruders and attacks that includes disclosure of data, masquerade, content modification and repudiation, though safeguarding Augmented Reality media remains an unfulfilled issue. This research takes a step towards encrypting augmented reality content by implementing the modified Hill cipher encryption approach. This is acquired by offering marker-based mobility of immersive 3 Dimensional media, that is scanned by the user to extract it; furthermore, users first verify their identity by entering the passcode or key to the modified Hill cipher to decrypt the data. Coupling cryptography and augmented reality will lead to a safe and secured meaningful engagement of the digital media over the networks.**

***Keywords—augmented reality, cryptography, modified hill cipher, encryption, decryption, key, markers, security***

## I. INTRODUCTION

Augmented Reality (AR) has been visualized as a digital world having virtual representation. However, augmented reality is much more, having applications in practically every field. It has the potential to alter how humans interact with technology. Take a closer look at how the form of games, learning, and communication has been changed [1]. This amalgamation of one technology with another technology creates a valuable asset to interact with. But what if the shared augmented reality data is not secure or vulnerable to potential cyber-attacks that include Phishing, Man-In-TheMiddle attack, Denial of Service attack, and Rootkits [2]? As a corollary, considering how to safeguard them using digital security methodologies is the basic requirement. This research aims to make the transfer of marker-based 3 Dimensional (3D) augmented reality data in a secured way using modified hill cipher cryptography methods. The original data is obscured by encrypting and decrypting it with a key. Authentication, Confidentiality, Integrity, and Non-repudiation are all requirements for the cryptosystems that will be fulfilled [3]. Humans adore augmented reality, yet it contains certain hidden dangers like hijacking of devices and its media, or introduction of viruses and worms from the internet. So, this is just the first step toward the security and protection of augmented reality technology.

## II. BACKGROUND

Immersive technology is about the creation of the digital and virtual world around us. It has piqued the curiosity of investors and the general public in the previous five years, with financial investment skyrocketing after Mark Zuckerberg purchased Oculus for 2 billion dollars. Many other businesses, such as Sony, Samsung, HTC, and Google, are currently investing heavily in Virtual Reality (VR) and Augmented Reality [4]. In respect of security breaches, form-jacking code is compromised on an average of 4,800 websites per month. Internal actors were present in 34 percent of data breaches in 2018. 71 percent of data breaches are motivated by monetary gain. Ransomware is responsible for approximately a quarter of all malware related events. The global average cost of a data breach is 3.86 million dollars.[5] In 2020, the country with the highest average total cost of a data breach was the United States at 8.64 million dollars. The average cost per lost or stolen record in a data breach is 150 dollars [6]. This defines the lack in the security system and the urgent need of security of all digital media and systems.

## III. SOME BASIC TERMINOLOGY

### A. Augmented Reality

Living life with a mobile device is all about the 2022 era and combining physical life with this virtualized technology to view new objects of different dimensions, presented in the actual world is what augmented reality is all about [7]. Augmented reality is a part of immersive technology. Augmented Reality is the fusion of the real world with the computer-generated or virtual world. This union of technologies has been created by augmenting the virtual world objects over the real world. There are four ways of creation namely marker-based augmented reality, markerless augmented reality, projection-based augmented reality, and superimposition–based augmented reality. The method used in the proposed research to create the 3 dimensional models, is marker based augmented reality technology that scans a trigger image to create the specified 3 dimensional model in an encrypted format using cryptographic method and can later be converted back to original 3 dimensional model[8].



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:45:49 UTC from IEEE Xplore. Restrictions apply.

## B. Cryptography

Cryptography is a technique that helps in achieving the confidentiality, authenticity, Integrity, availability and accessibility of data. This technology has been around for a long time, dating back to 1900 BC to protect the communication between the rulers. People at the time were also aware of the need for privacy and security. Various Italian and Papal states led to the rapid growth of cryptography techniques during and after the European Renaissance [9]. To decipher the secret codes, many analytical and hacking approaches were developed during this time frame. However, in present days, the privacy of individuals and organizations is provided through high-level cryptography methods, making sure that information sent is secured in a way that the authorized receiver can access this information.

# IV. Strength Weakness Opportunity Threat(SWOT) Analysis of Secured AR Model

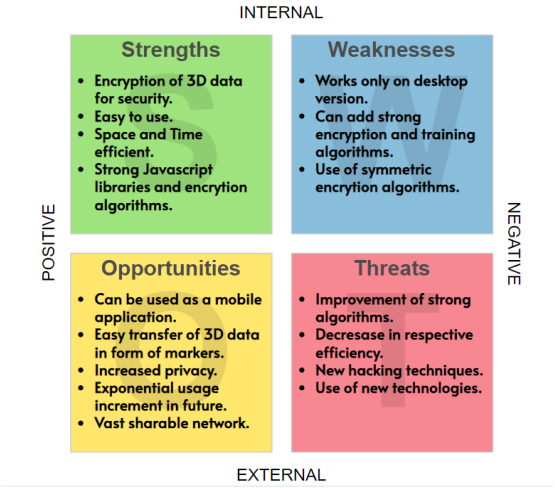



Fig. 1. SWOT matrix of AR model

## A. Strength

- *Encryption of 3 dimensional data for security:* Integrating Augmented Reality technology framework with cryptography using modified Hill cipher technique, which has a wide range of bits like 114 to 128, makes it the most reliable technique for security[10].
- *Easy to use: The proposed model just* requires the user to only scan the marker to create the 3 dimensional model[11].
- *Space and Time Efficient:* The use of the modified hill cipher technique over the 3 *dimensional* model requires less space and time efficiency is gained due to interaction of the image matrix and key matrix[12].
- *Strong JavaScript libraries and encryption algorithm:* To conquer the web structure, media, new technologies, HTML, JavaScript, AR.js, and Howler.js will work as strength[13][14].

## B. Weakness

- *Works only on the desktop version:* For now, only the desktop version of the proposed application is available.
- *Can add strong encryption and training algorithms:* New emerging strong algorithms can be used. And, merging the model with Machine Learning(ML) Algorithms can increase the efficiency.
- *Use of symmetric encryption algorithms: Due to the use of* symmetric encryption, the model becomes prone to known plaintext attack[15]

## C. Opportunity

- *Can be used as a mobile application:* Using application development tools, the proposed model can be used as a mobile application for sharing augmented reality media in a secure way.
- *Easy transfer of 3 dimensional data in form of markers:* Due to the presence of markers, the marker-based mobility of AR media is easy between the users.[16].
- *Increased privacy:* Encrypting the 3 dimensional images using cipher techniques, it decreases the vulnerability of media to the potential cyber attacks[17].
- *Exponential usage increment in future*: With new emerging tools and technology for digital media, Most preference will be on securing them from attacks.[18]
- *Vast shareable network:* Internet device all over the world creates a vast network for communication.[19]

## D. Threat

- *Improvement of strong algorithms:* The creation of more reliable algorithms will decrease the use of proposed algorithmic models due to efficiency factors.
- *Decrease in respective efficiency:* New emerging technologies, will decrease the respective efficiency of the proposed model.[20]
- *New hacking techniques:* Proposed model is prone to new cyber attack techniques [21].
- *Use of new technology:* Replacement of marker based AR with new AR methods can decrease the use of the proposed model [22].

# V. Methodology

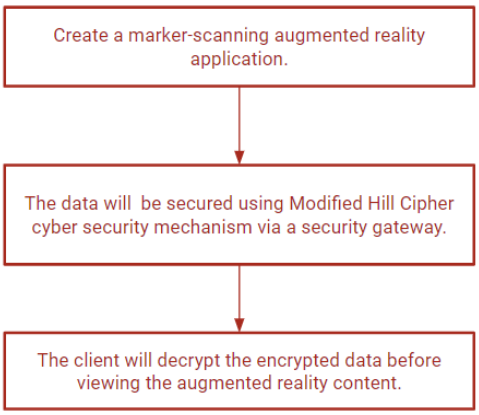



Fig. 2. Experimental design of proposed methodology



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:45:49 UTC from IEEE Xplore. Restrictions apply.

The notion described in this paper is based on an assessment of the current state of security breaches in the digital world, which has worsened over time, with scams and threats increasing exponentially and leading to data leaks to life threatening challenges over all sectors[23]. Furthermore, the increased use of unprotected interactive augmented reality work frames makes them vulnerable and exposes them to potential digital attacks by the hackers and intruders. In order to provide a security to the 3 dimensional media, a marker or 2 Dimensional barcode or marker is presented to the capturing device for scanning which carries the 3 dimensional model in it, the algorithm will scan the image, create the 3 dimensional image and convert it into an encrypted message using modified hill cipher technique[24], and to decipher the encrypted image, a keyword will be required that is provided by the sender, and using the decrypting modified hill cipher technique, original image will be captured. The efficiency of the model depends on the modified hill cipher algorithm used.

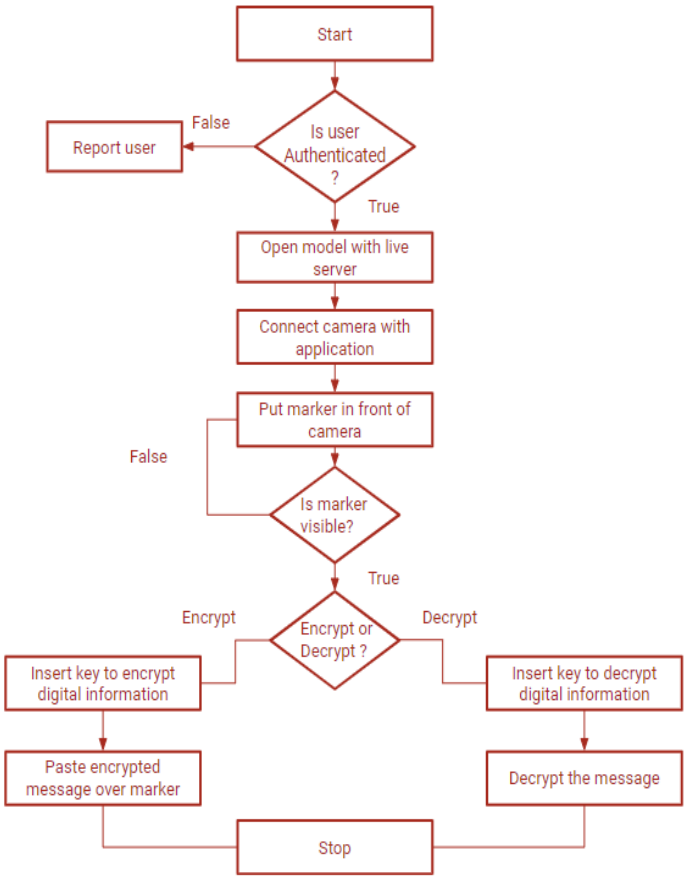



Fig. 3. Flowchart of proposed methodology

## VI. ALGORITHM

Let's assume

P = 3 dimensional plain object

C = 3 dimensional cipher object

Each P and C matrices are used to represent the 3 dimensional matrices where Red, Green, and Blue are the color layers.

Let p(x,y,z) and c(x,y,z) be the entry value at position (x,y,z) of the plain and cipher objects. p(x,y,z) and c(x,y,z) belong to 0,1,2,3,4,...255.

Let K is the key matrix, E and D represent encryption and decryption respectively.

Encryption: C = E(P,K) = E(R,K), E(G,K), E(B,K)
Decryption: P = D(C,K) = D(R,K), D(G,K), (B,K)

**Algorithm 1** Modified Hill Cipher On Augmented Reality Media.

**procedure** STEPS(Marker, Camera, key)
  Authenticate User.
  The Marker is clearly visible.
  Open model with the live server.
  Connect the camera with the application.
  Put the marker in front of the camera.

  **if** $Encrypt == True$ **then**
    Insert key to encrypt digital information.
    Paste encrypted message over the marker.
  **else**
    Insert key to decrypt digital information.
    Decrypt the message.
  **end if**
**end procedure**

## VII. RESULT

Fig. 4. 3 dimensional original image created from 2 dimensional marker



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:45:49 UTC from IEEE Xplore. Restrictions apply.

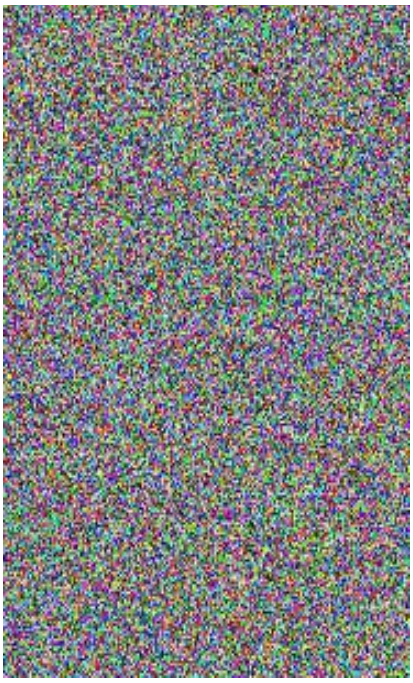


Fig. 5. Encrypted image by modified hill cipher technique

Fig. 6. Decrypted image by modified hill cipher technique

This research allows us to introduce a new aspect of security to the world of augmented reality and a protection strategy to immersive media. This method enables a sender to share their 3 dimensional framework without any risk of information leaks and potential breach of privacy. The most efficient encryption approach is being used to ensure smooth and secure transmission of data. Using a 3x3 matrix grid of integer numbers, produce a barcode or marker having greater than zero value of hamming code to create a 2 dimensional marker based 3 dimensional data. The bar code is intentionally designed in black and white to allow for clear presentation , quick image scanning and recognition. The actual 3 dimensional data is then encased using the modified Hill cipher cryptography approach on top of the marker. The following results show a 3 dimensional image originated from a 2 dimensional marker. After using the modified Hill cipher, the original image is encrypted and a secret or non-intelligible image will be generated. To make the encrypted image readable again i.e to get the original image at the receiver side, it is then decrypted using the reverse of modified hill cipher technique

## VIII. Conclusion

The growing transit of data over the network in the emerging digital era makes it more vulnerable to hackers and invaders[25] with variable effects in terms of financial threat, privacy threats, and criminal threats. Because of that same increasing mobility and threats, the digital content urges to be always protected by utilizing the security mechanisms used in this study. Data over the network is protected from hackers using the most economical and efficient cryptography technique[26]. In future, researchers may ponder upon improving and securing the mobility of augmented reality, virtual reality and mixed reality content to generate a trustworthy 3 *dimensional* immersive practical exposure.

## References


[1] R. Aggarwal and A. Singhal, ”Augmented Reality and its effect on our life,” 2019 9th International Conference on Cloud Computing, Data Science and Engineering (Confluence), 2019, pp. 510-515, doi: 10.1109/CONFLUENCE.2019.8776989

[2] Yuchong Li, Qinghui Liu, A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments, Energy Reports, Volume 7, 2021, Pages 8176-8186

[3] C. Thomas, N. Vickel, A. Laurent, Auckland University of Technology, New Zealand, Designing Collaborative Learning Environments Using Mobile AR, Association for the Advancement of Computing in Education (AACE) June 2015.

[4] C. Pietro, G. Irene, R. Mariano and R. Giuseppe, The Past, Present, and Future of Virtual and Augmented Reality Research: A Network and Cluster Analysis of the Literature 2018.

[5] R. Sobers, 134 Cybersecurity Statistics and Trends for 2021 16-03-2021.

[6] Rob Sobers, 98 Must-Know Data Breach Statistics

[7] What is augmented reality? (2020) newsCafe.

[8] Nicolocarpignoli artoolkit barcode markers collection (2022) https://github.com/nicolocarpignoli/artoolkitbarcode-markers-collection/tree/master.

[9] Origin of Cryptography tutorialspoint.

[10] F. Roesner, T. Kohno and D. Molnar, Security and privacy for augmented reality systems, Communications of the ACM, vol. 57, no. 4, pp. 88-96, 2014.

[11] Siltanen, Sanni. (2012). Theory and applications of marker based augmented reality.

[12] G. Akgun, ¨ Performance Analysis of Hill Cipher and Its Modifications.

[13] S. Delcev and D. Draskovic, ”Modern JavaScript frameworks: A Survey Study,” 2018 Zooming Innovation in Consumer Technologies Conference (ZINC), 2018, pp. 106-109, doi: 10.1109/ZINC.2018.8448444.

[14] A-Frame primitive element and applications.

[15] Chandra, Sourabh and Bhattacharyya, Siddhartha and Paira, Smita and Alam, Sk. (2014). A Study and Analysis on Symmetric Cryptography. 10.1109/ICSEMR.2014.7043664.




Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:45:49 UTC from IEEE Xplore. Restrictions apply.

[16] C. Guo, H. J. Wang and W. Zhu, Smart-phone attacks and defenses, HotNets III, 2014.

[17] L. Pombo and M. M. Marques, "Markerbased augmented reality application for mobile learning in an urban park: Steps to make it real under the EduPARK project," 2017 International Symposium on Computers in Education (SIIE), 2017, pp. 1-5, doi: 10.1109/SIIE.2017.8259669.

[18] Mariya James Future of Augmented Reality https://techdayhq.com/community/articles/future-ofaugmented-reality-ar.

[19] Arumugam, Saravanan and Subramanian, Sathya. (2019). A Review on Cyber Security and the Fifth Generation Cyberattacks. Oriental journal of computer science and technology. 12. 50-56. 10.13005/ojcst12.02.04.

[20] Abood, Omar and Guirguis, Shawkat. (2018). A Survey on Cryptography Algorithms. International Journal of Scientific and Research Publications. 8. 495-516. 10.29322/IJSRP.8.7.2018.p7978.

[21] Cheng, Jack and Chen, Keyu and Chen, Weiwei. (2017). Comparison of marker-based AR and markerless AR: A case study on indoor decoration system. 10.24928/JC3-2017/0231.

[22] Cheng, Long and Liu, Fang and Yao, Danfeng. (2017). Enterprise data breach: causes, challenges, prevention, and future directions: Enterprise data breach. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 7. e1211. 10.1002/widm.1211.

[23] Hill Cipher https://en.wikipedia.org/wiki/Hill cipher.

[24] F. Roesner, T. Kohno, T. Denning, R. Calo and B. C. Newell, Augmented reality: Hard problems of law and policy, Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, pp. 1283-1288, 2014.

[25] AMIN Mohammed2, DIAB Hossam2, "How to repair the Hill cipher".

[26] M. Kesima, Y. Ozarslan Augmented Reality in Education: Current Technologies and the Potential for Education 2018.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:45:49 UTC from IEEE Xplore. Restrictions apply.

# Entanglement in interacting quenched two-body coupled oscillator system

Sayantan Choudhury,[1,*] Rakshit Mandish Gharat,[2,†] Saptarshi Mandal,[3,‡] Nilesh Pandey,[4,§] Abhishek Roy,[5,‖] and Partha Sarker[6,¶]

[1]*International Centre for Theoretical Sciences, Tata Institute of Fundamental Research (ICTS-TIFR), Shivakote, Bengaluru 560089, India*
[2]*Department of Physics, National Institute of Technology Karnataka, Surathkal, Karnataka-575025, India*
[3]*Department of Physics, Indian Institute of Technology Kharagpur, Kharagpur-721302, India*
[4]*Department of Applied Physics, Delhi Technological University, Delhi-110042, India*
[5]*Department of Physics, Indian Institute of Technology Jodhpur, Karwar, Jodhpur-342037, India*
[6]*Department of Physics, University of Dhaka, Curzon Hall, Dhaka 1000, Bangladesh*

(Received 18 April 2022; accepted 26 June 2022; published 7 July 2022)

In this work, we explore the effects of a quantum quench on the entanglement measures of a two-body coupled oscillator system having quartic interaction. We use the invariant operator method, under a perturbative framework, for computing the ground state of this system. We give the analytical expressions for the total and reduced density matrix of the system having non-Gaussian, quartic interaction terms. Using this reduced density matrix, we show the analytical calculation of two entanglement measures viz., Von Neumann entanglement entropy using replica trick and Renyi entanglement entropy. Further, we give a numerical estimate of these entanglement measures with respect to the dimensionless parameter $(t/\delta t)$ and show its behavior in the three regimes, i.e., late time behavior, around the quench point and the early time behavior. We comment on the variation of these entanglement measures for different orders of coupling strength. The variation of Renyi entropy of different orders has also been discussed.

DOI: 10.1103/PhysRevD.106.025002

## I. INTRODUCTION

In recent years, the most important works in theoretical physics have been studied by blending the ideas of quantum many-body physics, quantum information theory and quantum field theory. The amalgamation of these branches of physics have been reviewed in [1–7]. These works mostly focus on studying the dynamical properties of quantum entanglement in time-dependent systems [8–17]. This dynamical entanglement can be characterized by computing von Neumann entanglement entropy and Renyi entanglement entropy. The traditional way of computing these entanglement measures involves constructing the reduced density matrix using the eigenstates of the time-dependent Hamiltonian.

One of the ways to compute the eigenstates of such time-dependent Hamiltonians is by constructing the Lewis-Resenfield invariant operator and this is often termed as invariant operator representation of the wave function [18]. Some works following this approach to compute the time-dependent eigenstates are [19–23]. Evolution of such time-dependent quantum states can be studied as a solution to the Ermakov-Milne-Pinney equation [24–26]. The invariant operator method is generalized for perturbed theories by computing time-independent perturbative corrections [27], considering adiabatic evolution [28,29] of the parameters.

For unperturbed Hamiltonians or free theories, the reduced density matrix once computed represents the Gaussian states. Entanglement can then be computed using these Gaussian states [30–32,32]. Computing von Neumann entropy for these Gaussian states employs the use of correlation functions defined using the framework of quantum field theory [33].

Entanglement in interacting theories has been studied using replica trick [34] in [35]. On the other hand, the perturbed entanglement entropy is computed using the path-integral approach in [36].

Most of the recent works in many-body physics have been about contemplating the behavior of entanglement for a system having a time-dependent parameter in

*sayantan.choudhury@icts.res.in, sayanphysicsisi@gmail.com
†rakshitmandishgharat.196ph018@nitk.edu.in
‡saptarshijhikra@gmail.com
§nilesh911999@gmail.com
‖roy.1@iitj.ac.in
¶sarker239@gmail.com

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.*

Hamiltonian. This time-dependent parameter can be varied quickly or slowly and hence the process is termed as a "quantum quench." Some of the most important works for studying entanglement-properties of extended systems undergoing a quantum quench are [37–46]. These quantum quenches can be thought of as protocols driving the system out-of-equilibrium [47–49]. These local equilibrium can then be analyzed in quenched systems by using reduced density matrix. In recent years, effects of quantum quenches have even been studied experimentally using cold atom systems [50–59]. Studying the effect of quantum quench in the case of interacting theories or perturbed theories is of prime importance [60].

In recent years, the study of coupled oscillators has been an area of active research. This is primarily due to the extensive use of such models in studying quantum and nonlinear physics [61–65], molecular chemistry [66–68] and biophysics [69–71]. Especially, in quantum physics, analyzing entanglement of coupled oscillators is of prime importance [72–76].

Motivated by the discussion given above, in this work we consider a toy model with a Hamiltonian of two coupled oscillators having quartic self-coupling term. The Hamiltonian for this system consists of a time-dependent quench profile. The eigenstates for this time-dependent Hamiltonian are computed using invariant operator method, in a perturbative framework. Further, the reduced density matrix (with quartic interaction terms) is constructed using time-independent perturbation theory. The dynamical von Neumann entropy and Renyi entropy are then derived using appropriate formulas for the obtained reduced density matrix. We comment on the behavior of these entanglement measures by varying the relevant parameters.

The organization of the paper is as follows:

(i) We start our discussion by providing an overview of the quench protocol and Hamiltonian of the system in Sec. II.

(ii) In Sec. III, the expression for eigenstates of the time-dependent Hamiltonian is computed using invariant-operator representation of wave function. Further, the expression for first order time-independent correction to the ground state of the Hamiltonian is also approximated in this section.

(iii) In Sec. IV, we use the ground state wave function with perturbative correction to compute the expression for reduced density matrix, with quartic interaction terms. von Neumann entanglement entropy is then computed by performing the replica trick over this reduced density matrix. Further, we show the analytically computed expression for Renyi entanglement entropy.

(iv) In Sec. V, we numerically evaluate the respective entanglement measures and plot them with respect to the dimensionless parameter $(t/\delta t)$. We comment on the parametric variation of these entanglement measures for each of the chosen three regimes.

(v) Section VI summarizes the conclusions we draw from the obtained results of this work with some interesting future prospects of our present work.

## II. THE SETUP AND THE QUENCH PROTOCOL

In this section we begin by discretizing the Hamiltonian for a scalar field theory with $\phi^4$ interaction term on a lattice. We show that the Hamiltonian then represents a family of infinite anharmonic oscillators with quartic couplings. In this article, we study a system of two coupled oscillators having quartic perturbation. Furthermore, we use normal mode basis to decouple the Hamiltonian so that we can compute the eigenstates for this system in a much simpler way, in upcoming section. Also, we mention the time-dependent quench profile chosen as the frequency of this Hamiltonian.

The Hamiltonian for a scalar field theory with a $\hat{\lambda}\phi^4$ interaction is given by [77],

$$\mathcal{H}=\frac{1}{2}\int d^{d-1}x\left[\pi(x)^2+(\nabla\phi(x))^2+m^2\phi(x)^2+\frac{\hat{\lambda}}{12}\phi(x)^4\right]. \tag{1}$$

Here $d$ is the space-time dimensions. We assume that the coupling $\hat{\lambda}\ll 1$, so that we can work in a perturbative framework. This theory can be discretized on a $d-1$ dimensional lattice, which is characterized by lattice spacing, $\delta$. It can be shown that, the discretized Hamiltonian becomes,

$$\mathcal{H}=\frac{1}{2}\sum_{\vec{n}}\left\{\frac{\pi(\vec{n})^2}{\delta^{d-1}}+\delta^{d-1}\left[\frac{1}{\delta^2}\sum_i(\phi(\vec{n})-\phi(\vec{n}-\hat{x}_i))^2+m^2\phi(\vec{n})^2+\frac{\hat{\lambda}}{12}\phi(\vec{n})^4\right]\right\}. \tag{2}$$

Here $\vec{n}$ denotes the spatial location of the points on lattice and $\hat{x}_i$ represent the unit vectors along the lattice. Further, we introduce the following substitutions to simplify the form of the Hamiltonian:

$$\begin{aligned}\hat{X}(\vec{n})&=\delta^{d/2}\phi(\vec{n}), \qquad \hat{P}(\vec{n})=\pi(\vec{n})/\delta^{d/2},\\ M&=\frac{1}{\delta}, \qquad \omega=m,\\ \eta&=\frac{1}{\delta}, \qquad \lambda=\frac{\hat{\lambda}}{24}\delta^{-d},\end{aligned} \tag{3}$$

where $\omega$ represents the frequency of individual oscillators and $\eta$ denotes intermass coupling. After these substitutions we get,

$$\mathcal{H}=\sum_{\vec{n}}\left\{\frac{\hat{P}(\vec{n})^2}{2M}+\frac{1}{2}M\left[\omega^2\hat{X}(\vec{n})^2+\eta^2\sum_i(\hat{X}(\vec{n})-\hat{X}(\vec{n}-\hat{x}_i))^2 +2\lambda\hat{X}(\vec{n})^4\right]\right\}. \tag{4}$$

The above Hamiltonian, in Eq. (4) represents a family of infinite coupled anharmonic oscillators. In this work we focus on the system representing two coupled oscillators and compute the entanglement for this system. Setting $M=1$, for simplicity, Eq. (4) can be specialized for case of two coupled oscillators,

$$H=\frac{1}{2}[p_1^2+p_2^2+\omega^2(x_1^2+x_2^2)+\eta^2(x_1-x_2)^2 +2\{\lambda(x_1^4+x_2^4)\}]. \tag{5}$$

Here $x_i$ and $p_i$, for $i=1,2$ denote the canonical coordinates of the respective oscillator following the standard commutation relation $[x_i,p_j]=i\delta^i_j$, while $\lambda$ denotes the coupling coefficient of $\phi^4$ interaction term.

The eigenstates of the above Hamiltonian Eq. (5), can easily be computed using normal coordinates defined as,

$$\begin{aligned}X_1&=(x_1+x_2)/\sqrt{2}\\ X_2&=(x_1-x_2)/\sqrt{2}\\ P_1&=(p_1+p_2)/\sqrt{2}\\ P_2&=(p_1-p_2)/\sqrt{2}.\end{aligned} \tag{6}$$

The unperturbed part of Hamiltonian when written using these normal coordinates decouples. One can then show that the total Hamiltonian of Eq. (5) in normal coordinates takes the following form:

$$H=\sum_{i=1}^{2}H_i+H_p,$$

where,

$$H_i(T)=\frac{1}{2}({P_i}^2+\omega_i^2(T){X_i}^2), \tag{7}$$

denotes the unperturbed and decoupled Hamiltonian for each of the two oscillators. On the other hand the perturbed Hamiltonian is given by,

$$\begin{aligned}H_p&=\lambda V=\lambda(x_1^4+x_2^4)\\ &=\lambda(X_1^4+X_2^4+6X_1^2X_2^2).\end{aligned} \tag{8}$$

This enables us to use $\lambda\phi^4$ time-independent perturbation theory and compute the eigenstates of the total Hamiltonian of Eq. (5).

We now consider the frequency $\omega$ in, Eq. (5) as a time-dependent quench profile. One of the most common quench profiles used in literature [78,79] is given by:

$$\omega^2(t/\delta t)=\omega_0^2\left[\tanh^2\left(\frac{t}{\delta t}\right)\right]. \tag{9}$$

Here $\omega_0$ can be interpreted as a free parameter and $\delta t$ is the quench parameter or the quench rate. The quench profile chosen here is such that it admits an exact solution for the mode functions given in [79] and the quench profile attains a constant value at late and early times. The dynamical process due to this profile happens in the $[-\delta t,\delta t]$ time window. We will set $t/\delta t=T$ and $\omega_0=1$. The respective frequencies in normal mode basis take the following form,

$$\omega_1=\omega(T)\quad\text{and}\quad\omega_2=\sqrt{\omega^2(T)+4\eta^2}, \tag{10}$$

where $\omega(T)$ is the quench profile Eq. (9).

Note that the unperturbed Hamiltonian of Eq. (7) is now time-dependent while the perturbed Hamiltonian of Eq. (8) acts as time-independent $\phi^4$ coupling applied on the two harmonic oscillators. In Sec. III, we show the analytical computation of ground state, $\Psi_{0,0}$ of the total Hamiltonian of two coupled anharmonic oscillators having $\lambda\phi^4$ perturbation. This ground state is used to derive the analytical expressions of the respective entanglement measures in Sec. IV.

## III. CONSTRUCTING WAVE FUNCTION FOR A $\phi^4$ QUENCH MODEL

In this section our prime objective is to construct the wave function approximated to first order in coupling constant $\lambda$. In Sec. III A we compute the eigenstates of decoupled and unperturbed Hamiltonian Eq. (7). These eigenstates are then used to construct the ground state of perturbed Hamiltonian Eq. (8), approximated to first perturbative order, in Sec. III B. Finally we compute the total wave function as ground state of total Hamiltonian Eq. (5).

### A. Eigenstates and eigenvalues for unperturbed Hamiltonian

As, the unperturbed Hamiltonian Eq. (7) decouples in the normal mode basis, the eigenstates for the unperturbed Hamiltonian are just the product of the eigenstates of respective oscillators in the normal-mode basis:

$$\psi^{(0)}_{n_1,n_2}(X_1,X_2,T)=\psi_{n_1}(X_1,T)\psi_{n_2}(X_2,T). \tag{11}$$

Since the unperturbed Hamiltonian consists of a time-dependent frequency scale, we now use a prescription often termed as the invariant representation in the literature [80], to get the unperturbed eigenstates.

We begin the invariant representation by listing the auxiliary equations. The solutions to these equations can

then be used to compute the coupled wave function. The auxiliary equations can be written as:

$$\ddot{\sigma}_j - \sigma_j \dot{\gamma}_j + \omega_j^2(T)\sigma_j = 0$$
$$\sigma_j \ddot{\gamma}_j + 2\dot{\sigma}_j \dot{\gamma}_j = 0. \tag{12}$$

Here, $j = 1, 2$ and $\sigma_j(T)$ and $\gamma_j(T)$ are time-dependent factors for each of the two coupled oscillators. Also, $\dot{\gamma}_j = \partial_T \gamma_j$, $\dot{\sigma}_j = \partial_T \sigma_j$ and $\ddot{\sigma}_j = \partial_T^2 \sigma_j$. The subscript $j$ denotes the oscillator for which the respective parameter is mentioned. The computation of explicit values of $\sigma(T)$ and $\gamma(T)$ is discussed in Appendix A. Note that we have suppressed the time-dependence throughout this section. Next, we define the creation $(a_j^\dagger)$ and annihilation $(a_j)$ operators given by,

$$a_j = \frac{1}{\sqrt{2\dot{\gamma}_j}}\left[\dot{\gamma}_j\left(1 - i\frac{\dot{\sigma}_j}{\sigma_j \dot{\gamma}_j}\right)X_j + iP_j\right]$$
$$a_j^\dagger = \frac{1}{\sqrt{2\dot{\gamma}_j}}\left[\dot{\gamma}_j\left(1 + i\frac{\dot{\sigma}_j}{\sigma_j \dot{\gamma}_j}\right)X_j - iP_j\right]. \tag{13}$$

Here, $j = 1, 2$. One can show that these operators satisfy the commutation relation $[a_i, a_j^\dagger] = \delta_j^i$. The creation and annihilation operators can be used to define invariant operator for the respective decoupled Hamiltonian,

$$I_j = \Omega_j\left(a_j^\dagger a_j + \frac{1}{2}\right). \tag{14}$$

Here, $j = 1, 2$. On the other hand, $\Omega_j = \sigma_j^2 \dot{\gamma}_j$, is an invariant quantity with respect to time. The construction of this invariant operator Eq. (14) has been briefly outlined in Appendix B. The invariant operator has its own spectrum and eigenstates. The eigenstates of invariant operator can be used to formulate the wave functions for each decoupled Hamiltonian. The outline of the same is given in Appendix B. Using equation (B4) for $n_1$, $n_2 = 0$, one can show that the ground state of unperturbed Hamiltonian is given by,

$$\psi_{0,0}^{(0)} = \sqrt{\frac{g_1 g_2}{\pi}}\exp\left[-i\frac{\gamma_1 + \gamma_2}{2}\right]$$
$$\times \exp\left[-\frac{1}{2}(g_1^2(1 - id)X_1^2 + g_2^2(1 - if)X_2^2)\right]. \tag{15}$$

where, the coefficients $g_1$, $g_2$, $d$, and $f$ are given by,

$$g_1 = \sqrt{\dot{\gamma}_1}, \quad g_2 = \sqrt{\dot{\gamma}_2}, \quad d = \frac{\dot{\sigma}_1}{\dot{\gamma}_1 \sigma_1}, \quad f = \frac{\dot{\sigma}_2}{\dot{\gamma}_2 \sigma_2}. \tag{16}$$

Next we emphasize that the eigenvalues of the unperturbed decoupled Hamiltonians in Eq. (7) will have a time dependent factor [27]. These eigenvalues for each of the decoupled Hamiltonians are given as:

$$\langle \psi_{n_j} | H_i | \psi_{n_j} \rangle = W_j(T)\left[n_j + \frac{1}{2}\right], \tag{17}$$

where, $j = 1, 2$. Here $W_j(T)$ is the time-dependent factor for each oscillator given by,

$$W_j(T) = \frac{\dot{\gamma}_j}{2}\left(\frac{\dot{\sigma}_i + \sigma_i^2\omega_i^2 + \sigma_i^2\dot{\gamma}_i}{\sigma_i^2 \dot{\gamma}_i{}^2}\right), \tag{18}$$

where $j = 1, 2$. Using the above eigenvalues one can write the energy eigenvalue for the unperturbed state of two coupled oscillators, Eq. (B4) as:

$$\langle \psi_{n_1,n_2}^{(0)} | H | \psi_{n_1,n_2}^{(0)} \rangle = W_1(T)\left(n_1 + \frac{1}{2}\right) + W_2(T)\left(n_2 + \frac{1}{2}\right). \tag{19}$$

### B. Ground state of two coupled oscillators with first order-$\phi^4$ perturbation

Using time-independent perturbation theory, one can show that the first order perturbative correction to the ground state of two-coupled oscillators is,

$$\psi_{0,0}^{(1)} = \sum_{(n_1,n_2)\neq(0,0)} \frac{\langle \psi_{n_1,n_2}^{(0)} | V | \psi_{0,0}^{(0)} \rangle \times \psi_{n_1,n_2}^{(0)}.}{\langle \psi_{0,0}^{(0)} | H | \psi_{0,0}^{(0)} \rangle - \langle \psi_{n_1,n_2}^{(0)} | H | \psi_{n_1,n_2}^{(0)} \rangle} \tag{20}$$

Using the form of perturbed Hamiltonian Eq. (8) and the time dependent eigenvalues Eq. (19), the above expression when evaluated becomes,

$$\psi_{0,0}^{(1)} = -\frac{3(g_1^2 + g_2^2)\psi_{0,2}^{(0)}}{4\sqrt{2}hg_1^2 g_2^2} - \frac{3\psi_{0,4}^{(0)}}{8\sqrt{2}hg_2^2} - \frac{3(g_1^2 + g_2^2)\psi_{2,0}^{(0)}}{4\sqrt{2}gg_1^2 g_2^2}$$
$$- \frac{3\psi_{2,2}^{(0)}}{2(2g + 2h)2g_1^2 g_2^2} - \frac{3\psi_{4,0}^{(0)}}{8\sqrt{2}gg_1^2}. \tag{21}$$

Here,

$$g = \left(\frac{\dot{\sigma}_1 + \sigma_1^2\omega_1^2 + \Omega_1\dot{\gamma}_1}{\Omega_1}\right),$$
$$h = \left(\frac{\dot{\sigma}_2 + \sigma_2^2\omega_2^2 + \Omega_2\dot{\gamma}_2}{\Omega_2}\right). \tag{22}$$

The explicit form of the first order correction can be computed using the expression of unperturbed eigenstates, Eq. (B4). The total wave function for ground state of total Hamiltonian Eq. (5), corrected to first order of time-independent $\lambda\phi^4$ perturbation, is given by: $\Psi_{0,0} = \psi_{0,0}^{(0)} + \lambda\psi_{0,0}^{(1)}$. Using Eq. (15) and Eq. (21) while approximating, the coupling constant $\lambda \ll 1$ we can express the final form of the wave function in normal mode basis as:

$$\Psi_{0,0}(X_1,X_2)=\left(\frac{g_1^2g_2^2}{\pi^2}\right)^{\frac{1}{4}}e^{-i\frac{(\gamma_1+\gamma_2)}{2}}\exp\left[-\frac{1}{2}(1-id)g_1^2X_1^2-\frac{1}{2}(1-if)g_2^2X_2^2+\lambda(A_1+A_2X_1^2+A_3X_2^2+A_4X_1^4+A_5X_2^4+A_6X_1^2X_2^2)\right]. \tag{23}$$

The coefficients $A_i$ for $i=1$ to $i=6$ are mentioned in a table given in Appendix D. The above wave function Eq. (23) represents the ground state of total Hamiltonian in Eq. (5), of the system of two coupled oscillators with $\lambda\phi^4$ perturbation. We take note of the fact that all variables, aside from the coordinates, $X_1$, $X_2$ and coupling constant, $\lambda$ in the wave function, Eq. (23) are functions of timescale $T$. The wave function is then dependent on both $t$ and $\delta t$. This explicit time dependence can be evaluated by computing $\sigma_i$ and $\gamma_i$, shown in Appendix A.

## IV. ANALYTICAL CALCULATION OF ENTANGLEMENT MEASURES

In the previous Secs. II and III, we computed the ground state wave function for a system of two coupled bosonic oscillators with a $\phi^4$ first-order perturbative correction for Hamiltonian having a quenched frequency-profile. In this section our prime objective is to show the analytical steps to calculate entanglement measures, viz., von Neumann entanglement entropy and Renyi entropy.

In Sec. IV A reduced density matrix for the system of two coupled oscillators is constructed using the wave function Eq. (23). To compute von Neumann entropy using replica trick [34,35] as well as Renyi entropy, the trace of reduced density matrix should be evaluated, this is shown in Sec. IV B. Finally using the appropriate formulas we show the computation of the respective entanglement measures in Sec. IV C.

### A. Density matrix for perturbed wave function

We begin by transforming the wave function $\Psi(X_1,X_2)$ given in Eq. (23) to $\Psi(x_1,x_2)$, i.e., we transform the normal coordinates back to space-time coordinates using Eq. (6). We mention four new symbols:

$$P=\frac{1}{2}(1-id)g_1^2;\qquad P^*=\frac{1}{2}(1+id)g_1^2$$
$$Q=\frac{1}{2}(1-if)g_2^2;\qquad Q^*=\frac{1}{2}(1+if)g_1^2. \tag{24}$$

The wave function in spacetime coordinates is then represented by:

$$\begin{aligned}\Psi(x_1,x_2)=&\left(\frac{g_1^2g_2^2}{\pi^2}\right)^{\frac{1}{4}}e^{-i(\gamma_1+\gamma_2)/2}\exp\left\{-\frac{P}{2}(x_1^2+x_2^2+2x_1x_2)-\frac{Q}{2}(x_1^2+x_2^2-2x_1x_2)+\lambda\left[A_1+\frac{A_2}{2}(x_1^2+x_2^2+2x_1x_2)\right.\right.\\&+\frac{A_3}{2}(x_1^2+x_2^2-2x_1x_2)+\frac{A_4}{4}(x_1^4+x_2^4+4x_1^3x_2+4x_1x_2^3+6x_1^2x_2^2)\\&\left.\left.+\frac{A_5}{4}(x_1^4+x_2^4-4x_1^3x_2-4x_1x_2^3+6x_1^2x_2^2)+\frac{A_6}{4}(x_1^4+x_2^4-2x_1^2x_2^2)\right]\right\}.\end{aligned} \tag{25}$$

The complex conjugate of the above given wave function is denoted by $\Psi^*(x_1',x_2')$. Using the conjugate of the wave function in Eq. (25), we can construct the total density matrix for the system of two oscillators by $\rho(x_1,x_2,x_1',x_2')=\Psi(x_1,x_2)\Psi^*(x_1',x_2')$. One can easily show that the density matrix is given as:

$$\begin{aligned}\rho(x_1,x_2,x_1',x_2')=&\left(\frac{g_1g_2}{\pi}\right)\exp\left\{-\frac{P}{2}(x_1^2+x_2^2+2x_1x_2)-\frac{P^*}{2}(x_1'^2+x_2'^2+2x_1'x_2')-\frac{Q}{2}(x_1^2+x_2^2-2x_1x_2)-\frac{Q^*}{2}(x_1'^2+x_2'^2-2x_1'x_2')\right.\\&+\lambda\left[2A_1+\frac{A_2}{2}(x_1^2+x_1'^2+x_2^2+x_2'^2+2x_1x_2+2x_1'x_2')+\frac{A_3}{2}(x_1^2+x_1'^2+x_2^2+x_2'^2-2x_1x_2-2x_1'x_2')\right.\\&+\frac{A_4}{4}(x_1^4+x_1'^4+x_2^4+x_2'^4+4x_1^3x_2+4x_1'^3x_2'+4x_1x_2^3+4x_1'x_2'^3+6x_1^2x_2^2+6x_1'^2x_2'^2)\\&+\frac{A_5}{4}(x_1^4+x_1'^4+x_2^4+x_2'^4-4x_1^3x_2-4x_1'^3x_2'-4x_1x_2^3-4x_1'x_2'^3+6x_1^2x_2^2+6x_1'^2x_2'^2)\\&\left.\left.+\frac{A_6}{4}(x_1^4+x_1'^4+x_2^4+x_2'^4-2x_1^2x_2^2-2x_1'^2x_2'^2)\right]\right\}.\end{aligned} \tag{26}$$

The reduced density matrix can be computed using total density matrix, shown in Eq. (26) by tracing over the coordinates of second oscillator, i.e., by setting $x_2'=x_2$ and computing $\rho(x_1,x_1',x_2)$. The reduced density matrix can then be evaluated as:

$$\rho(x_1, x_1') = \int_{-\infty}^{\infty} \rho(x_1, x_1', x_2) dx_2. \tag{27}$$

We mention the final form of the reduced density matrix which can be used to calculate the entanglement measures as:

$$\rho(x_1,x_1') = N\left(\frac{g_1 g_2}{\pi}\right)\sqrt{\frac{\pi}{A}}\exp\Bigg\{(2\lambda A_1+\lambda\alpha_0)+\left(\frac{-(P+Q)}{2}+\frac{(Q-P)^2}{4A}+\lambda\alpha_1\right)x_1^2+\left(\frac{-(P^*+Q^*)}{2}+\frac{(Q^*-P^*)^2}{4A}+\lambda\alpha_2\right)x_1'^2$$
$$+\left(\frac{(Q-P)(Q^*-P^*)}{2A}+\lambda\alpha_3\right)x_1x_1'+\lambda[\alpha_4 x_1^4+\alpha_5 x_1'^4+\alpha_6 x_1^3 x_1'+\alpha_7 x_1 x_1'^3+\alpha_8 x_1^2 x_1'^2]\Bigg\}. \tag{28}$$

Here $N$ is the normalization factor, while,

$$A = \left(\frac{g_1^2+g_2^2}{2}\right)$$
$$B = ((Q-P)x_1 + (Q^*-P^*)x_1'). \tag{29}$$

The values of coefficients, $\alpha_i$ for $i=0$ to $i=8$, are listed in Appendix D.

### B. Computing the value of $tr(\rho^n)$

In this subsection we outline analytical steps to compute the expression for $tr(\rho^n)$ where $\rho$ is the reduced density matrix of Eq. (28). The reduced density matrix, given by Eq. (28) is clearly representing a non-Gaussian state. The calculation of entanglement entropy for such a non-Gaussian state is shown in [35]. We follow a similar method and modify the same derivation to compute entropy for the state representing Eq. (28).

According to the definition of trace, considering that $x_{n+1}=x_1$, one can write:

$$tr(\rho^n) = \int dx_1 dx_2 \cdots dx_n \rho(x_1,x_2)\rho(x_2,x_3)\cdots\rho(x_n,x_1), \tag{30}$$

where, $\rho(x_i, x_{i+1})$, for $i=1$ to $n$, represents the reduced density matrix given in Eq. (28). The product of density matrices in Eq. (30), when evaluated gives:

$$tr(\rho^n) = N^n\left(\frac{g_1g_2}{\pi}\right)^n\left(\sqrt{\frac{\pi}{A}}\right)^n e^{n(2\lambda A_1+\lambda\alpha_0)}\int d^n x \exp\Bigg\{\left[\frac{-(P+P^*+Q+Q^*)}{2}+\frac{(Q-P)^2+(Q^*-P^*)^2}{4A}+\lambda(\alpha_1+\alpha_2)\right]\sum_{i=1}^n x_i^2$$
$$+\left[\frac{(Q-P)(Q^*-P^*)}{2A}+\lambda\alpha_3\right]\sum_{i=1}^n x_i x_{i+1}+\lambda\left[(\alpha_4+\alpha_5)\sum_{i=1}^n x_i^4+(\alpha_6 x_{i-1}^3+\alpha_7 x_{i+1}^3)\sum_{i=1}^n x_i+\alpha_8\sum_{i=1}^n x_i^2\sum_{i=1}^n x_{i+1}^2\right]\Bigg\}. \tag{31}$$

We modify the above expression by using new coefficients $\beta_1, \beta_2 \ldots \beta_7$ tabulated in Appendix D. Note that each $\beta_i$ for $i=1$ to 7 is evaluated by substituting $P$, $Q$, $A$, and $B$ defined in Eq. (24) and Eq. (29). One can then show that,

$$tr(\rho^n) = N^n\left(\frac{g_1g_2}{\pi}\right)^n\left(\sqrt{\frac{\pi}{A}}\right)^n e^{n(2\lambda A_1+\lambda\alpha_0)}\int d^n x\Bigg\{\exp\left[(\beta_1+\lambda\beta_2)\sum_{i=1}^n x_i^2+(\beta_3+\lambda\beta_4)\sum_{i=1}^n x_i x_{i+1}\right]$$
$$\times\exp\Bigg\{\lambda\left[\beta_5\sum_{i=1}^n x_i^4+(\alpha_6 x_{i-1}^3+\alpha_7 x_{i+1}^3)\sum_{i=1}^n x_i+\beta_6\sum_{i=1}^n x_i^2\sum_{i=1}^n x_{i+1}^2\right]\Bigg\}\Bigg\}. \tag{32}$$

The integral in the above Eq. (32) can be solved using the steps shown in Appendix C.

To evaluate normalization factor $N$ one needs to set $tr(\rho(x_1,x_1'))=1$ for the reduced density matrix in Eq. (28). Using $\mu$ and $\xi$ defined in Appendix C the normalization factor is given by,

$$N^n = \left(\frac{g_1g_2}{\pi}\right)^{-n}\left(\sqrt{\frac{\pi}{A}}\right)^{-n} e^{-n(2\lambda A_1+\lambda\alpha_0)}\left(\frac{\beta}{\pi}\right)^{\frac{n}{2}}|1-\mu|^n\left(1-\frac{3\lambda(\beta_5+\beta_6+\beta_7)}{4\beta^2(1-\mu)^4}\right)^n. \tag{33}$$

Substituting (33) and Eq. (C8) in Eq. (C12) from Appendix C, one can obtain the final value of trace of $n$th order of the reduced density matrix of Eq. (28), as:

$$tr(\rho^n) = \frac{|1-\mu|^n}{|1-\mu^n|}\left(1 - \frac{3\lambda(\beta_5+\beta_6+\beta_7)}{4\xi^2(1-\mu)^4}\right)^n \{1 + n\lambda[3\beta_5(M_{11}^{-1})^2 + \beta_6(M_{11}^{-1})^2 + 2(M_{12}^{-1})^2 + 3\beta_7(M_{11}^{-1}M_{12}^{-1})]\}. \quad (34)$$

### C. Entanglement measures

In this subsection we compute the two entanglement measures, viz., von Neumann entanglement entropy and Renyi entropy using the respective formulas for the reduced density matrix of Eq. (28).

The von Neumann entanglement entropy for a given density matrix $\rho$ is computed as, $S_{VN} = -tr(\rho \ln \rho)$. As we know the explicit $n-$ dependence of $tr(\rho^n)$ from Eq. (34), we instead use replica trick often given as [34],

$$S_{VN} = -\lim_{n\to 1}\frac{\partial}{\partial n} tr(\rho^n). \quad (35)$$

Substituting the respective values of matrix inverses of, Eq. (C13) in Eq. (34) one can show that the von Neumann entropy is given by,

$$S_{VN} = -\frac{\mu\ln\mu + (1-\mu)\ln(1-\mu)}{(1-\mu)} + \lambda\left[\frac{3\mu\ln\mu}{\xi^2(\mu+1)(\mu-1)^5}\beta_5 + \frac{\ln\mu(1+\mu+\mu^2)}{\xi^2(\mu+1)(\mu-1)^5}\beta_6 + \frac{3\ln\mu(1+\mu)}{4\xi^2(\mu-1)^5}\beta_7\right]. \quad (36)$$

Using Eq. (C2), the above expression for von Neumann entropy can be approximated to first order in coupling constant $\lambda$ as,

$$S_{VN} = -\frac{C_1\ln C_1 + (1-C_1)\ln(1-C_1)}{(1-C_1)} + \lambda\left[-\frac{C_1C_2\ln C_1}{(1-C_1)^2} - \frac{C_2\ln C_1}{(1-C_1)} + \frac{3C_1\ln C_1}{C_3^2(C_1+1)(C_1-1)^5}\beta_5 + \frac{\ln C_1(1+C_1+C_1^2)}{C_3^2(C_1+1)(C_1-1)^5}\beta_6 + \frac{3\ln C_1(1+C_1)}{4C_3^2(C_1-1)^5}\beta_7\right]. \quad (37)$$

The Renyi entropy of order $n$ can be evaluated using,

$$S_R = \frac{1}{1-n}\ln[tr(\rho^n)]. \quad (38)$$

Substituting the respective values of matrix inverses, Eq. (C13) in Eq. (34) while using Eq. (C2) one can show that the Renyi entropy is given by,

$$S_R = \frac{1}{1-n}\left\{n\ln(1-C_2-\lambda C_3) - \ln(1-(C_2+\lambda)^n) + n\lambda\left[\left(-\frac{3}{4C_3^2(1-C_1)^4} + \frac{(C_1^{2n}-1)^2}{4C_3^2(1-C_1^n)^4(C_1^2-1)^2}\right)\beta_5 + \left(-\frac{3}{4C_3^2(1-C_1)^4} + \frac{(C_1^{2n}-1)^2}{4C_3^2(1-C_1^n)^4(C_1^2-1)^2} + 2\frac{(C_1^n+C_1^2)^2}{4C_1^2C_3^2(C_1^2-1)^2(C_1^n-1)^2}\right)\beta_6 + \left(-\frac{3}{4C_3^2(1-C_1)^4} + \frac{3(C_1^{2n}-1)(C_1^n+C_1^2}{4C_1C_3^2(1-C_1^n)^2(C_1^2-1)^2(C_1^n-1)}\right)\beta_7\right]\right\}. \quad (39)$$

Note that the coefficients $C_i$ for $i = 1, 2, 3, 4$ in Eq. (37) and Eq. (39) arise due to the analytical steps shown in Appendix C. These coefficients are tabulated in Appendix D.

Using the values of von Neumann entropy, Eq. (37) and Renyi entropy, Eq. (39) one can verify, using first order approximation in coupling constant $\lambda$,

$$\lim_{n\to 1} S_R = S_{VN}.$$

Note that the final formulas of von Neumann as well as Renyi entropies depend on $\beta_i$ for $i = 1, 2\ldots 7$. These coefficients $\beta_i$, given in the table of IV B depend on timescale $t$ and $\delta t$. Entanglement measures therefore depend on these timescales. We check this time-dependence by computing numerical values of both entanglement measures in section V.

## V. NUMERICAL RESULTS

In this section we numerically evaluate von Neumann and Renyi entanglement entropy measures computed for the quench setup of two coupled oscillators using Eq. (37) and Eq. (39). As mentioned before each factor in the derived formulas for entanglement measures explicitly depends on $\sigma_i(t,\delta t)$ and $\gamma_i(t,\delta t)$. The values of $\sigma_i(t,\delta t)$ and $\gamma_i(t,\delta t)$ are computed by solving auxiliary equations, shown in, Eq. (12) which is outlined in Appendix A. Analytically solving the differential equation of Eq. (A1), is very complicated and hence we set some initial conditions to numerically evaluate the solution to this equation.

We begin by considering the coupling coefficient in the Hamiltonian of the coupled oscillators, Eq. (5) as $\eta = 0.5$. We further set the invariant quantities in Eq. (14) as, $\Omega_1 = \Omega_2 = 1$. To obtain the constants $A$, $B$, and $C$ mentioned in Eq. (A4) we first compute $\sigma_i(t,\delta t)$ and $\gamma_i(t,\delta t)$ at $t \to 0$. Next we, set $d(t \to 0) = f(t \to 0) = 0$, defined in, Eq. (16) and $\sigma_i(0,\delta t) = 1$.

Using these initial conditions we obtain values of $A$, $B$, and $C$ which can be inserted in Eq. (A4). $\rho_i(t,\delta t)$ and $\gamma_i(t,\delta t)$ are then used to get numerical values of von Neumann and Renyi entanglement entropies, for the aforementioned initial conditions. Using the numerical values of $S_{VN}$ and $S_R$, we parametrize four different plots for a chosen timescale. We have varied the dimensionless parameter $(t/\delta t)$ from 0.1 to 0.7 in steps of 0.05. The ratio is then plotted on the x-axis of the respective figures. We term the value of $(t/\delta t) = 1$ as the "quench point" represented by a red dotted line in all the respective figures. Using the values of $(t/\delta t)$ we divide all the plots in three different regions. The first region shaded as red, is marked for values of $(t/\delta t) < 0.8$. This region shows the "early-time behavior" of the respective entanglement measures, when the quench rate $\delta t$ is varied in a way so as to keep $(t/\delta t) \ll 1$. The next region, shaded as yellow is marked by two equal intervals to the right as well as left of quench point, precisely for values of $(t/\delta t)$ between 0.8 and 1.2. This region represents the values of entanglement measures for $(t/\delta t) \approx 1$ and hence is termed as the region "around the quench point." The last region shaded as blue is marked for values of $(t/\delta t) > 1.2$. This region shows the "late-time behavior" of the respective entanglement measures, when the quench rate $\delta t$ is varied in a way so as to keep $(t/\delta t) \gg 1$.

In Fig. 1 we have plotted the von Neumann (V-N) entropy for two coupled oscillators having quartic self-coupling with respect to the dimensionless parameter $(t/\delta t)$ for different orders of $\lambda$. We observe that the computed values of V-N entropy are negative for the chosen timescale for $\lambda > 10^{-5}$. We begin by plotting the values of V-N entropy by decreasing the order of $\lambda$, starting from

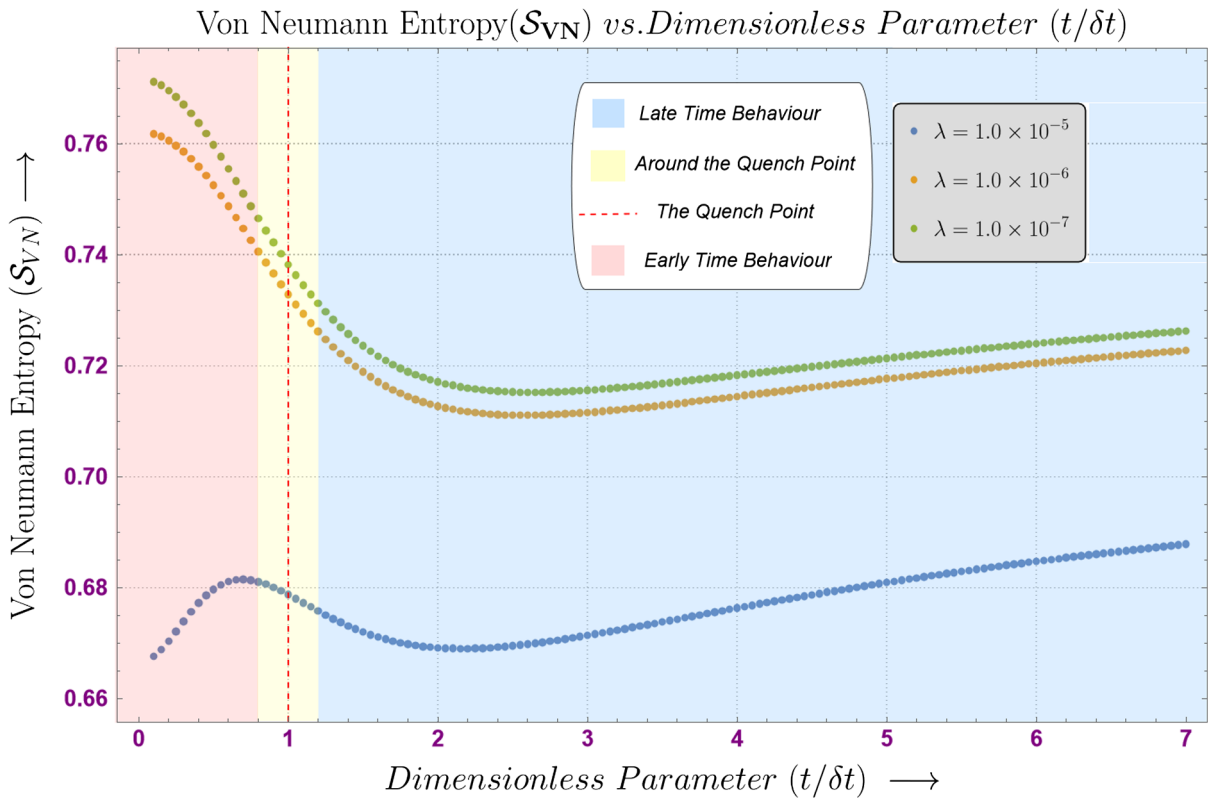



FIG. 1. Variation of the Von Neumann entropy ($S_{VN}$) with respect to the dimensionless parameter $(t/\delta t)$ for different orders of the coupling constant $\lambda$ for two coupled oscillators with quartic perturbation.

$\lambda = 10^{-5}$. We see that for $\lambda = 10^{-5}$ initially the V-N entropy grows for very small values of $(t/\delta t) < 0.6$. It can therefore be inferred from the plot that V-N entropy (for $\lambda = 10^{-5}$) increases in most of the region covering the early-time behavior, shaded as blue. Further, in the range $0.6 < (t/\delta t) < 2.2$ the entropy decreases gradually. Thus, the whole region near to the quench-point, shaded as yellow, shows a decreasing value of V-N entropy. After $(t/\delta t) > 2.2$ the V-N entropy increases monotonically and shows a thermalizing behavior for large values of $(t/\delta t)$. Most of the late-time behavior therefore, shows thermalizing behavior of V-N entropy.

In case of $\lambda = 10^{-6}$ and $\lambda = 10^{-7}$ we observe from the plot in Fig. 1, that there is a decrease in von Neumann entropy up to $(t/\delta t) < 2.6$. Hence, the whole early-time behavior region as well as the region around the quench point, shaded as red and yellow respectively, show a decreasing trend in V-N entropy. This trend is in contrast to the same for $\lambda = 10^{-5}$. When we move further toward larger values of the dimensionless parameter $(t/\delta t)$ we again see a thermalizing behavior for both the coupling constants. This region is shaded as blue and shows trend similar to that of $\lambda = 10^{-5}$. Another observation which we can make from this graph is that as we decrease the order of the coupling constant $\lambda$ the von Neumann entropy increases.

In Fig. 2 we have plotted the Renyi entropies for two coupled oscillators having quartic self-coupling with respect to the dimensionless parameter $(t/\delta t)$ for different orders of $\lambda$. We observe that the computed values of Renyi entropy are negative for the chosen timescale for $\lambda > 10^{-4}$. We begin by plotting the values of Renyi entropy by decreasing the order of coupling constant, starting from $\lambda = 10^{-4}$. We see that for $\lambda = 10^{-4}$, initially the Renyi entropy grows for very small value of $(t/\delta t) < 0.8$. Hence, the early time-behavior of the system shows an increasing trend in values of Renyi entropy, shaded by red color. Further, in the range $0.8 < (t/\delta t) < 2.2$ the entropy decreases gradually. The region around the quench point, shaded as yellow, shows decreasing trend in Renyi entropy. After $(t/\delta t) > 2.2$ the Renyi entropy increases monotonically and shows a thermalizing behavior for large values of $(t/\delta t)$. Most of the late-time behavior of the system shows the thermalization trend in Renyi entropy. This region is shaded by blue color. This behavior is similar to that of V-N entropy for $\lambda = 10^{-5}$ shown in Fig. 1.

In case of $\lambda = 10^{-5}$ and $\lambda = 10^{-6}$ we observe from the graph that there is a decrease in Renyi entropy up to $(t/\delta t) < 2.6$. Hence, the early-time behavior as well as behavior of the system around the quench point results in decreasing values of Renyi entropy, shaded as red and yellow respectively in the Fig. 2. This trend is in contrast to

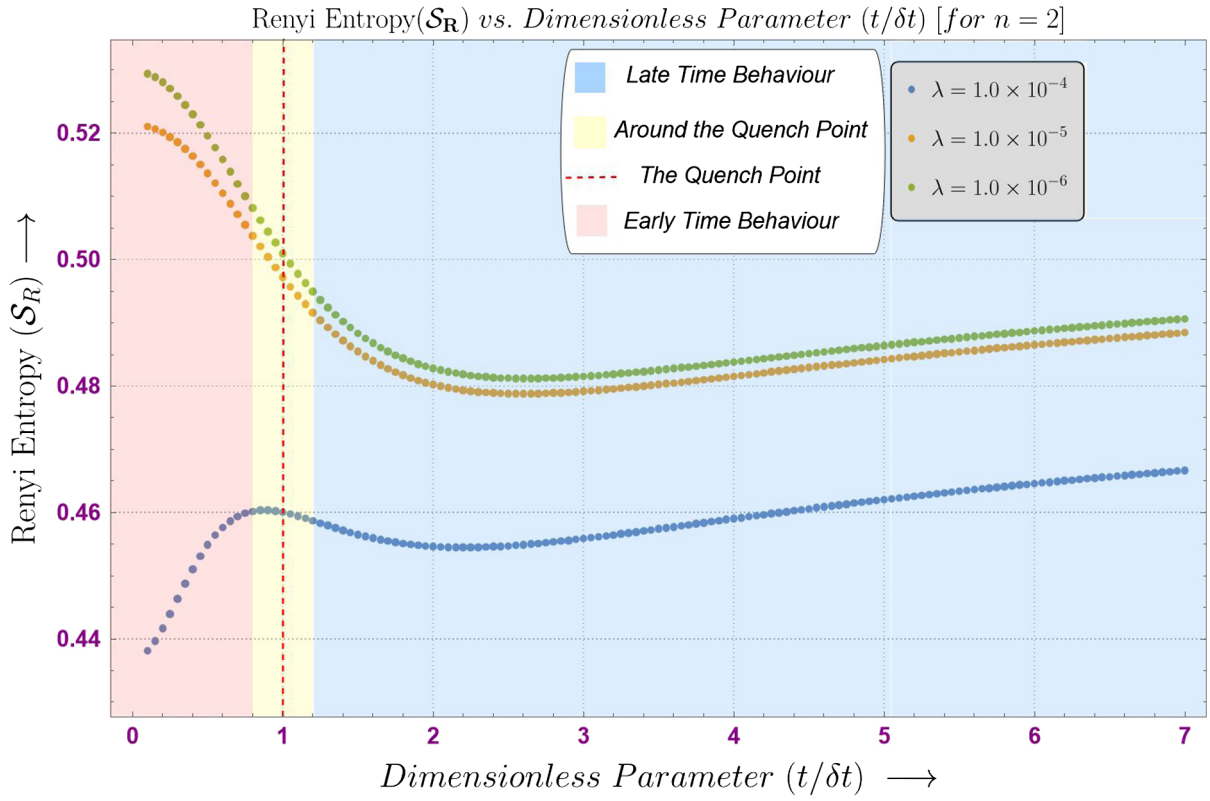



FIG. 2. Variation of the Renyi entropy ($S_R$) for $n = 2$, with respect to the dimensionless parameter $(t/\delta t)$ for different orders of the coupling constant $\lambda$ for two coupled oscillators with quartic perturbation.

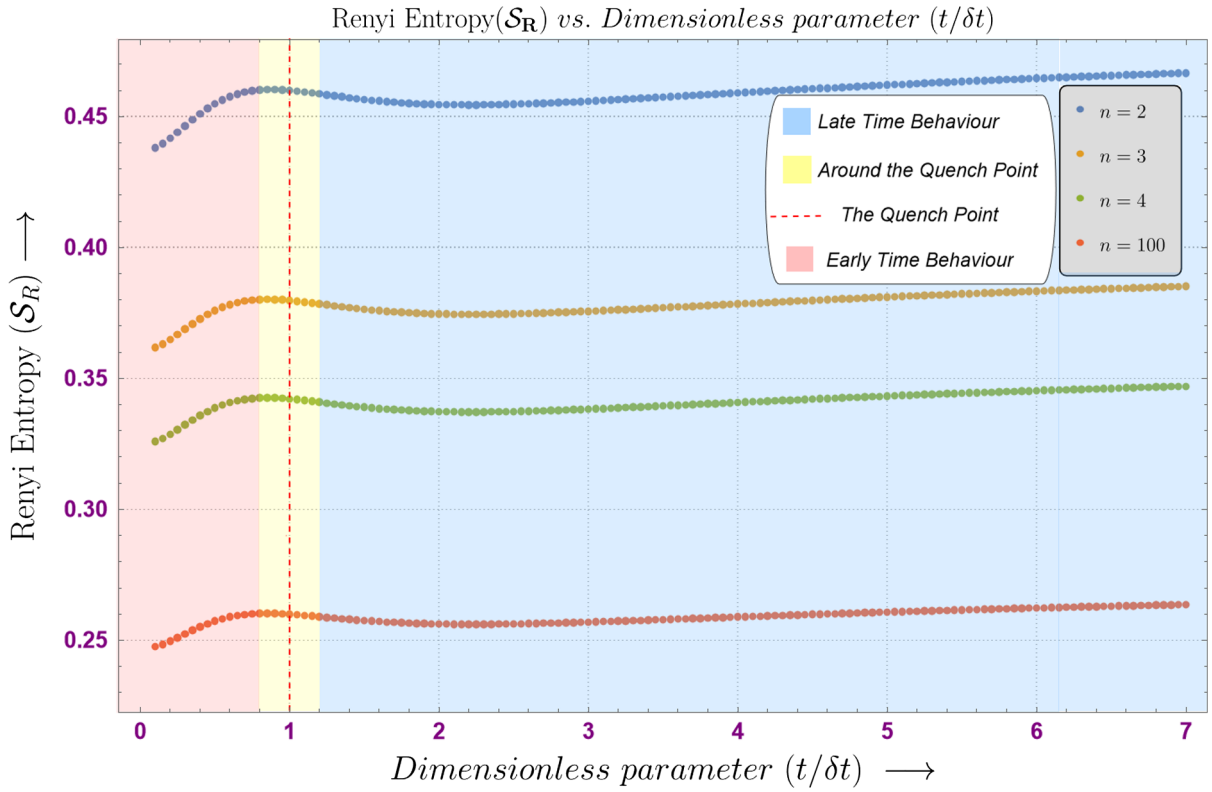



FIG. 3. Variation of the Renyi entropy $(S_R)$ for $\lambda = 10^{-4}$, with respect to the dimensionless parameter $(t/\delta t)$ for different values of $n$ for two coupled oscillators with quartic perturbation.

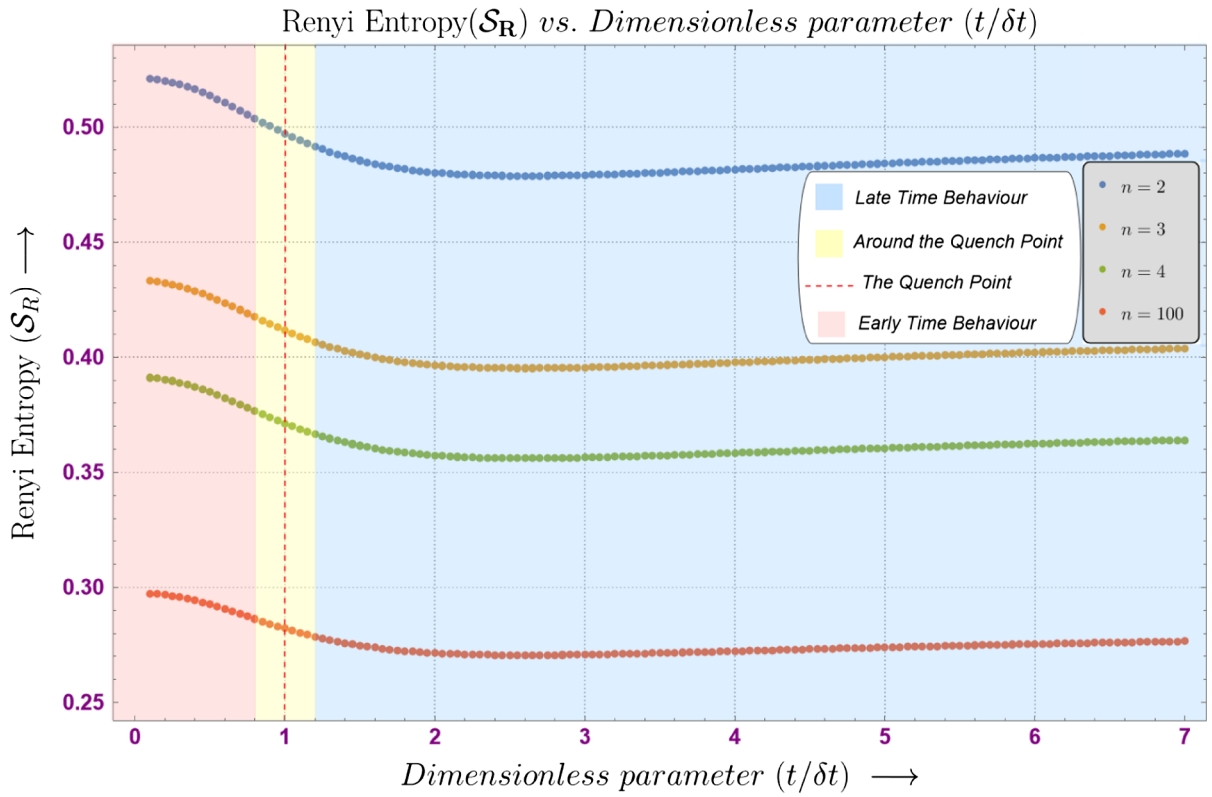



FIG. 4. Variation of the Renyi entropy $(S_R)$ for $\lambda = 10^{-5}$, with respect to the dimensionless parameter $(t/\delta t)$ for different values of $n$ for two coupled oscillators with quartic perturbation.

that of $\lambda = 10^{-4}$. When we move further toward larger values of the dimensionless parameter $(t/\delta t)$ we again see a thermalizing behavior for both the coupling constants. The late-time behavior of the system is, mostly characterized by thermalization of Renyi entropy. This is shown by blue region in Fig. 2. This behavior is similar to that of V-N entropy for $\lambda = 10^{-6}, 10^{-7}$ in Fig. 1. Another observation which we can make from this graph is that as we decrease the order of the coupling constant $\lambda$ the Renyi entropy increases.

Hence, as we are decreasing the order of $\lambda$ the plots of von Neumann entropy and that of Renyi entropy show a similar behavior with respect to each other, given that the order of $\lambda$ in the case of von Neumann entropy is one lower than that in Renyi entropy.

In the Fig. 3 we have plotted the Renyi entropies for two coupled oscillators having quartic self-coupling with respect to the dimensionless parameter $(t/\delta t)$ for different orders of Renyi entropy, i.e., for different values of $n$, set at $\lambda = 10^{-4}$. The early time behavior shows an increasing trend in the value of Renyi entropy for the chosen values of $n = 2, 3, 4$. The entropy then decreases covering the region around the quench. Most of the late time behavior of the system shows thermalizing nature of Renyi entropy. It is clear that this scaling behavior is retained for large value of $n = 100$.

Figure 4 shows parametric variation for different orders of Renyi entropy, i.e., for different values of $n$, set at $\lambda = 10^{-5}$. The early time behavior as well as the behavior of system near the quench point, shows a decreasing trend in values of Renyi entropy. Most of the late-time behavior is characterized again by thermalization of Renyi entropy for chosen values of $n = 2, 3, 4$. Again, this scaling behavior is retained for large value of $n = 100$.

## VI. CONCLUSION

The concluding remarks of this work are appended below pointwise:

(i) Focusing on a system of two coupled oscillators with quartic perturbation, we have derived analytical expressions of von Neumann entanglement entropy and Renyi entropy, undergoing a quantum quench.

(ii) First we have computed the expression for eigenstates of unperturbed Hamiltonian using invariant operator method. Using this expression we have approximated the first order $\lambda\phi^4$ correction for the total Hamiltonian of the system. Since the Hamiltonian is time-dependent due to the chosen quench profile as the frequency of the oscillators, it is quite evident that the total Hamiltonian can be quantized by using solutions to the Ermakov-Milne-Pinney equation. The ground state of the total Hamiltonian of the system, having $\phi^4$ interaction term, is then used to derive analytical expressions for the respective entanglement measure.

(iii) Next we have mentioned the reduced density matrix for the ground state of the above described system of coupled oscillators. This reduced density matrix, clearly represents non-Gaussian state due to presence of quartic interaction terms. We deal with this non-Gaussian terms by constructing a quartic tensor and computing the trace of $n$th order of reduced density matrix.

(iv) Finally, we employ the use of replica trick for computing von Neumann entanglement entropy. Further, Renyi entropy was computed using the standard formula, depending on the reduced density matrix. The analytical expression for these entanglement measures is time-dependent as all the coefficients in the respective expression depend on solutions of Ermakov-Milne-Pinney equation.

(v) Using the numerically evaluated values of von Neumann entropy and Renyi entropy we studied the variation of these entanglement measures with respect to the dimensionless parameter $(t/\delta t)$ specifying three regions: early-time behavior, the behavior around the quench point and the late-time behavior.

(vi) From these numerical results, we find that both von Neumann entropy and Renyi entropy delicately depend on the order of coupling constant $\lambda$. Evidently there exists a respective threshold order of $\lambda$ beyond which $\lambda$ if increased, we do not get positive values of both von Neumann and Renyi entanglement entropies, for chosen values of $(t/\delta t)$. For the respective threshold order of $\lambda$ we observe same scaling behavior in both von Neumann entropy and Renyi entropy. This scaling behavior can be characterized by a trend of increasing values of the entanglement measure for early times while in the region around the quench point the behavior shows a decreasing trend in these values.

(vii) As the order of $\lambda$ is decreased below the respective threshold order we get another scaling behavior of both von Neumann and Renyi entropies. This scaling behavior can be characterized by a trend of decreasing values of the respective entanglement measure for both early-times and around the quench point regions. Both scaling behaviors show thermalizing behavior of the respective entanglement measures at very late times.

(viii) It is quite clear from the plots that as we decrease the order of $\lambda$ the value of both entanglement measures increases.

(ix) Also, for a given order of coupling constant $\lambda$ von Neumann entropy thermalizes at higher values compared to that of Renyi entropy.

(x) Next we find that for a particular order of $\lambda$ we obtain same scaling behavior for different orders of the Renyi entropy. However, the value of Renyi entropy decreases as we increase the order of Renyi entropy.The particular scaling behavior is retained

even for the case of very high order of Renyi entropy.

*Future prospects:*

(i) In the present article, we have analyzed the effects of quantum quench on the entanglement entropy for a system consisting of two coupled oscillators with quartic perturbation. Of course, this study of entanglement entropy and quantum quench can be generalized to that of $N$-coupled oscillators. For $N \to \infty$, in the continuous limit, it would be interesting to explore the effects of quantum quench on entanglement in the context of interacting field theory.

(ii) One of the latest developments in research in the field of high energy physics is, the study of circuit complexity [81–87]. There are some works which are focused on relating the complexity with quantum entanglement [79,88–91]. The study of the same might turn out to be intriguing in the case of interacting quenched field theories.

(iii) Hence, it would be interesting to explore the connection between quantum circuit complexity and entanglement and check its consistency with the CA and CV [92–95] proposal.

## ACKNOWLEDGMENTS

The Visiting Post Doctoral research fellowship of S. C. is supported by the J. C. Bose National Fellowship of Director, Professor Rajesh Gopakumar, ICTS, TIFR, Bengaluru. S. C. also would like to thank ICTS, TIFR, Bengaluru for providing the work friendly environment. S. C. also thanks all the members of our newly formed virtual international non-profit consortium Quantum Aspects of the Space-Time & Matter (QASTM) for elaborative discussions. R. M. G., S. M., N. P., A. R., and P. S. would like to thank the members of the QASTM Forum for useful discussions.

## APPENDIX A: COMPUTING EXPLICIT NUMERICAL VALUES OF $\sigma_i(T)$ AND $\gamma_i(T)$

Using auxiliary conditions given by Eq. (12) we briefly show the steps to compute $\sigma_1$ and $\gamma_1$. We begin by rearranging Eq. (12) for one of the oscillators,

$$\ddot{\sigma_1} + \omega_1^2(T)\sigma_1 = \frac{\Omega_1}{\sigma_1^3}. \tag{A1}$$

The above second order differential equation is often termed as, Ermakov-Milne–Pinney equation [24–26]. This equation can be solved numerically to obtain $\sigma_1(T)$. Since $T = (t/\delta t)$, the solution will clearly be function of both $t$ and $\delta t$. We assume that the form of solution of the above equation gives us a linear combination,

$$\sigma_1(t,\delta t) = c_1 z_1(t,\delta t) + c_2 z_2(t,\delta t). \tag{A2}$$

Here, $c_1$ and $c_2$ are numerical constants, while $z_1$ and $z_2$ are treated as two complex-valued solutions of Eq. (A1). We will consider only $z_1$ as one of the solutions. Using the form of quench profile Eq. (9), the computed value of one of the solutions is,

$$z_1(t,\delta t) = [e^{\frac{2t}{\delta t}}]^{-\frac{1}{2}(i\delta t)}[e^{\frac{2t}{\delta t}}+1]^{\frac{1}{2}(\sqrt{1-4\delta t^2}+1)}\,{}_2F_1\left[\frac{1}{2}(\sqrt{1-4\delta t^2}+1), \frac{1}{2}(-2i\delta t+\sqrt{1-4\delta t^2}+1); 1-i\delta t; -e^{\frac{2t}{\delta t}}\right]. \tag{A3}$$

Here, ${}_2F_1$ represents the hypergeometric function. Since $z_1$ is complex valued we can write, $z_1 = y_1 + iy_2$ such that $y_1$ and $y_2$ are now two real-linearly independent equations. We give an outline of steps shown in [96] for numerical solution of Eq. (A1) using these linearly independent equations. This solution is guaranteed to be of the form,

$$\sigma_1(t,\delta t) = [\sqrt{Ay_1^2(t,\delta t)t + 2By_1(t,\delta t)y_2(t,\delta t) + Cy_2^2(t,\delta t)}]. \tag{A4}$$

The next step is to determine the constants in Eq. (A4). These are fixed by setting the condition $AC - B^2 = \Omega_1^2$. These steps give us the explicit value of $\sigma_1(t,\delta t)$. One can repeat these steps by inserting the respective parameters (of second oscillator) to find $\sigma_2(t,\delta t)$. Since, $\Omega_i = \sigma_i^2\dot{\gamma}_i$, the explicit value of $\gamma_i$ is computed by using value of Eq. (A4),

$$\gamma_i(t,\delta t) = \int_0^t \frac{\Omega_i}{\sigma_i^2(t,\delta t)}\,dt \tag{A5}$$

Inserting the values of $\sigma_i(t,\delta t)$ and $\gamma_i(t,\delta t)$ it is clear that the wave function, in Eq. (23) becomes a function of both $t$ and $\delta t$. Note that we conceal this functional dependence in all the sections until the entropy is numerically evaluated in Sec. V.

## APPENDIX B: AN OUTLINE OF INVARIANT OPERATOR REPRESENTATION

In Sec. III we defined $I_j$ as an operator in Eq. (14). We outline in this appendix a few important steps for constructing this operator and the way to find eigenstates of this operator. Note that the subscript $j = 1, 2$ represent the parameters described for the oscillators having spatial coordinates: $X_1$ and $X_2$ respectively.

The operator $I_j$ is constructed such that it satisfies [18],

$$\frac{\partial I_j}{\partial T} + \frac{1}{i}[I_j, H_j], \tag{B1}$$

where $H_j$ for $j = 1, 2$ represents the respective decoupled Hamiltonian for each oscillator (7). One can show that the operator given in Eq. (14) satisfies the above condition.

It is assumed that the invariant $I_j$ is one of a complete set of commuting observables for respective $H_j$. This guarantees that there is a complete set of eigenstates for each $I_j$ defined in Eq. (14). We refer to $u_{0_j}$ for $j = 1, 2$ as the ground state for the spectrum of the respective invariant operator. These ground states of the respective invariant operators can be determined using the condition, $a_j u_{0_j} = 0$ where $a_j$ is the respective annihilation operator Eq. (13). When evaluated, the expression for ground state of the invariant operator $I_j$ is given by,

$$u_{0_j} = \left(\frac{\dot{\gamma}_j}{\pi}\right)^{1/4} \exp\left[-\frac{\dot{\gamma}_j}{2}(-i\dot{\sigma}_j\sigma_j\dot{\gamma}_j)X_j^2\right]. \tag{B2}$$

Using the ground states and the respective creation operators $a_j^\dagger$ one can then show that, the $n$th eigenstate of the invariant-operator $I_j$ is given by,

$$u_{n_j} = \frac{1}{\sqrt{n!}}(a_j^\dagger)^n u_{0_j} = \left(\frac{1}{2^{n_j} n_j!}\right)\left(\frac{\dot{\gamma}_j}{\pi}\right)^{1/4} \times \exp\left[\dot{\gamma}_j\left(1 - \frac{i\dot{\sigma}_j}{\sigma_j\dot{\gamma}_j}\right)X_j^2\right]\mathbf{H}_{n_j}\left[\sqrt{\dot{\gamma}_j}X_j\right]. \tag{B3}$$

Here, $j = 1, 2$ and $\mathbf{H}_{n_j}$ represents the Hermite polynomial of order $n_j$. Using the eigenstates of invariant operator (B3), one can compute the wave functions of the decoupled Hamiltonians [80]. It can be shown that the computed wave functions take the form: $\psi_{n_j} = e^{i\alpha_{n_j}} u_{n_j}$, as solutions to Schrodinger's equations for respective $H_j$, where $\alpha_{n_j} = -(1/2 + n_j)$; for $j = 1, 2$. The eigenstates for unperturbed Hamiltonian for the coupled oscillator system can further be computed as $\psi^{(0)}_{n_1,n_2} = \psi_{n_1} \times \psi_{n_2}$.

Using Eq. (B3) one can then show that,

$$\psi^{(0)}_{n_1,n_2} = \sqrt{\frac{\dot{\gamma_1}^2\dot{\gamma_1}^2}{2^{n_1+n_2}n_1!n_2!\pi}}\exp\left[-i\frac{(2n_1+1)\gamma_1 + (2n_2+1)\gamma_2}{2}\right]\exp\left[-\frac{1}{2}\dot{\gamma}_1\left(-\frac{i\dot{\sigma_1}}{\dot{\gamma}_1\sigma_1}\right)X_1^2 - \frac{1}{2}\dot{\gamma}_2\left(-\frac{i\dot{\sigma_2}}{\dot{\gamma}_2\sigma_2}\right)X_2^2\right] \times \mathbf{H}_{n_1}[\sqrt{\dot{\gamma_1}}X_1]\mathbf{H}_{n_2}[\sqrt{\dot{\gamma_2}}X_2]. \tag{B4}$$

This equation (B4) represents the eigenstates for the unperturbed Hamiltonian of two coupled oscillators having a quenched frequency profile. In III B we compute the first order time-independent correction to the ground state of above equation.

## APPENDIX C: COMPUTING INTEGRALS IN $tr(\rho^n)$

In the integral of Eq. (32) we have separated both Gaussian and non-Guassian parts. In this appendix we give detailed outline of solving both Gaussian and non-Gaussian contributions and finally combine them to compute the integral in (32).

The Gaussian part of the integrand can be parametrized by considering a quadratic coefficient matrix similar to the case in [35]. Using the values of $\beta_i$ for $i = 1$ to $i = 7$ defined in Sec. IV B, this coefficient matrix is defined as:

$$M_{ij} = -2(\beta_1 + \lambda\beta_2)\delta_{ij} - (\beta_3 + \lambda\beta_4)(\delta^i_{j+1} + \delta^{i+1}_j). \tag{C1}$$

We further modify the above defined matrix by introducing two new symbols $\mu$ and $\xi$. We choose these variables so that they satisfy,

$$\xi(1+\mu^2) = -(\beta_1 + \lambda\beta_2)$$
$$2\xi\mu = \beta_3 + \lambda\beta_4. \tag{C2}$$

We consider the following explicit solution of these equations, approximated to first order in $\lambda$:

$$\mu = C_1 + \lambda C_2$$
$$\xi = C_3 + \lambda C_4. \tag{C3}$$

The values of newly defined coefficients $C_1, C_2 \ldots C_4$ are tabulated in Appendix D. Using matrix $M_{ij}$ defined in Eq. (C1), one can recover the Gaussian part of Eq. (32) as shown below,

$$\exp\left[-\frac{1}{2}x^i M_{ij} x_j\right] = \exp\left[(\beta_1 + \lambda\beta_2)\sum_{i=1}^n x_i^2 + (\beta_3 + \lambda\beta_4)\sum_{i=1}^n x_i x_{i+1}\right]. \tag{C4}$$

Moving on to the non-Guassian part in Eq. (32), we further define a quartic tensor as:

$$T_{ijkl} = \lambda[\beta_5\delta_{ij}\delta_{jk}\delta_{kl} + (\alpha_7\delta_l^{k-1} + \alpha_6\delta_l^{k+1})\delta_{ij}\delta_{jk} + \beta_6\delta_{ij}\delta_k^{j+1}\delta_{kl}]. \tag{C5}$$

Using Eq. (C5), one can recover the remaining part of Eq. (32) as shown below,

$$\exp[x^i x^j x^k x^l T_{ijkl}] = \exp\left\{\lambda\left[\beta_5\sum_{i=1}^{n} x_i^4 + (\alpha_6 x_{i-1}^3 + \alpha_7 x_{i+1}^3)\sum_{i=1}^{n} x_i + \beta_6\sum_{i=1}^{n} x_i^2\sum_{i=1}^{n} x_{i+1}^2\right]\right\}. \tag{C6}$$

Using the expression shown in Eq. (C4) and the non-Gaussian contribution from Eq. (C6), one can parametrize the integrand in Eq. (32), as shown below:

$$tr(\rho^n) = N^n\left(\frac{g_1 g_2}{\pi}\right)^n\left(\sqrt{\frac{\pi}{A}}\right)^n e^{n(2\lambda A_1+\lambda\alpha_0)}\int d^n x \exp\left[-\frac{1}{2}x^i M_{ij}x_j\right]\sum_{p=0}^{\infty}\frac{1}{i!}(x^i x^j x^k x^l T_{ijkl})^p. \tag{C7}$$

Further, we define a Gaussian partition function, $Z_0$, given by,

$$Z_0 = \int d^n x \exp\left[-\frac{1}{2}x^i M_{ij}x_j\right] = \sqrt{\frac{(2\pi)^n}{\det M}} = \frac{\left(\sqrt{\frac{\pi}{\xi}}\right)^n}{|1-\mu^n|}, \tag{C8}$$

where det $M$ denotes the determinant of matrix $M_{ij}$ of Eq. (C1). Using the above partition function the summed over tensor in Eq. (C7) can be transformed to a correlator as shown below retaining the form of perturbative expansion,

$$tr(\rho^n) = N^n\left(\frac{g_1 g_2}{\pi}\right)^n\left(\sqrt{\frac{\pi}{A}}\right)^n e^{n(2\lambda A_1+\lambda\alpha_0)}Z_0\sum_{p=0}^{\infty}\frac{1}{p!}\langle\langle x^{i_1}x^{j_1}x^{k_1}x^{l_1}\ldots x^{i_p}x^{j_p}x^{k_p}x^{l_p}\rangle\rangle T_{i_1 j_1 k_1 l_!}\ldots T_{i_p j_p k_p l_p}. \tag{C9}$$

We simplify the above expression in Eq. (C9) using a generating functional $J$, as shown below,

$$Z(J) = \frac{1}{Z_0}\int d^n x \exp\left[-\frac{1}{2}x^i M_{ij}x_j + J_i x_i\right] = \exp\left[\frac{1}{2}J_i M_{ij}^{-1}J_j\right]. \tag{C10}$$

As shown in [35], correlator of Eq. (C9) computed using the above Eq. (C10) then becomes,

$$\langle\langle x_{i_1}\ldots xi_{2m}\rangle\rangle = \frac{\delta^{2m}}{\delta J_{i_1}\ldots\delta J_{i_{2m}}}Z(J)\bigg|_{J=0} = \frac{1}{2^m m!}\sum_{\sigma\in S_G}(M^{-1})_{i_{\sigma(1)\sigma(2)}}\ldots(M^{-1})_{i_{\sigma(2n-1)\sigma(2n)}}. \tag{C11}$$

Here $G$ is the quotient group which can be defined to reduce the sum significantly. Note that for a $4m$ point correlator function the chosen quotient group gives rise to three different permutations. A more detailed discussion about finding the quotient group $G$ can be found in [35].

Using the value of quartic tensor from Eq. (C4) and the correlator from Eq. (C11), one can simplify Eq. (C9) as,

$$tr(\rho^n) = N^n Z_0\left(\frac{g_1 g_2}{\pi}\right)^n\left(\sqrt{\frac{\pi}{A}}\right)^n e^{n(2\lambda A_1+\lambda\alpha_0)}\{1 + n\lambda[(3\beta_5+\beta_6)(M_{11}^{-1})^2 + 3\beta_7 M_{11}^{-1}M_{12}^{-1} + 2\beta_6(M_{12}^{-1})^2]\}. \tag{C12}$$

One can check that the matrix inverses are [35],

$$M_{11}^{-1} = \frac{(\mu^{2n}-1)}{2(1-\mu^n)^2\xi(\mu^2-1)}; \qquad M_{12}^{-1} = \frac{(\mu^n+\mu^2)}{2\mu\xi(\mu^2-1)(\mu^n-1)}. \tag{C13}$$

## APPENDIX D: TABULATED VALUES OF COEFFICIENTS

In this appendix, the values of various coefficients we have used in some steps to compute the analytical expression of entanglement measures, are tabulated in respective tables.

(i) We begin by listing the values of $A_i$ for $i = 1, 2 \ldots 6$ in Eq. (23) of Sec. III B in the table given below.

| $A_i$ | Coefficient of $A_i$ |
|---|---|
| $A_1$ | $\frac{3}{16}\left(-\frac{8\sigma_1^2\sigma_2^2}{\sigma_1^2\omega_1^2+\sigma_2^2\omega_2{}^2+\dot{\sigma}_1+\dot{\sigma}_2+\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}+\frac{3\sigma_1^4+4\sigma_2^2\sigma_1^2}{\sigma_1^2\omega_1^2+\dot{\sigma}_1+\frac{1}{\sigma_1^2}}+\frac{3\sigma_2^4+4\sigma_1^2\sigma_2^2}{\sigma_2^2\omega_2^2+\dot{\sigma}_2+\frac{1}{\sigma_2^2}}\right)$ |
| $A_2$ | $\frac{3}{4}\sigma_1^2\left(\frac{4\sigma_2^2}{\sigma_1^2\omega_1^2+\sigma_2^2\omega_2^2+\dot{\sigma}_1+\dot{\sigma}_2+\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}-\frac{\sigma_1^2+2\sigma_2^2}{\sigma_1^2\omega_1^2+\dot{\sigma}_1+\frac{1}{\sigma_2^2}}\right)$ |
| $A_3$ | $\text{or}\frac{3\sigma_1^2\sigma_2^2}{\sigma_1^2\omega_1^2+\sigma_2^2\omega_2^2+\dot{\sigma}_1+\dot{\sigma}_2+\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}-\frac{3(\sigma_2^4+2\sigma_1^2\sigma_2^2)}{4(\sigma_2^2\omega_2^2+\dot{\sigma}_2+\frac{1}{\sigma_2^2})}$ |
| $A_4$ | $-\frac{\sigma_1^4}{4(\sigma_1^2\omega_1^2+\dot{\sigma}_1+\frac{1}{\sigma_1^2})}$ |
| $A_5$ | $-\frac{\sigma_2^4}{4(\sigma_2^2\omega_2^2+\dot{\sigma}_2+\frac{1}{\sigma_2^2})};$ |
| $A_6$ | $-\frac{6\sigma_1^2\sigma_2^2}{\sigma_1^2\omega_1^2+\sigma_2^2\omega_2^2+\dot{\sigma}_1+\dot{\sigma}_2+\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}};$ |

(ii) Next, we tabulate the values of coefficients $\alpha_i$ for $i = 1, 2 \ldots 8$ in Eq. (28) of Sec. IV A.

| $\alpha_i$ | Coefficient of $\alpha_i$ |
|---|---|
| $\alpha_0$ | $\frac{(A_2+A_3)}{2A}+\frac{3(A_4+A_5+A_6)}{8A^2}$ |
| $\alpha_1$ | $\frac{(A_2+A_3)}{2}+\frac{(3A_4+3A_5-A_6)}{4A}+\frac{(A_2-A_3)(Q-P)}{2A}+\frac{(A_2+A_3)(Q-P)^2}{4A^2}+\frac{3(A_4-A_5)(Q-P)}{4A^2}+\frac{3(A_4+A_5+A_6)}{8A^3}(Q-P)^2$ |
| $\alpha_2$ | $\frac{(A_2+A_3)}{2}+\frac{(3A_4+3A_5-A_6)}{4A}+\frac{(A_2-A_3)(Q^*-P^*)}{2A}+\frac{(A_2+A_3)(Q^*-P^*)^2}{4A^2}+\frac{3(A_4-A_5)(Q^*-P^*)}{4A^2}+\frac{3(A_4+A_5+A_6)}{8A^3}(Q^*-P^*)^2$ |
| $\alpha_3$ | $\frac{(A_2-A_3)}{2A}(Q+Q^*-P-P^*)+\frac{(A_2+A_3)(Q-P)(Q^*-P^*)}{2A^2}+\frac{3(A_4-A_5)(Q+Q^*-P-P^*)}{4A_2}+\frac{3(A_4+A_5+A_6)(Q-P)(Q^*-P^*)}{4A^3}$ |
| $\alpha_4$ | $\frac{(A_4+A_5+A_6)}{4}+\frac{(A_4-A_5)(Q-P)}{2A}+\frac{(3A_4+A_5-A_6)(Q-P)^2}{8A^2}+\frac{(A_4-A_5)(Q-P)^3}{8A^3}+\frac{(A_4+A_5+A_6)(Q-P)^4}{32A^4}$ |
| $\alpha_5$ | $\frac{(A_4+A_5+A_6)}{4}+\frac{(A_4-A_5)(Q^*-P^*)}{2A}+\frac{(3A_4+A_5-A_6)(Q^*-P^*)^2}{8A^2}+\frac{(A_4-A_5)(Q^*-P^*)^3}{8A^3}+\frac{(A_4+A_5+A_6)(Q^*-P^*)^4}{32A^4}$ |
| $\alpha_6$ | $\frac{(A_4-A_5)(Q^*-P^*)}{2A}+\frac{(3A_4+3A_5-A_6)(Q-P)(Q^*-p^*)}{4A^2}+\frac{(A_4-A_5)[3(Q-P)^2(Q^*-P^*)+(Q-P)^3}{8A^3}+\frac{(A_4+A_5+A_6)(Q-P)^3(Q^*-P^*)}{8A^4}$ |
| $\alpha_7$ | $\frac{(A_4-A_5)(Q-P)}{2A}+\frac{(3A_4+3A_5-A_6)(Q-P)(Q^*-p^*)}{4A^2}+\frac{(A_4-A_5)[3(Q-P)(Q^*-P^*)^2+(Q^*-P^*)^3}{8A^3}+\frac{(A_4+A_5+A_6)(Q-P)(Q^*-P^*)^3}{8A^4}$ |
| $\alpha_8$ | $\frac{(3A_4+3A_5-A_6)[(Q^*-P^*)^2+(Q-P)^2]}{8A^2}+\frac{(A_4-A_5)[3(Q-P)^2(Q^*-P^*)+3(Q-P)(Q^*-P^*)^2]}{8A^3}+\frac{(A_4+A_5+A_6)6(Q-P)^2(Q^*-P^*)^2}{32A^4}$ |

(iii) The values of $\beta_i$ for $i = 1, 2 \ldots 7$ in Eq. (32) of Sec. IV B are tabulated in the below given table.

| $\beta_i$ | Coefficient of $\beta_i$ |
|---|---|
| $\beta_1$ | $-A+\frac{(g_2^2-g_1^2)^2-(dg_1^2-fg_2^2)^2}{8A}$ |
| $\beta_2$ | $A_2+A_3+\frac{3A_4+3A_5-A_6}{2A}+(\frac{A_2-A_3}{2A})(g_2^2-g_1^2)+\frac{3(A_4-A_5)(g_2^2-g_1^2)}{4A^2}+[\frac{3(A_4+A_5+A_6)}{16A^3}+\frac{A_2+A_3}{8A^2}][(g_2^2-g_1^2)^2-(dg_1^2-fg_2^2)^2]$ |
| $\beta_3$ | $\frac{(g_2^2-g_1^2)^2+(dg_1^2-fg_2^2)^2}{8A}$ |
| $\beta_4$ | $[\frac{A_2-A_3}{2A}+\frac{3(A_4-A_5)}{4A^2}][g_2^2-g_1^2]+[\frac{(A_2+A_3)}{8A^2}+\frac{3(A_4+A_5+A_6)}{16A^3}][(g_2^2-g_1^2)^2+(dg_1^2-fg_2^2)^2]$ |
| $\beta_5$ | $\frac{A_4+A_5+A_6}{2}+\frac{(A_4-A_5)(g_2^2-g_1^2)}{2A}+\frac{3A_4+3A_5-A_6}{16A^2}[(g_2^2-g_1^2)^2-(dg_1^2-fg_2^2)^2]+\frac{A_4-A_5}{64A^3}[2(g_2^2-g_1^2)^3-6(g_2^2-g_1^2)(dg_1^2-fg_2^2)^2]$ $+\frac{A_4+A_5+A_6}{512A^4}[2(g_2^2-g_1^2)^4+2(dg_1^2-fg_2^2)^4-12(g_2^2-g_1^2)^2(dg_1^2-fg_2^2)^2]$ |
| $\beta_6$ | $(\frac{3A_4+3A_5-A_6}{16A^2})[(g_2^2-g_1^2)^2-(dg_1^2-fg_2^2)^2]+(\frac{3(A_4-A_5)}{32A^3})[(g_2^2-g_1^2)^3+(g_2^2-g_1^2)(dg_1^2-fg_2^2)^2]+(\frac{6(A_4+A_5+A_6)}{512A^4})[(g_2^2-g_1^2)^4$ $+(dg_1^2-fg_2^2)^4+2(g_2^2-g_1^2)^2(dg_1^2-fg_2^2)^2]$ |
| $\beta_7$ | $\frac{(A_4-A_5)(g_2^2-g_1^2)}{2A}+(\frac{3A_4+3A_5-A_6}{8A^2})[(g_2^2-g_1^2)^2+(dg_1^2-fg_2^2)]+(\frac{A_4-A_5}{32A^3})[(g_2^2-g_1^2)^3]+(\frac{A_4+A_5+A_6}{64A^4})[(g_2^2-g_1^2)^4-(dg_1^2-fg_2^2)^4]$ |

(iv) The last table, given below contains values of $C_i$ for $i = 1, 2, 3, 4$ in Eq. (C3) of Appendix C.

| $C_i$ | Value of $C_i$ |
|---|---|
| $C_1$ | $-\frac{\beta_1}{\beta_3}+\frac{(\beta_1^2-\beta_3^2)^{\frac{1}{2}}}{\beta_3}$ |
| $C_2$ | $-\frac{\beta_2}{\beta_3}+\frac{(\beta_1\beta_2-\beta_3\beta_4)}{\beta_3(\beta_1^2-\beta_3^2)^{\frac{1}{2}}}+\frac{\beta_1\beta_4}{\beta_3^2}-\frac{\beta_4(\beta_1^2-\beta_3^2)^{\frac{1}{2}}}{\beta_3^2}$ |
| $C_3$ | $\frac{\beta_3}{2C_1}$ |
| $C_4$ | $\frac{\beta_4}{2C_1}-\frac{C_2\beta_3}{2C_1^2}$ |


[1] P. Calabrese and J. Cardy, Entanglement entropy and quantum field theory, J. Stat. Mech. (2004) P06002.

[2] E. Witten, Notes on some entanglement properties of quantum field theory, Rev. Mod. Phys. **90**, 045003 (2018).

[3] T. Nishioka, Entanglement entropy: Holography and renormalization group, Rev. Mod. Phys. **90**, 035007 (2018).

[4] D. Blanco, Quantum information measures and their applications in quantum field theory, arXiv:1702.07384.

[5] M. Headrick, Lectures on entanglement entropy in field theory and holography, arXiv:1907.08126.

[6] L. Amico, R. Fazio, A. Osterloh, and V. Vedral, Entanglement in many-body systems, Rev. Mod. Phys. **80**, 517 (2008).

[7] J. I. Cirac, Entanglement in many-body quantum systems, arXiv:1205.3742.

[8] A. Lakshminarayan and V. Subrahmanyam, Multipartite entanglement in a one-dimensional time-dependent Ising model, Phys. Rev. A **71**, 062334 (2005).

[9] H. F. Song, S. Rachel, C. Flindt, I. Klich, N. Laflorencie, and K. Le Hur, Bipartite fluctuations as a probe of many-body entanglement, Phys. Rev. B **85**, 035409 (2012).

[10] C. J. Bardeen, Time dependent correlations of entangled states with nondegenerate branches and possible experimental realization using singlet fission, J. Chem. Phys. **151**, 124503 (2019).

[11] A. Sivaramakrishnan, Entanglement entropy with a time-dependent Hamiltonian, Phys. Rev. D **97**, 066003 (2018).

[12] P. Caputa, G. Mandal, and R. Sinha, Dynamical entanglement entropy with angular momentum and U(1) charge, J. High Energy Phys. 11 (2013) 52.

[13] E. Canovi, E. Ercolessi, P. Naldesi, L. Taddia, and D. Vodola, Dynamics of entanglement entropy and entanglement spectrum crossing a quantum phase transition, Phys. Rev. B **89**, 104303 (2014).

[14] P. Jacquod and C. Petitjean, Decoherence, entanglement and irreversibility in quantum dynamical systems with few degrees of freedom, Adv. Phys. **58**, 67 (2009).

[15] S. Akhtar, S. Choudhury, S. Chowdhury, D. Goswami, S. Panda, and A. Swain, Open quantum entanglement: A study of two atomic system in static patch of de Sitter space, Eur. Phys. J. C **80**, 748 (2020).

[16] S. Choudhury and S. Panda, Quantum entanglement in de Sitter space from stringy axion: An analysis using $\alpha$ vacua, Nucl. Phys. **B943**, 114606 (2019).

[17] S. Choudhury and S. Panda, Entangled de Sitter from stringy axionic Bell pair I: An analysis using Bunch–Davies vacuum, Eur. Phys. J. C **78**, 52 (2018).

[18] H. R. Lewis and W. B. Riesenfeld, An exact quantum theory of the time-dependent harmonic oscillator and of a charged particle in a time-dependent electromagnetic field, J. Math. Phys. (N.Y.) **10**, 1458 (1969).

[19] K. Andrzejewski, Dynamics of entropy and information of time-dependent quantum systems: Exact results, Quantum Inf. Process. **21**, 117 (2022).

[20] B. Khantoul, A. Bounames, and M. Maamache, On the invariant method for the time-dependent non-Hermitian Hamiltonians, Eur. Phys. J. Plus **132**, 258 (2017).

[21] S. Choudhury, Cosmological geometric phase from pure quantum states: A study without/with having Bell's inequality violation, arXiv:2105.06254.

[22] J.-R. Choi, Coherent and squeezed states for light in homogeneous conducting linear media by an invariant operator method, Int. J. Theor. Phys. **43**, 2113 (2004).

[23] H. Kanasugi and H. Okada, Systematic treatment of general time-dependent harmonic oscillator in classical and quantum mechanics, Prog. Theor. Phys. **93**, 949 (1995).

[24] V. P. Ermakov, Second-order differential equations: Conditions of complete integrability, *Applicable Analysis and Discrete Mathematics* (University of Belgrade, Serbia, 2008), pp. 123–145.

[25] W. Milne, The numerical determination of characteristic numbers, Phys. Rev. **35**, 863 (1930).

[26] E. Pinney, The nonlinear differential equation $y'' + p(x)y + cy^{-3} = 0$, Proc. Am. Math. Soc. **1**, 681 (1950).

[27] J. R. Choi, Perturbation theory for time-dependent quantum systems involving complex potentials, Front. Phys. **8**, 189 (2020).

[28] W. Wang, S. C. Hou, and X. X. Yi, Adiabatic evolution under quantum control, Ann. Phys. (Amsterdam) **327**, 1293 (2012).
[29] M.-Y. Ye, X.-F. Zhou, Y.-S. Zhang, and G.-C. Guo, Two kinds of quantum adiabatic approximation, Phys. Lett. A **368**, 18 (2007).
[30] K. Audenaert, J. Eisert, M. B. Plenio, and R. F. Werner, Entanglement properties of the harmonic chain, Phys. Rev. A **66**, 042327 (2002).
[31] A. Jafarizadeh and M. A. Rajabpour, Bipartite entanglement entropy of the excited states of free fermions and harmonic oscillators, Phys. Rev. B **100**, 165135 (2019).
[32] T. J. Osborne and M. A. Nielsen, Entanglement in a simple quantum phase transition, Phys. Rev. A **66**, 032110 (2002).
[33] R. D. Sorkin, Expressing entropy globally in terms of (4D) field-correlations, J. Phys. Conf. Ser. **484**, 012004 (2014).
[34] C. Holzhey, F. Larsen, and F. Wilczek, Geometric and renormalized entropy in conformal field theory, Nucl. Phys. **B424**, 443 (1994).
[35] Y. Chen, L. Hackl, R. Kunjwal, H. Moradi, Y. K. Yazdi, and M. Zilhão, Towards spacetime entanglement entropy for interacting theories, J. High Energy Phys. 11 (2020) 114.
[36] V. Rosenhaus and M. Smolkin, Entanglement entropy: A perturbative calculation, J. High Energy Phys. 12 (2014) 179.
[37] P. Calabrese and J. Cardy, Quantum quenches in extended systems, J. Stat. Mech. (2007) P06008.
[38] P. Basu and S. R. Das, Quantum quench across a holographic critical point, J. High Energy Phys. 01 (2012) 103.
[39] A. Buchel, L. Lehner, R. C. Myers, and A. van Niekerk, Quantum quenches of holographic plasmas, J. High Energy Phys. 05 (2013) 067.
[40] S. R. Das, D. A. Galante, and R. C. Myers, Universal Scaling in Fast Quantum Quenches in Conformal Field Theories, Phys. Rev. Lett. **112**, 171601 (2014).
[41] S. R. Das, D. A. Galante, and R. C. Myers, Universality in fast quantum quenches, J. High Energy Phys. 02 (2015) 167.
[42] S. R. Das, D. A. Galante, and R. C. Myers, Smooth and fast versus instantaneous quenches in quantum field theory, J. High Energy Phys. 08 (2015) 073.
[43] S. R. Das, D. A. Galante, and R. C. Myers, Quantum quenches in free field theory: Universal scaling at any rate, J. High Energy Phys. 05 (2016) 164.
[44] V. Alba and P. Calabrese, Entanglement dynamics after quantum quenches in generic integrable systems, SciPost Phys. **4**, 017 (2018).
[45] S. Ghosh, K. S. Gupta, and S. C. L. Srivastava, Entanglement dynamics following a sudden quench: An exact solution, Europhys. Lett. **120**, 50005 (2017).
[46] S. Ghosh, K. S. Gupta, and S. C. L. Srivastava, Exact relaxation dynamics and quantum information scrambling in multiply quenched harmonic chains, Phys. Rev. E **100**, 012215 (2019).
[47] A. Polkovnikov, K. Sengupta, A. Silva, and M. Vengalattore, Colloquium: Nonequilibrium dynamics of closed interacting quantum systems, Rev. Mod. Phys. **83**, 863 (2011).
[48] C. Gogolin and J. Eisert, Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems, Rep. Prog. Phys. **79**, 056001 (2016).
[49] P. Calabrese, F. H. L. Essler, and G. Mussardo, Introduction to 'quantum integrability in out of equilibrium systems', J. Stat. Mech. (2016) 064001.
[50] T. Langen, T. Gasenzer, and J. Schmiedmayer, Prethermalization and universal dynamics in near-integrable quantum systems, J. Stat. Mech. (2016) 064009.
[51] T. Kinoshita, T. Wenger, and D. Weiss, A quantum Newton's cradle, Nature (London) **440**, 900 (2006).
[52] S. Hofferberth, I. Lesanovsky, B. Fischer, T. Schumm, and J. Schmiedmayer, Non-equilibrium coherence dynamics in one-dimensional Bose gases, Nature (London) **449**, 324 (2007).
[53] S. Trotzky, Y. A. Chen, A. Flesch, I. P. McCulloch, U. Schollwöck, J. Eisert, and I. Bloch, Probing the relaxation towards equilibrium in an isolated strongly correlated one-dimensional Bose gas, Nat. Phys. **8**, 325 (2012).
[54] M. Gring, M. Kuhnert, T. Langen, T. Kitagawa, B. Rauer, M. Schreitl, I. Mazets, D. A. Smith, E. Demler, and J. Schmiedmayer, Relaxation and prethermalization in an isolated quantum system, Science **337**, 1318 (2012).
[55] M. Cheneau, P. Barmettler, D. Poletti, M. Endres, P. Schauß, T. Fukuhara, C. Gross, I. Bloch, C. Kollath, and S. Kuhr, Light-cone-like spreading of correlations in a quantum many-body system, Nature (London) **481**, 484 (2012).
[56] F. Meinert, M. J. Mark, E. Kirilov, K. Lauber, P. Weinmann, A. J. Daley, and H.-C. Nägerl, Quantum Quench in an Atomic One-Dimensional Ising Chain, Phys. Rev. Lett. **111**, 053003 (2013).
[57] T. Langen, R. Geiger, M. Kuhnert, B. Rauer, and J. Schmiedmayer, Local emergence of thermal correlations in an isolated quantum many-body system, Nat. Phys. **9**, 640 (2013).
[58] T. Fukuhara, P. Schauß, M. Endres, S. Hild, M. Cheneau, I. Bloch, and C. Gross, Microscopic observation of magnon bound states and their dynamics, Nature (London) **502**, 76 (2013).
[59] T. Fukuhara, A. Kantian, M. Endres, M. Cheneau, P. Schauß, S. Hild, D. Bellem, U. Schollwöck, T. Giamarchi, C. Gross, I. Bloch, and S. Kuhr, Quantum dynamics of a mobile spin impurity, Nat. Phys. **9**, 235 (2013).
[60] S. Sotiriadis and J. Cardy, Quantum quench in interacting field theory: A self-consistent approximation, Phys. Rev. B **81**, 134305 (2010).
[61] D. Han, Y. Kim, and M. Noz, Illustrative example of Feynman's rest of the universe, Am. J. Phys. **67**, 61 (1999).
[62] J. S. Prauzner-Bechcicki, Two-mode squeezed vacuum state coupled to the common thermal reservoir, J. Phys. A **37**, L173 (2004).
[63] D. Han, Y. S. Kim, and M. E. Noz, Linear canonical transformations of coherent and squeezed states in the wigner phase space. iii. two-mode states, Phys. Rev. A **41**, 6233 (1990).
[64] Y. S. Kim, Observable Gauge Transformations in the Parton Picture, Phys. Rev. Lett. **63**, 348 (1989).
[65] F. Iachello and S. Oss, Model of *n* Coupled Anharmonic Oscillators and Applications to Octahedral Molecules, Phys. Rev. Lett. **66**, 2976 (1991).

[66] S. Ikeda and F. Fillaux, Incoherent elastic-neutron-scattering study of the vibrational dynamics and spin-related symmetry of protons in the $khco_3$ crystal, Phys. Rev. B **59**, 4134 (1999).
[67] F. Fillaux, Quantum entanglement and nonlocal proton transfer dynamics in dimers of formic acid and analogues, Chem. Phys. Lett. **408**, 302 (2005).
[68] M. Delor, S. Archer, T. Keane, A. Meijer, I. Sazanovich, G. Greetham, M. Towrie, and J. Weinstein, Directing the path of light-induced electron transfer at a molecular fork using vibrational excitation, Nat. Chem. **9**, 1099 (2017).
[69] E. Romero, R. Augulis, V. Novoderezhkin, M. Ferretti, J. Thieme, D. Zigmantas, and R. van Grondelle, Quantum coherence in photosynthesis for efficient solar energy conversion, Nat. Phys. **10**, 676 (2014).
[70] F. D. Fuller, J. Pan, A. Gelzinis, V. Butkus, S. S. Senlik, D. E. Wilcox, C. F. Yocum, L. Valkunas, D. Abramavicius, and J. P. Ogilvie, Vibronic coherence in oxygenic photosynthesis, Nat. Chem. **6**, 706 (2014).
[71] A. Halpin, P. Johnson, R. Tempelaar, R. Murphy, J. Knoester, T. Jansen, and R. Miller, Two-dimensional spectroscopy of a molecular dimer unveils the effects of vibronic coupling on exciton coherences, Nat. Chem. **6**, 196 (2014).
[72] D. N. Makarov, Coupled harmonic oscillators and their quantum entanglement, Phys. Rev. E **97**, 042203 (2018).
[73] J.-Y. Kao and C.-H. Chou, Quantum entanglement in coupled harmonic oscillator systems: From micro to macro, New J. Phys. **18**, 073001 (2016).
[74] A. Jellal, F. Madouri, and A. Merdaci, Entanglement in coupled harmonic oscillators studied using a unitary transformation, J. Stat. Mech. (2011) P09015.
[75] M. S. Abdalla, M. Abdel-Aty, and A.-S. F. Obada, Degree of entanglement for anisotropic coupled oscillators interacting with a single atom, J. Opt. B **4**, 396 (2002).
[76] M. B. Plenio, J. Hartley, and J. Eisert, Dynamics and manipulation of entanglement in coupled harmonic systems with many degrees of freedom, New J. Phys. **6**, 36 (2004).
[77] R. A. Jefferson and R. C. Myers, Circuit complexity in quantum field theory, J. High Energy Phys. 10 (2017) 107.
[78] P. Caputa, S. R. Das, M. Nozaki, and A. Tomiya, Quantum quench and scaling of entanglement entropy, Phys. Lett. B **772**, 53 (2017).
[79] H. A. Camargo, P. Caputa, D. Das, M. P. Heller, and R. Jefferson, Complexity as a Novel Probe of Quantum Quenches: Universal Scalings and Purifications, Phys. Rev. Lett. **122**, 081601 (2019).
[80] K. H. Yeon, H. J. Kim, C. I. Um, T. F. George, and L. N. Pandey, Wave function in the invariant representation and squeezed-state function of the time-dependent harmonic oscillator, Phys. Rev. A **50**, 1035 (1994).
[81] K. Adhikari, S. Choudhury, S. Kumar, S. Mandal, N. Pandey, A. Roy, S. Sarkar, P. Sarker, and S. S. Shariff, Circuit complexity in $\mathcal{Z}_2\mathcal{EEFT}$, arXiv:2109.09759.
[82] K. Adhikari, S. Choudhury, and A. Roy, Krylov complexity in quantum field theory, arXiv:2204.02250.
[83] K. Adhikari, S. Choudhury, H. N. Pandya, and R. Srivastava, PGW circuit complexity, arXiv:2108.10334.
[84] S. Choudhury, A. Mukherjee, N. Pandey, and A. Roy, Causality constraint on circuit complexity from $\mathcal{COSMOEFT}$, arXiv:2111.11468.
[85] K. Adhikari and S. Choudhury, Cosmological krylov complexity, arXiv:2203.14330.
[86] P. Bhargava, S. Choudhury, S. Chowdhury, A. Mishara, S. P. Selvam, S. Panda, and G. D. Pasquino, Quantum aspects of chaos and complexity from bouncing cosmology: A study with two-mode single field squeezed state formalism, SciPost Phys. Core **4**, 026 (2021).
[87] R. A. Jefferson and R. C. Myers, Circuit complexity in quantum field theory, J. High Energy Phys. 10 (2017) 107.
[88] K. Adhikari, S. Choudhury, S. Chowdhury, K. Shirish, and A. Swain, Circuit complexity as a novel probe of quantum entanglement: A study with black hole gas in arbitrary dimensions, Phys. Rev. D **104**, 065002 (2021).
[89] S. Choudhury, S. Chowdhury, N. Gupta, A. Mishara, S. P. Selvam, S. Panda, G. D. Pasquino, C. Singha, and A. Swain, Circuit complexity from cosmological Islands, Symmetry **13**, 1301 (2021).
[90] J. Eisert, Entangling Power and Quantum Circuit Complexity, Phys. Rev. Lett. **127**, 020501 (2021).
[91] S. D. Mathur, Three puzzles in cosmology, Int. J. Mod. Phys. D **29**, 2030013 (2020).
[92] D. Stanford and L. Susskind, Complexity and shock wave geometries, Phys. Rev. D **90**, 126007 (2014).
[93] L. Susskind, Addendum to computational complexity and black hole horizons, Fortschr. Phys. **64**, 24 (2016).
[94] D. A. Roberts, D. Stanford, and L. Susskind, Localized shocks, J. High Energy Phys. 03 (2015) 051.
[95] L. Susskind and Y. Zhao, Switchbacks and the bridge to nowhere, arXiv:1408.2823.
[96] S. Mukherjee, A. Ghose Choudhury, and P. Guha, Generalized damped Milne-Pinney equation and Chiellini method, arXiv:1603.08747.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/362061618

# Examining Systemic Risk using Google PageRank Algorithm: An Application to Indian Non-Bank Financial Companies (NBFCs) Crisis

**Article** *in* International Journal of Mathematical, Engineering and Management Sciences · July 2022

DOI: 10.33889/IJMEMS.2022.7.4.037

CITATIONS
0

READS
19

**2 authors**, including:

Anurag Chaturvedi
Delhi Technological University
**2** PUBLICATIONS **0** CITATIONS

SEE PROFILE

All content following this page was uploaded by Anurag Chaturvedi on 18 July 2022.

The user has requested enhancement of the downloaded file.

International Journal of Mathematical, Engineering and Management Sciences
Vol. 7, No. 4, 575-588, 2022
https://doi.org/10.33889/IJMEMS.2022.7.4.037

# Examining Systemic Risk using Google PageRank Algorithm: An Application to Indian Non-Bank Financial Companies (NBFCs) Crisis

**Anurag Chaturvedi**
University School of Management & Entrepreneurship,
Delhi Technological University,
East Delhi Campus, Vivek Vihar, Phase-2, Delhi-110095, India.
*Corresponding author*: canurag17@gmail.com

**Archana Singh**
Delhi School of Management,
Delhi Technological University,
Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India.
E-mail: archanasingh205@gmail.com

(Received on January 8, 2022; Accepted on June 9, 2022)

**Abstract**
In the recent financial crises, attention has shifted towards "too-central-to-fail" to recognize the sources of systemic risk. The NBFC Crisis of 2018-19 adversely affected other financial institutions and the real economy of India. The NBFCs crisis highlighted the role of smaller institutions in perpetuating and amplifying the crisis. Thus, the present study models the interconnection of NBFCs with the rest of financial institutions using a complex Granger-causality network based on returns data. The PageRank algorithm identifies the central and important nodes and ranks financial institutions in pre-crisis and crisis periods. The financial institutions are also ranked based on the maximum percentage loss suffered during the crises. Using non-parametric rank-based regression, the PageRank ranking of financial institutions in the pre-crises period (explanatory variable) is regressed with the ranking of financial institutions based on maximum percentage loss suffered by them during the crises period (dependent variable) along with Leverage and Size as control variables. We found that PageRank from pre-crisis can significantly identify most financial institutions that suffered loss during NBFCs crises even in the presence of control variables.

**Keywords**- Complex financial network, Network centrality, Pagerank centrality, Systemic risk, Non-banking financial companies, Early warning signal, Too-central-to-fail.

## 1. Introduction

Over the past decade, the frequency and severity of crises have increased. The globalization and financial integration of economies have also ensured that crises from a particular region quickly spread to another part of the world. The Global Financial Crisis (GFC) of 2007-08 highlighted the role of "too-big-to-fail" institutions in exacerbating the crisis due to their size and operations. The identification of systemically important financial institutions (SIFIs) post GFC also relied on “too-big-to-fail” methodology. It does not consider the complex relationships between the financial institutions. Many smaller institutions amplified and propagated the shocks during the GFC through their complex credit intermediation and maturity transformation. Thus "too-connected-to-fail" and "too-central-to-fail" institutions pose a more significant risk to the financial system. "Too-big-to-fail" focuses on big financial institutions and thus mainly uses data from the nodes. "Too-connected-to-fail" focuses on financial institutions' relationships and thus prioritizes information from edges connected to big nodes. "Too-central-to-fail" focuses on information from both nodes and edges of all financial institutions in the network. It helps in discovering the intricate relationship which can become critical in distress. The "too-central-to-fail" approach explores the network centrality measures like eigenvector centrality, betweenness centrality, Katz centrality, and PageRank centrality.

Thus, recently "too-central-to-fail" has replaced the "too-big-to-fail" and "too-interconnected-to-fail" in identifying systemic nodes.

The earliest research of (Danielsson & De Vries, 2000; Hartmann et al., 2004; Lehar, 2005; Gray et al., 2007; Chan-Lau et al., 2009; Huang et al., 2009; Acharya et al., 2010; Adrian and Brunnermeier, 2011; Brownlees and Engle, 2012; Hautsch et al., 2015) based on historical data and often identifies "too-big-to-fail" institutions. They work best when systemic risk is well represented by historic data and does not consider the simultaneous losses experienced by newly connected parts due to rapid financial innovations. The modern financial system is a complex network of interconnected institutions at many levels. Therefore, the complex network-based systemic risk measures like like Billio et al.(2012) PCAS and Granger-causal network; Diebold and Yilmaz (2014) variance decomposition; Battiston et al. (2012) DebtRank and Härdle et al. (2016) TENET (Tail Event-driven NETwork) have gained importance as they can capture and simulate time-varying intricate relationships between financial institutions and based on "too-interconnected-to-fail" hypothesis. However, the theoretical literature is inconclusive whether dense interconnection makes financial networks more resilient to shocks or make it more fragile by amplifying a large negative shock (Allen and Gale, 2000; Freixas et al., 2000; Dasgupta, 2004; Brunnermeier and Pedersen, 2009; Acemoglu et al., 2015; and Minoiu et al., 2015). Thus, it is necessary to study the role of institutions in contributing to the systemic risk of the network from "too-central-to-fail" perspective.

Taking a cue from the study done in social networks, it was established that not all nodes in financial networks contribute effectively to the spread of contagion. Thus, studying the topological position of nodes becomes critical. Billio et al. (2012) used Eigenvector and Closeness centrality measures to study GFC systemic institutions. Thurner and Poledna (2013) proposed modified Katz centrality based on DebtRank to study systemic risk of the hypothetical network. Kuzubaş et al.(2014) calculated a centrality measure for the overnight money market to study the Turkish Banking Crisis of 2000. Wang and Huang (2021) used network centrality measures to study the tail dependence of the Chinese financial network from 2009 to 2018. Xu and Corbett (2020) calculated FIRank, a measure of interconnectedness based on the PageRank algorithm, for ranking countries on the financial interconnectedness with the global bank-lending network. Yun et al. (2019) used the PageRank algorithm to simulate the network. PageRank captures network topology better than balance sheet measures like CoVaR and MES. Thus, we propose to use PageRank centrality measure to study the critical nodes of systemic importance in the financial network.

The study's objective is to apply the PageRank algorithm approach in examining the systemic risk of financial institutions by taking the NBFC crisis of 2018-19 as a systemic event. Reserve Bank of India (RBI), the central bank of India, defines NBFC "as a registered company under Company Act, 1956 and engaged in the business of loans and advances, acquisition of shares/ stocks/ bonds/ debentures/ securities issued by Government or local authority or other marketable securities like leasing, hire-purchase, insurance business, chit funds"[1]. Unlike banks, they are subjected to light touch regulations by RBI. Often Banks use NBFCs for regulatory arbitrage, and most NBFCs are either subsidiaries or have ownership of Banks. NBFCs do not have federal support or deposit insurance like banks in a crisis. NBFC crises started in June 2018 when one of the subsidiaries of Infrastructure Leasing and Financial Services (IL&FS), classified as "Systemically Important Non-Deposit taking NBFC" (NBFC-ND-SI), defaulted on its debt papers. This contagious effect quickly spread to other NBFCs, and by the end of the third week of September 2018, the top 15 NBFCs are estimated to have lost over Rs 75000cr in market capitalization[2]. In Oct 2018, the benchmark index NIFTY and SENSEX hit their six-month low[3]. It was called India's Lehman moment, comparing events preceding the Global Financial Crisis. The crises prevailed over 2019, and many NBFCs

like Reliance Home Finance, Reliance Commercial Finance, etc., were declared bankrupt or shut down their business. As the crises intensified, Mutual Funds started dumping the Commercial Papers (CPs) of stressed NBFCs creating liquidity shock and further intensifying crises. This development has led to a concern about the systemic risk possessed by NBFCs which are smaller institutions than Banks.

We adopted the Granger-causal relationships of Billio et al. (2012) to construct the complex network of financial institutions. Mensah and Premaratne (2017), Zhang and Broadstock (2018), and Lai and Hu (2020) have also used Granger-causal network to study the interconnectedness and systemic risk of the financial institutions. We calculated the PageRank centrality score using Page et al. (1999) and Yun et al. (2019) methodology for the financial institutions. Taking the NBFC crisis of 2018-19 as a systemic event, we compared topological structures of pre-crisis and crisis periods. To explore the predictive property, we regressed the PageRank of financial institutions in the pre-crisis period on the maximum percentage loss of the institutions during the crisis period by controlling the firm's level factors. (Fong et al., 2020; Jin, 2020) extended complex networks to non-bank financial institutions. However, this study is the first to extended the applicability of the "too-central-to-fail" concept in modeling the systemic crisis limited to the non-bank financial institutions. The study is also the first to model the Indian NBFC crisis using the PageRank algorithm. The study uses real-time market data, which is publicly available, making it a dynamic and responsive measure. It also does not depend on balance sheet data which generally comes with a lag. Thus, regulators can use the proposed PageRank measure as an alarm or early warning signal to detect systemic risk in the network.

In the remaining paper, Section 2 develops empirical framework and hypothesis; Section 3 defines data, Section 4 empirical analysis of developed indicators on the Indian NBFC crisis; Section 5 discusses the usefulness of study and policy recommendation, and Section 6 presents the critical conclusion derived from the study.

## 2. Empirical Framework

### 2.1 Granger-Causal Complex Network

The pairwise Granger-causality test (Billio et al.2012) measures the dynamic propagation of shocks to the financial system. The rolling window (sub-periods) of 52 weeks performs the Granger-causality tests and builds the network parameters.

Let $r_t^i$ and $r_t^j$ be the two stationary log return time series of financial institutions assumed to have zero mean. If $r_t^j$ contain information that helps in predicting $r_t^i$ beyond the information that is contained in lagged values of $r_t^i$ alone then $r_t^j$ is said to be "granger-cause" $r_t^i$.

$$r_{t+1}^i = a^i {r_t}^i + b^{ij} r_t^j + e_{t+1}^j$$
$$r_{t+1}^j = a^j {r_t}^j + b^{ji} r_t^i + e_{t+1}^i \quad (1)$$

where, $e_{t+1}^j$, $e_{t+1}^i$ are uncorrelated residual series assumed to be white noise and $a^i, b^{ij}, a^j, b^{ji}$ are coefficients of the model. Then $r_t^j$ Granger-causes $r_t^i$ if $b^{ij}$ is different from zero. BIC (Bayesian Information Criteria) is used to determine the number of lags in the model.

## 2.2 Control for Heteroskedasticity and Return Autocorrelations

As equation (1) is a regression equation with Ordinary Least Square (OLS) estimates, so error term $\epsilon_t$may have conditional heteroskedasticity and serial correlations. It results in inconsistent OLS estimates. Financial assets return shows volatility clustering phenomena leading to persistence in amplitudes of price changes. Thus, a baseline Generalized Auto-Regressive Conditional Heteroskedasticity GARCH (1,1) model is used for our log-returns of financial institutions. Let $r_t^i$ be log-return series of institution $i$ and $a_t^i = r_t^i - \mu_i$ be the innovation of institution $i$ at time $t$. Then $a_t^i$ and $\sigma_{it}^2$ is conditional on the financial system information and $\mu_i, \omega_i, \alpha_i, \beta_i$ are the coefficients of the model.

$$a_t^i = \sigma_{it}\epsilon_t^i \ , \epsilon_t \sim WN(0,1),\ \sigma_{it}^2 = \omega_i + \alpha_i a_{t-1}^2 + \beta_i \sigma_{it-1}^2 \quad (2)$$

To control the heteroskedasticity and return autocorrelation among the institutions, the Granger-causality test in equation (1) is performed on $\tilde{r}_t^i = r_t^i / \hat{\sigma}_{it}$ where $\hat{\sigma}_{it}$ is estimated using the GARCH(1,1) model as defined in (2)

Granger causality helps in modeling time-varying and complex relationships among financial institutions. An adjacency matrix of the network of N financial institutions is defined as

$$(j \rightarrow i) = \begin{cases} 1, & if\ j\ Granger\ causes\ i \\ 0, & otherwise \end{cases} \quad (3)$$

and define $(j \rightarrow j) = 0$.

## 2.3 PageRank Centrality

PageRank algorithm (Page et al., 1999) ranks the financial institution based on relative importance. The incoming link to a node is seen as a vote of support, and thus, that node becomes a democracy where other nodes vote for importance by linking to them. Yun et al. (2019) used PageRank to measure the centrality of financial institutions from a "too-central-to-fail" perspective. Based on the same method, we calculate the effect matrix whose entity ($e_{ijt}$) uses F-values of the granger causality network to account for the wider variations. We define the effect weight of each financial institution as $E_{ijt} = \frac{e_{ijt}}{\sum_i e_{ijt}}$ where $e_{ijt}$ and $E_{ijt}$ denotes the extent of the effect and effect weight respectively by financial institution $i$ on financial institution $j$ at time $t$. To obtain the PageRank

$$PageRank_{it} = \frac{1-\alpha}{N} + \alpha \sum E_{ijt} PageRank_{jt} \quad (4)$$

where, $PageRank_{it}$ is the rank of the firm $i$ at time $t$, $\alpha$ is the damping factor and is generally set to 0.85, and $N$ is the total number of financial institutions in the system. PageRank is always the positive value and a higher PageRank indicates that the institution has a higher contribution to the systemic risk.

## 2.4 Hypothesis Formulation

Based on the PageRank centrality measure, the financial institutions are ranked in descending order. For detecting stressed financial institutions during the crises, the variable maximum percentage

loss(Max%Loss) is defined as the difference between the market capitalization of the institution(Asset Under Management (AUM) in case of Mutual Funds) at the beginning of the crisis(i.e.,the start of June 2018) and the minimum market capitalization during the entire shadow bank crisis period (June 2018 to Dec 2019) divided by the market capitalization or fund size at the beginning of the crisis period.

The financial institutions are ranked based on Max%Loss, where rank 1 means that the financial institution has suffered maximum loss during financial crises.

**Hypothesis Testing:** The Kendall τ(1938) statistic detects the relationship between the rankings based on PageRank centrality and Max%Loss. Kendall τ(1938) is a non-parametric test to measure the ordinal association between two measured quantities. $\boldsymbol{\tau}$lies between -1 & 1. The $\boldsymbol{\tau}$ is also corrected for the ties between rankings. Kendall$\boldsymbol{\tau}$ is even less sensitive to outliers and has superior statistical power to Pearson's coefficient which is important for systemic events.
The purpose of the present study is to identify and quantify financial crisis periods and to determine the predictive power of PageRank centrality measure in predicting which institutions will suffer maximum loss in out-of-sample tests. Thus, the following hypothesis is formed:

$$H_\alpha : PageRank\ score\ is\ significant\ determinant\ of\ Max\%Loss.$$

## 2.5 Multivariate Regression

We used the following out-of-sample rank-regression to study the predictive power of the PageRank centrality measure.

$$Max\%Loss_{i,crisis} = \alpha + \beta_1 PageRank_{i,pre-crisis} + \beta_2 Size_{i,pre-crisis} + \beta_3 Leverage_{i,pre-crisis} + \varepsilon \quad (5)$$

Here $i$ represents the financial institution. The dependent variable is Max%Loss, an institution that suffered during the crisis. The primary explanatory variable is the *PageRank* score of the institution measured in the pre-crisis period. We controlled the firm-specific factors that can affect the institutions' systemic risk by using the Size and Leverage of the institutions in the pre-crisis period. *Size* is defined as the natural logarithm of total assets, ln$(Total\ Assets)$, and *Leverage* as the ratio of debt to equity $Debt/Equity$. We expect a positive and significant $\beta_1$ even in the presence of control variables *Size* and *Leverage*.

## 3. Data and Empirical Results

The study employs the Private Banks (PB), Public Sector Banks (PSB), Non-Banking Financial Companies (NBFC), Housing Finance Companies (HFC), and Liquid Debt Mutual Funds (LDMF). HFCs are also NBFCs whose primary business is to provide financing to housing. HFCs were modeled separately as they suffered maximum distress during the NBFC crisis. NBFCs depend on bank borrowings and Commercial Papers to meet working capital for operations. LDMFs are a major buyer of NBFCs' Commercial Papers along with banks. The short-term fund is used to lend for long-term infrastructure, housing, construction, etc., causing asset-liability mismatch (ALM) issues.

The study uses weekly return data for PBs, PSBs, NBFCs, HFCs, and Net Asset Value (NAV) return data for LDMFs. The period selected is from January 2017- January 2020. S&P BSE Finance Index is used to draw a sample of PB, PSB, HFC & NBFCs. Association of Mutual Funds of India (AMFI) report is used to select LDMFs. The weekly return series data is taken from CMIE Prowess, and NAV return data from ACE MF.

**Table 1.** Composition of financial institutions.

| Constituents | # Number | % of representation of total Market Capitalization for Banks, HFCs, and NBFCs, and Asset Under Management (AUM) for LDMFs as of March 2020 |
|---|---|---|
| Private Banks (PB) | 8 | 85 |
| Public Sector Banks (PSB) | 10 | 89 |
| Housing Finance Companies (HFC) | 5 | 99 |
| Non-Banking Finance Company (NBFC) | 21 | 69 |
| Liquid Debt Mutual Funds (LDMF) | 8 | 74 |

Table 2 shows the summary statistics of weekly logarithmic returns of financial institutions. Dickey and Fuller (1981) and Phillips and Perron (1988) tests check series stationarity. The stationarity is a necessary condition for applying pairwise Granger-causality tests. All the financial institutions time series are covariance stationary. The negative mean return and negative skewness imply loss to the investors. Kurtosis greater than three means log return distribution is more peaked and fatter tailed than the standard distributions. It implies that there is more chance of observing extreme or abnormal returns. Thus, the institutions with a negative mean return, negative skewness, and large kurtosis are crucial for our analysis.

**Table 2.** Summary statistics of financial institutions.

| Financial Institutions | Abbreviation | Mean(%) | Std_Dev(%) | Min(%) | Max(%) | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|
| Axis Bank Ltd. | AXIS | 0.334292 | 3.70687 | -13.996 | 12.72232 | -0.11101 | 1.45042 |
| City Union Bank Ltd. | CUB | 0.471559 | 3.396569 | -9.33992 | 9.942288 | 0.010289 | 0.503474 |
| H D F C Bank Ltd. | HDFCB | 0.481075 | 2.034737 | -5.18879 | 6.090392 | 0.212894 | 0.495926 |
| I C I C I Bank Ltd. | ICICI | 0.54173 | 3.802105 | -8.77222 | 15.52305 | 0.56947 | 1.634147 |
| Indusind Bank Ltd. | INDUS | 0.230609 | 4.031364 | -15.7273 | 18.27098 | 0.158926 | 3.310951 |
| Kotak Mahindra Bank Ltd. | KOTAK | 0.534872 | 2.618217 | -8.15476 | 10.36817 | 0.19658 | 1.376781 |
| Federal Bank Ltd. | FED | 0.212609 | 4.201613 | -10.9183 | 15.28349 | 0.236453 | 0.772422 |
| Yes Bank Ltd. | YES | -0.9896 | 9.194739 | -35.5217 | 26.08091 | -0.66113 | 2.221234 |
| Bank Of Baroda | BOB | -0.2638 | 5.645859 | -18.7244 | 25.68177 | 0.689325 | 3.243198 |
| Bank Of India | BOI | -0.26256 | 6.55626 | -18.1873 | 29.5352 | 0.807366 | 2.655518 |
| Canara Bank | CAN | -0.12967 | 5.975111 | -20.9416 | 28.93743 | 0.581479 | 3.554909 |
| Central Bank Of India | CBI | -0.96251 | 7.132521 | -25.6834 | 31.6876 | 0.397422 | 4.420071 |
| Indian Bank | INDB | -0.45014 | 6.624989 | -19.5819 | 20.44384 | 0.21244 | 0.865743 |
| Indian Overseas Bank | IOB | -0.51134 | 4.455384 | -10.4094 | 15.64055 | 0.670257 | 1.355178 |
| Punjab National Bank | PNB | -0.38755 | 6.517907 | -22.2267 | 41.73586 | 1.163536 | 10.45713 |
| State Bank Of India | SBI | 0.182267 | 4.461946 | -11.7805 | 24.79216 | 1.07117 | 5.627177 |
| Uco Bank | UCO | -0.4425 | 5.747935 | -10.883 | 35.93031 | 1.982775 | 9.266191 |
| Union Bank Of India | UNION | -0.5128 | 6.645952 | -17.3674 | 33.3118 | 0.788049 | 3.588716 |
| Can Fin Homes Ltd. | CANFIN | 0.137797 | 5.027593 | -20.8047 | 15.03889 | 0.049389 | 1.573127 |
| Dewan Housing Finance Corpn. Ltd. | DHFL | -1.69372 | 11.18042 | -63.1482 | 24.15523 | -2.29752 | 8.92448 |
| G I C Housing Finance Ltd. | GICH | -0.29614 | 5.686067 | -21.432 | 20.97907 | -0.01383 | 1.959854 |
| Housing Development Finance Corpn. Ltd. | HDFC | 0.439927 | 2.777253 | -7.11769 | 10.45762 | 0.339777 | 1.061961 |
| L I C Housing Finance Ltd. | LICH | -0.13259 | 3.95153 | -12.2675 | 13.19275 | -0.01456 | 0.499158 |
| Bajaj Finserv Ltd. | BAJ_FIN | 0.804814 | 3.833633 | -11.0913 | 10.97736 | -0.08996 | 0.427177 |
| I I F L Finance Ltd. | IIFL | -0.36727 | 8.90523 | -80.5081 | 18.6136 | -4.52792 | 39.37111 |
| J M Financial Ltd. | JMFIN | 0.194178 | 5.458024 | -18.5688 | 16.88881 | -0.00852 | 0.544158 |
| Religare Enterprises Ltd | REL | -1.11492 | 10.39327 | -27.3293 | 43.54213 | 0.545469 | 2.280063 |
| IFCI Ltd. | IFCI | -0.85108 | 6.335381 | -23.8156 | 25.21311 | 0.225625 | 2.737318 |
| Power Finance Corpn. Ltd. | PFC | -0.00504 | 4.83824 | -13.9083 | 16.82958 | 0.092799 | 0.529145 |
| R E C Ltd. | REC | 0.096642 | 5.057822 | -19.1621 | 18.93926 | -0.03719 | 1.665538 |
| Tourism Finance Corpn. Of India Ltd. | TFCI | 0.141956 | 6.052502 | -24.2699 | 21.83576 | -0.14111 | 4.973959 |
| Bajaj Finance Ltd. | BAJ_FIN | 1.041482 | 4.017809 | -11.6054 | 14.39224 | 0.170782 | 1.192955 |
| Bajaj Holdings & Invst. Ltd. | BAJ_HL | 0.412572 | 3.126099 | -13.4939 | 11.0087 | -0.02793 | 2.890414 |

**Table 2.** *Continued*.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Cholamandalam Investment & Finance Co. Ltd. | CHOLA | 0.305698 | 4.138868 | -8.91946 | 10.54501 | 0.301148 | -0.31483 |
| Edelweiss Financial Services Ltd. | EDEL | 0.122921 | 7.995963 | -29.4308 | 23.13902 | -0.18346 | 1.396137 |
| L & T Finance Holdings Ltd. | L&TF | 0.214884 | 5.450814 | -15.5262 | 22.03432 | 0.224353 | 1.473308 |
| Magma Fincorp Ltd. | MAG | -0.34237 | 6.491625 | -24.4396 | 24.24678 | 0.081554 | 2.100321 |
| Mahindra & Mahindra Financial Services Ltd. | M&MF | 0.170572 | 4.903248 | -19.7095 | 21.59613 | 0.235862 | 2.768684 |
| Manappuram Finance Ltd. | MANA | 0.674046 | 5.201224 | -18.0184 | 14.07662 | -0.04754 | 0.673289 |
| Motilal Oswal Financial Services Ltd. | MOTOS | 0.30579 | 5.649899 | -13.0923 | 19.22658 | 0.440702 | 0.394769 |
| Muthoot Finance Ltd. | MUTH | 0.652315 | 4.031772 | -12.552 | 13.17436 | 0.134566 | 0.08237 |
| Shriram City Union Finance Ltd. | SH_CIT | -0.15036 | 3.924391 | -10.0771 | 15.48557 | 0.287507 | 1.616644 |
| Shriram Transport Finance Co. Ltd. | SH_TR | 0.194179 | 5.139068 | -15.5961 | 18.76867 | 0.222944 | 1.06192 |
| Tata Investment Corpn. Ltd. | TICL | 0.233727 | 3.481101 | -10.1634 | 15.77827 | 0.969773 | 3.321327 |
| Axis Liquid Fund- Reg(G) | AXIS_M | 0.127874 | 0.018069 | 0 | 0.205384 | -1.56439 | 17.49282 |
| Baroda Liquid Fund(G) | BAR_MF | 0.127583 | 0.01892 | 0 | 0.197111 | -1.46409 | 13.06267 |
| HDFC Liquid Fund(G) | HDFC_M | 0.124672 | 0.018655 | 0 | 0.177808 | -1.78485 | 11.73002 |
| ICICI Pru Liquid Fund(G) | SBI_M | 0.125309 | 0.018249 | 0 | 0.219017 | -0.87274 | 18.3344 |
| ICICI Pru Liquid Fund(G) | ICICI_M | 0.126709 | 0.01856 | 0 | 0.222609 | -1.12925 | 16.72627 |
| L&T Liquid Fund(G) | L&T_MF | 0.126896 | 0.018636 | 0 | 0.187435 | -1.79365 | 12.95456 |
| BNP Paribas Liquid Fund(G) | BNP_MF | 0.127038 | 0.017928 | 0 | 0.180221 | -2.07157 | 14.76169 |
| Aditya Birla SL Liquid Fund(G) | ABSL_M | 0.12766 | 0.018222 | 0 | 0.208723 | -1.60752 | 16.1296 |

## 4. Empirical Analysis and Results

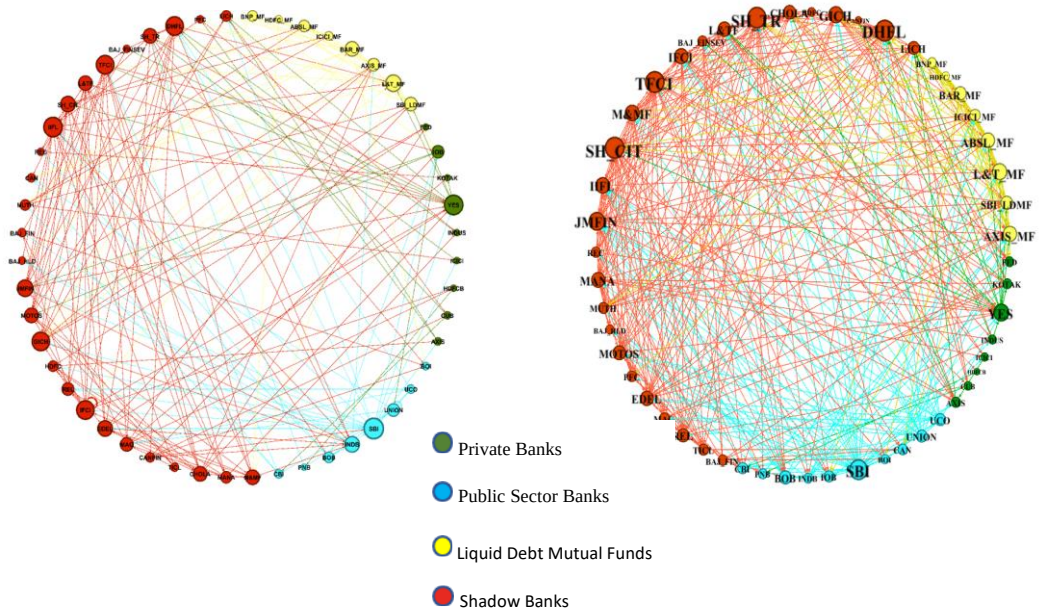
This section empirically investigates the use of PageRank centrality in measuring the institutions' systemic risk. We constructed the complex financial network using a rolling window (sub-periods) of 52 weeks. The pre-crisis period is selected from June 2017-May 2018, and the crisis period from Aug 2018- July 2019. We compared the topological graph and PageRank scores of pre-crisis and crisis periods. Then PageRank scores of the pre-crisis period are rank-regressed with Max%Loss during the crisis period.

### 4.1 Comparison of Topological Structures

The pre-crisis period (Figure 1a) and crisis period (Figure 1b) are compared using the complex network where nodes represent financial institutions and edges represent significant granger-causal connections. The node size represents its degree, i.e., a larger node connects to more networks. Here, the green node is a PB, the blue node is a PSB, the red node is NBFCs, and the yellow node is LDMFs. As it is a directed network, the edges assume color same as the source node. So, if the NBFC granger causes a PB, the edge color will be red. The complex graph of the crisis period is denser than the pre-crisis period. Also, financial institutions have higher interconnectedness during the crisis period.

During the pre-crisis period, the network has an optimal level of interconnectedness arising due to the financial institutions' common exposures and operational activities. The pre-crisis period represents a typical situation where NBFCs perform credit intermediation and risk transformation of the liquidity provided by the banks and liquid debt mutual funds. Also, the pre-crisis complex graph has higher risk absorbing and diversifying capabilities, making it shockproof to small distress events. During the crisis period, the interconnections among institutions got intensified. This reduced the diversification benefit, and any shock to the system gets amplified and spread like contagion. The denser network represents a high probability of risk spillover. In both pre-crisis and crisis periods, certain institutions have a higher degree and are highly interconnected with other

institutions. However, not all of them suffered higher loss during the crisis period. Thus, visualizing complex graph help in picking the nodes that have high interlacing with other financial institutions. However, we need to identify central nodes which create greater externality in the system.

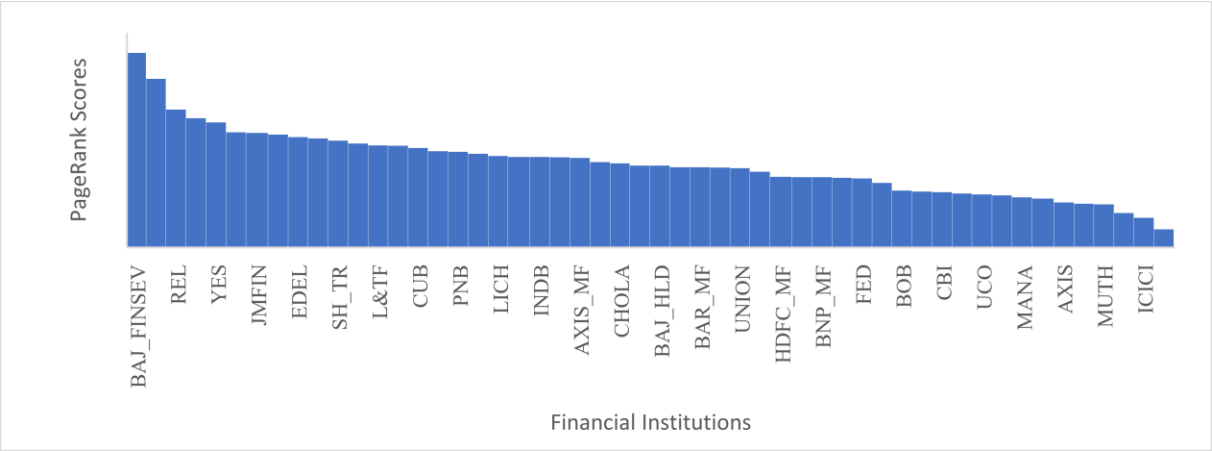



(a) (b)

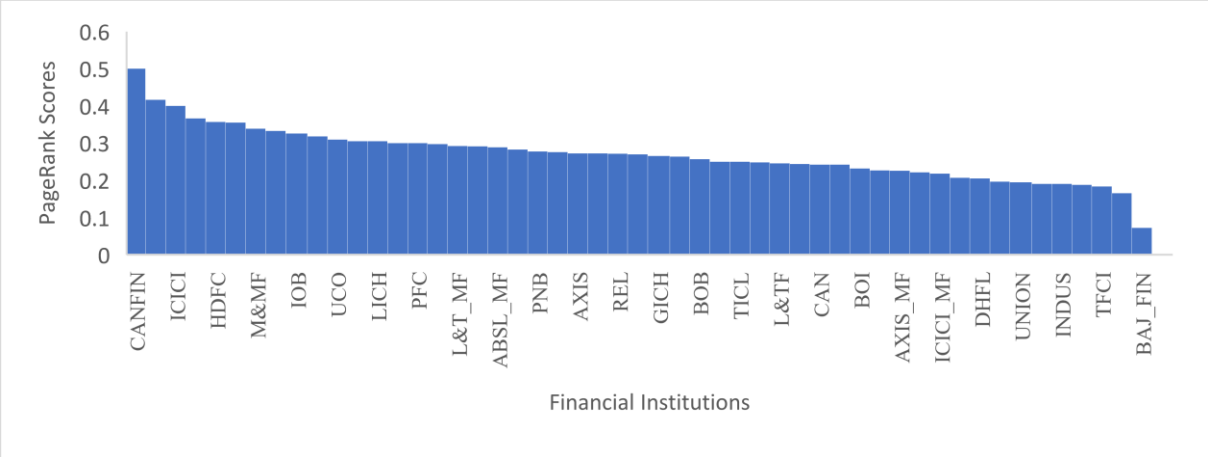
**Figure 1.** Complex Granger-causality NBFC network comparison for (a) the pre-crisis period (b) the crisis period. The edges have color same as the source node.

## 4.2 Comparison of PageRank Centrality Scores

Figure 2(a) and Figure 2(b) shows the PageRank centrality of the institutions in both the tranquil and crisis period, respectively. The PageRank centrality score of the institution in the financial network represents its ability to receive incoming links from high centrality neighbors. The algorithm dilutes the incoming centrality in proportion to the outgoing links from that high centrality neighbors. Unlike eigenvector centrality, in PageRank, the incoming connection from a parsimonious (low degree node) node is worthier than connections from a high degree centrality node. Thus, PageRank also represents systemic vulnerability and, in a true sense, captures the "too-central-to-fail" fallacy. PageRank scores are higher in the crisis period than the tranquil period. There are five shadow bank institutions, SH_CIT, TFCI, SH_TR, DHFL, and MANA, in the crisis period, in the top ten, and the rest are banks and mutual funds.




(a)




(b)

**Figure 2.** Page Rank Centrality for (a) Pre-crisis period (b) Crisis period.

## 4.3 Hypothesis Testing

The financial institutions are ranked from 1 to 52 based on PageRank score for the pre-crisis period June 2017-May 2018. The entire NBFC crisis period from June 2018 to Dec 2019 is chosen as an out-of-sample period. Then financial institutions are ranked from 1 to 52 based on the maximum percentage financial loss (Max%Loss) suffered by each financial institution in the NBFC crisis from June 2018 to Dec 2019. The financial institution with the highest measure value is ranked one, and the lowest is 52.

Using Max%Loss ranking in crisis period as dependent variable and PageRank of financial institutions in the pre-crisis period as an explanatory variable, we reported coefficient, t-statistic, $Kendall\ \tau$, and its significance in Table 3. Using the same methodology, we also reported the rank-regression test statistic for the control variable Size and Leverage.

**Table 3.** Out-of-Sample PageRank centrality on univariate rank regression.

| | Coeff | t-statistic | Kendall $\tau$ |
|---|---|---|---|
| PageRank_Centrality | 0.4226 | 3.297 | 0.288*** |
| Size | -0.2258 | -1.409 | -0.127 |
| Leverage | -0.1875 | -1.295 | -0.130 |

Note:*p<0.1; **p<0.05; ***p<0.01

## 4.4 Multivariate Regression Result

Table 4 reported the out-of-sample analysis of the PageRank measure. Using non-parametric multivariate rank-based regression. The PageRank centrality is statistically significant in predicting Max%Loss even after adding control variable Size and Leverage. The asset size and leverage have negative coefficients in all the regression and are not statistically significant in predicting the Max%Loss. That means the smaller institutions suffered more losses compared to bigger institutions. It may be because NBFCs are smaller institutions than private banks and public sector banks. Also, banks have a bigger capital buffer to sustain the financial distress. It also proves that measures based on "too-big-to-fail" and "too-interconnected-to-fail" that take size into account may not detect the systemic financial institutions during the crisis.

**Table 4.** Out-of-sample analysis. Parameter estimates of multivariate rank regression of Max%Loss as dependent variable and PageRank centrality as independent variable, size and leverage as control variables.

| **Dependent Variable** | | | |
|---|---|---|---|
| **Max%Loss** | **1** | **2** | **3** |
| PageRank_Centrality | 0.452*** | 0.419** | 0.412** |
| | (0.005) | (0.010) | (0.011) |
| Size | | -0.089 | -0.137 |
| | | (0.572) | (0.399) |
| Debt | | | -0.150 |
| | | | (0.336) |
| Constant | 12.355 | 15.478 | 29.283 |
| | (0.028) | (0.037) | (0.045) |
| Observations | 52 | 52 | 52 |
| Adjusted R2 | 0.160 | 0.169 | 0.187 |

Note:*p<0.1; **p<0.05; ***p<0.01

Table 5 presents the top twenty institutions from Largest to Smallest based on Max%Loss from the entire crisis period June 2018 to Dec 2019 and PageRank, Size and Leverage from the pre-crisis period June 2017-May 2018. There are twelve institutions of PageRank centrality score, three institutions based on Size of assets and seven institutions based on Leverage

**Table 5.** Comparison of top twenty financial institutions based on Max%Loss from crisis period and Granger-causal-network-based measures from the pre-crisis period.

| **Crisis Period** | **Precrisis Period Institution Ranking (From Highest to Lowest) On Network Parameters** | | |
|---|---|---|---|
| **Max%Loss** | **Size** | **Leverage** | **Page_Rank Centrality** |
| DHFL | SBI | SBI | BAJ_FINSEV |
| IIFL | HDFCB | REC | IOB |
| YES | ICICI | LICH | REL |
| CBI | PNB | BAJ_FIN | GICH |
| TFCI | BOB | CUB | YES |
| IFCI | AXIS | IFCI | TFCI |
| MAG | CAN | MAG | JMFIN |
| EDEL | BOI | EDEL | DHFL |

**Table 5.** *Continued*.

| | | | |
|---|---|---|---|
| JMFIN | UNION | REL | EDEL |
| INDB | HDFC | IIFL | SH_CIT |
| MOTOS | CBI | MUTH | SH_TR |
| L&TF | YES | FED | KOTAK |
| SH_CIT | PFC | CANFIN | L&TF |
| GICH | KOTAK | MANA | L&T_MF |
| M&MF | REC | UNION | CUB |
| SH_TR | INDB | IOB | M&MF |
| L&T_MF | IOB | JMFIN | PNB |
| CHOLA | INDUS | KOTAK | HDFCB |
| BAR_MF | UCO | CAN | LICH |
| REL | LICH | CHOLA | SBI_LDMF |

## 5. Discussions

The study combined and extended the work of Billio et al. (2012) and Yun et al. (2019) to identify the systemic risk of institutions using the PageRank centrality approach. The PageRank centrality score can predict systemic institutions in line with the study of Xu and Corbett (2020) and Yun et al. (2019). The empirical results proved that asset size is not a significant predictor of systemic results and supports previous studies (Billio et al., 2012; Yun et al., 2019; and Wang and Huang, 2021). It also extended (Fong et al., 2020; Jin, 2020) in modeling the interconnectedness of non-bank financial companies or shadow banks using the complex network. After the NBFC crisis, the Reserve Bank of India (RBI) implemented a regulatory framework based on the systemic significance of shadow institutions. RBI classified systemic importance as a function of shadow lenders' size, activity, and risk profile[3,4] As shadow institutions operate in niche areas, there cannot be a one-size-fits-all regulatory framework, which can hamper their operational activities. Also shadow banks have smaller size as compared to banks. However, their activities of credit intermediation and maturity transformation make their role significant in financial network. Thus, a measure which is based on size or leverage will not reflect about the systemicity of the institutions.The PageRank score based on centrality measure can identify systemic risk of institutions based on their interconnectedness and position in financial network without getting affected by their size and leverage. Thus, it also supports (Billio et al., 2012; Lai and Hu, 2020; Mensah and Premaratne,2017; Zhang and Broadstock, 2018) for using Granger-causal network graphs in modeling complex financial relationships between the institutions. The PageRank centrality based on Granger-causal relationship serves as an unconditional, dynamic, and parsimonious way of generating an early-warning signal for measuring institutions' systemic risk.

## 6. Conclusion

The Global Financial Crisis has established that "too-central-to-fail" is a more significant concern for the systemic risk than "too-big-to-fail" and "too-connected-to-fail" institutions. As the frequency and severity of the crisis are increasing and the role of smaller institutions in spreading the crisis is becoming prominent, the too-central-to-fail approach proves a valuable measure in identifying critical nodes. To this end, we propose PageRank centrality measure on Granger-causal financial network to identify systemic institutions using the "too-central-to-fail" approach.

We used the Indian NBFCs crisis of 2018-19 as systemic event to test the applicability of PageRank centrality scores in identifying systemic institutions and testing the out-of-sample property of PageRank score in identifying institutions that may suffer most during the crisis. The complex network graph based on significant Granger-causal relationship between asset returns of the financial institutions has been studied for pre-crisis and crisis period. The topological structure of the crisis period is denser than the pre-crisis period indicating that institutions are more tightly coupled during the crisis period. Comparing

PageRank centrality measures, the financial institutions that hold prestigious positions due to their financial linkages have higher systemic vulnerability during the crisis. These critical positions give some financial institutions a dual role of risk transmitter and risk receptor, making them systemically important nodes. The rank-based regression shows that the PageRank centrality measure has out-of-sample properties in predicting market capitalization loss of institutions in the crisis period. The PageRank is a significant predictor of systemic institutions even after introducing the firm-level control variables. The high commonality in institution ranking on PageRank centrality measure from the pre-crisis period and market capitalization loss from the crisis period demonstrated its use as an alarm or early warning signal for detecting systemic risk. Regulators can swiftly isolate these critical nodes in the crisis and prevent the contagion from spreading to the entire financial network.

There are some critical limitations of this study and the scope of future research. First, it considers only the listed financial institutions. Secondly, it is susceptible to window size and frequency of the return, whether daily, weekly, or monthly and significance level of Granger-causality. Future research can explore the more extensive dataset using firm-related micro factors and economy-related macro factors along with the PageRank measure. The PageRank scores can also be compared with other centrality scores like Eigenvector centrality to test the "too-central-to-fail" hypothesis.

**Conflict of Interest**
The authors confirm that there is no conflict of interest to declare for this publication.

**Acknowlwdgements**
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors would like to thank the editor and anonymous reviewers for their comments that help improve the quality of this work. The authors would also like to thank the session chairs and participants of ICMTEA-2021 for providing valuable feedback and suggestions.

## References

Acemoglu, D., Ozdaglar, A., & Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *American Economic Review*, *105*(2), 564-608.

Adrian, T., & Brunnermeier, M.K. (2011). *CoVaR* (No. w17454). National Bureau of Economic Research.

Allen, F., & Gale, D. (2000). Financial contagion. *Journal of Political Economy*, *108*(1), 1-33.

Battiston, S., Puliga, M., Kaushik, R., Tasca, P., & Caldarelli, G. (2012). Debtrank: Too central to fail? financial networks, the fed and systemic risk. *Scientific Reports*, *2*(1), 1-6.

Billio, M., Getmansky, M., Lo, A.W., & Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, *104*(3), 535-559.

Brownlees, C.T., & Engle, R. (2012). Volatility, correlation and tails for systemic risk measurement. *Available at SSRN*, *1611229*.

Brunnermeier, M.K., & Oehmke, M. (2013). Bubbles, financial crises, and systemic risk. *Handbook of the Economics of Finance*, *2*, 1221-1288.

Brunnermeier, M.K., & Pedersen, L.H. (2009). Market liquidity and funding liquidity. *The Review of FInancial Studies*, *22*(6), 2201-2238.

Chan-Lau, J.A. (2018). Systemic centrality and systemic communities in financial networks. *Quantitative Finance and Economics*, *2*(2), 468-496.

Chan-Lau, J., Espinosa, M., & Solé, J. (2009). Co-risk measures to assess systemic financial linkages. *IMFworking*, *2*, 38.

Danielsson, J., & De Vries, C. G. (2000). Value-at-risk and extreme returns. *Annales d'Economie et de Statistique*, 239-270.

Dasgupta, A. (2004). Financial contagion through capital connections: A model of the origin and spread of bank panics. *Journal of the European Economic Association*, *2*(6), 1049-1084.

Dickey, D.A., & Fuller, W.A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: journal of the Econometric Society*, 1057-1072.

Fong, T.P.W., Sze, A.K.W., & Ho, E.H.C. (2021). Assessing cross-border interconnectedness between shadow banking systems. *Journal of International Money and Finance*, *110*, 102278.

Freixas, X., Parigi, B.M., & Rochet, J.C. (2000). Systemic risk, interbank relations, and liquidity provision by the central bank. *Journal of Money, Credit and Banking*, 611-638.

Gray, D.F., Merton, R.C., & Bodie, Z. (2007). New framework for measuring and managing macrofinancial risk and financial stability.DOI 10.3386/w13607.

Härdle, W.K., Wang, W., & Yu, L. (2016). Tenet: Tail-event driven network risk. *Journal of Econometrics*, *192*(2), 499-513.

Hartmann, P., Straetmans, S., & Vries, C.D. (2004). Asset market linkages in crisis periods. *Review of Economics and Statistics*, *86*(1), 313-326.

Hautsch, N., Schaumburg, J., & Schienle, M. (2015). Financial network systemic risk contributions. *Review of Finance*, *19*(2), 685-73.

Huang, X., Zhou, H., & Zhu, H. (2009). A framework for assessing the systemic risk of major financial institutions. *Journal of Banking & Finance*, *33*(11), 2036-2049.

Jin, L. (2020). Risks' cross-contagion between shadow banks based on complex network theory: evidence from China. *Kybernetes*. *50*(6), 1900-1918. https://doi.org/10.1108/K-04-2020-0236.

Kendall, M.G. (1938). A new measure of rank correlation. *Biometrika*, *30*(1/2), 81-93.

Kuzubaş, T. Ömercikoğlu, I., & Saltoğlu, B. (2014). Network centrality measures and systemic risk: An application to the Turkish financial crisis. *Physica A: Statistical Mechanics and its Applications*, *405*, 203-215.

Lai, Y., & Hu, Y. (2021). A study of systemic risk of global stock markets under COVID-19 based on complex financial networks. *Physica A: Statistical Mechanics and its Applications*, *566*, 125613.

Lehar, A. (2005). Measuring systemic risk: A risk management approach. *Journal of Banking & Finance*, *29*(10), 2577-2603.

Mensah, J.O., & Premaratne, G. (2017). Systemic interconnectedness among Asian banks. *Japan and the World Economy*, *41*, 17-33.

Minoiu, C., Kang, C., Subrahmanian, V.S., & Berea, A. (2015). Does financial connectedness predict crises? *Quantitative Finance*, *15*(4), 607-624.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*. Stanford InfoLab.

Pedersen, L.H., Acharya, V., Philippon, T., & Richardson, M. (2010). Measuring systemic risk. *NYU Working Paper*.

Phillips, P.C., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75*(2), 335-346.

Rajan, R.G. (2009). Too systemic to fail: consequences, causes and potential remedies. *Written Statement to the Senate Banking Committee Hearings*, *6*, 8.

Thurner, S., & Poledna, S. (2013). DebtRank-transparency: Controlling systemic risk in financial networks. *Scientific Reports*, *3*(1), 1-7.

Wang, D., & Huang, W.Q. (2021). Centrality-based measures of financial institutions' systemic importance: A tail dependence network view. *Physica A: Statistical Mechanics and its Applications*, *562*, 125345.

Xu, Y., & Corbett, J. (2020). What a network measure can tell us about financial interconnectedness and output volatility. *Journal of the Japanese and International Economies*, *58*, 101105.

Yun, T.S., Jeong, D., & Park, S. (2019). "Too central to fail" systemic risk measure using PageRank algorithm. *Journal of Economic Behavior & Organization*, *162*, 251-272.

Zhang, D., & Broadstock, D.C. (2020). Global financial crisis and rising connectedness in the international commodity markets. *International Review of Financial Analysis*, *68*, 101239.

## Notes

[1]All you wanted to know about NBFCs. https://www.rbi.org.in/Scripts/FAQView.aspx?Id=92 (Accessed on 30 Oct 2021).

[2]Non-Banking Finance Companies: The Changing Landscape www.pwc.in Contents 2016. (Accessed on 30 Oct 2021).

[3]Sonavane, R.N. (2021, Sep 24). NBFCs lose over Rs 75,000 crore in 2 sessions on liquidity fears. https://www.livemint.com/Money/ANLzO6NSMNjuzZcCqY2nSN/NBFC-stocks-fall-further-on-liquidity-fears.html (Accessed on Oct 30, 2021).

[4]Panda, S. (2021, Oct 22). Scale-based regulatory framework for NBFCs optimal approach: RBI Deputy Guv. https://www.business-standard.com/article/economy-policy/scale-based-regulatory-framework-for-nbfcs-optimal-approach-rbi-deputy-guv-121102200704_1.html (Accessed on Nov 11, 2021).

[5]Gopakumar, G. (2020, Sep 16). Vulnerability of NBFC sector still a concern: RBI governor. https://www.livemint.com/industry/banking/vulnerability-of-nbfc-sector-still-a-concern-rbi-governor-11600241941061.html#box_11600241941061 (Accessed on Nov 11, 2021).

Original content of this work is copyright © International Journal of Mathematical, Engineering and Management Sciences. Uses under the Creative Commons Attribution 4.0 International (CC BY 4.0) license at https://creativecommons.org/licenses/by/4.0/



View publication stats

*2022 3rd International Conference for Emerging Technology (INCET)*
*Belgaum, India. May 27-29, 2022*

# Exploring Multimodal Sentiment Analysis through Cartesian Product approach using BERT Embeddings and ResNet-50 encodings and comparing performance with pre-existing models

Dr. Aruna Bhat
*Associate Professor, Department of CSE*
*Delhi Technological University*,
Delhi, India
aruna.bhat@dtu.ac.in

Rahul Mahar
*Delhi Technological University, CSE*
Delhi, India
rahulmahar_2k18co275@dtu.ac.in

Rahul Punia
*Delhi Technological University, CSE*
Delhi, India
rahulpunia_2k18co276@dtu.ac.in

Rahul Srivastava
*Delhi Technological University, CSE*
Delhi, India
rahulsrivastava_2k18co277@dtu.ac.in

2022 3rd International Conference for Emerging Technology (INCET) | 978-1-6654-9499-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/INCET54531.2022.9825245

***Abstract* - Multimodal Sentiment Analysis has become a popular interest of research in Machine Learning. It not only provides better results and context inference but also provides a good alternative for uni-modal analysis. The past few years have seen many new datasets for this task and state-of-the-art models dealing with the intermodular and intramodular dynamics of the task. This paper aims at exploring the pre-existing research on the topic of Multimodal Sentiment analysis and combining it with the idea of Sentence Embedding and Video encoding using state-of-the-art models, using the technique of classifying cartesian-product of extracted features, discussing experiments, and doing the qualitative analysis of the results.**

***Keywords - Multimodal Sentiment Analysis, Tensor Fusion Network, BERT, ResNet-50, Cartesian Product***

## I. INTRODUCTION

Multimodal Sentiment Analysis is a somewhat new field that has emerged in recent years but is gradually picking up speed as more and more data is being collected and research giants trying their hands on this task as well. The idea of multimodal sentiment analysis is somewhat similar to sentiment analysis, the difference being combining multiple modes of data (corresponding to the same instance in a dataset), extracting their features and using transformations on these features to improve the feature richness of input to a classification model. Over the past few years, many approaches have been developed, few of which will be discussed in this paper, specifically aiming at improving the accuracy and metric scores for this task. Some researchers have exploited the fact that the parent task i.e Sentiment Analysis has seen some remarkably good models in their respective domain of mode, and can be used either as feature extractors or combined (transfer learning) to create much more powerful and multi-domain models. On the contrary, some papers have also taken the path of treating it as a completely different domain, and not an extension of Sentiment Analysis, and have achieved state-of-the-art results as well. Our focus in this paper would be to follow the first approach and use BERT and ResNet-50 for feature extraction to form a transformed input for the classification model and determine whether it results in better scores than uni-modular classification.

## II. LITERATURE REVIEW

### A. BERT and Idea of Sentence Embedding

BERT[11] stands for Bidirectional Encoder Representations from Transformer and is one of the most successful language models developed by Google Research back in 2018. It uses Transformers [13], an attention mechanism that learns contextual relations between tokens (or sub-words) in a text.

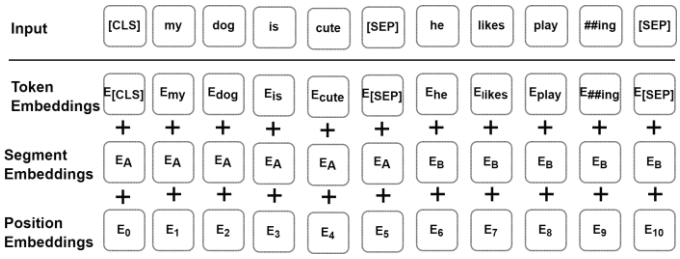



Fig. 1. BERT sequence representation.

It was the first model to use the method of pre-training with a pseudo task to identify masked input tokens. As opposed to directional models (like RNN, LSTM [15], GRU [19]), the Transformer [13] encoder which forms the heart of BERT [11] reads the entire sequence of words at once, removing and freeing the text from its directionality. Therefore it is considered bidirectional (or non-directional). This feature allows the model to learn the context of a word based on all of its surroundings (left and right of the word), hence making BERT [11] an excellent Sentence Encoder.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:22:36 UTC from IEEE Xplore. Restrictions apply.

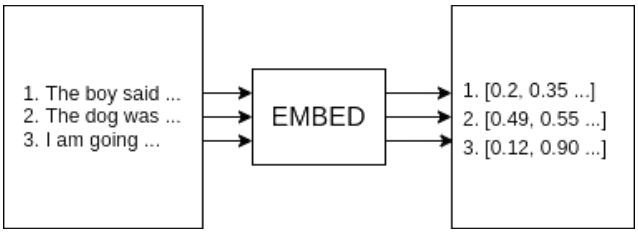



Fig. 2. Sentence Embedding

Sentence Embedding is the method of transforming a sentence or sequence of words into a sequence of numbers/floats with reduced or increased length. Sentence embedding has been explored since the beginning of Natural Language Processing, and several methods have hence been created. The most trivial is the Bag-Of-Words model with Count Vectorization, used generally with Naive Bayes Classifier. BERT[11] Encodes a sentence into a sequence of size (768,)

### B. *ResNet Image/Video Encoding*

ResNet-50[14] is the current state-of-the-art model for various computer vision tasks like Image Classification, Object Detection, Image Segmentation, etc. It consists of various stages consisting of convolution and identity blocks. This type of architecture makes the number of trainable parameters about 23 million.

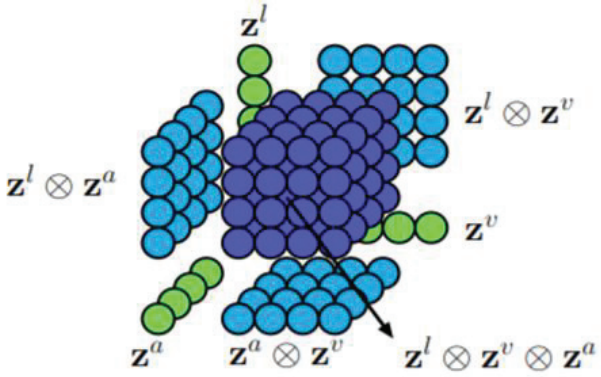



Fig. 3. ResNet-50 model

The reason for the state-of-the-art results of ResNet-50[14] is the exceptionally powerful feature extraction that is being performed in the intermediate layers and residual connections. This feature makes it ideal for feature extraction. Many models working in different domains of computer vision deploy ResNet-50 [14] as a feature extractor in the intermediate steps.

Video analysis is similar to image analysis in the way that video is an array of images, stacked w.r.t time at which they have to be shown. Hence, ResNet-50 [14] can be deployed to extract features from video data as well. The ResNet-50 [14] uses an average pooling layer, with an output size of (2048,), which is the layer used for feature extraction.

## III. RELATED WORK

Tensor Fusion Network (TFN) [1] was a novel approach introduced in 2017, which aimed at multimodal sentiment analysis using uni-modular techniques and merging them. The paper used multiple techniques for feature extraction like CNN [20] (for video features), LSTM [15] and CNN-LSTM (for text features), and Mel Spectrograms (for acoustic analysis), which were state-of-the-art at the time as uni-modular techniques.

### A. *Methods*

The TFN [1] works using 3 steps given below :

- LSTMs [15] and CNNs [20] are used to extract uni-modular features and output the encoded data.
- A 3-fold Cartesian product (tensor-product) of the embeddings generated above is calculated in order to pair every mode of data and this acts as input for the next step
- A simple neural network is trained with input as the cartesian product and sigmoid or softmax output is calculated, thus completing the classification pipeline

Broadly, tensor fusion model combines the following combinations of Audio(A), Video (V), and Text (T) :

$$A + T + V + A.T + V.T + A.T.V$$

Fig. 4. Tensor Fusion Network

The LSTMs (Fig 5) convert the input text sequence into a sequence of floats, essentially a form of Sentence Embedding discussed earlier using the equation :

$$\sim c_t = tanh(w_c[h_{t-1}, x_t] + b_c)$$

$$c_t = f_t * c_{t-1} + i_t * \sim c_t$$

$$h_t = o_t * tanh(c^t)$$

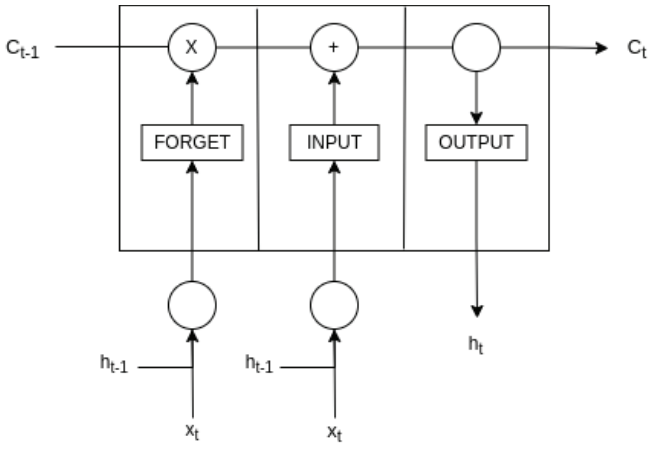



Fig. 5. An LSTM Cell

The TFN [1] uses 50-dimensional GloVe Embeddings [12] for word embeddings before being fed to the LSTM [15] network.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:22:36 UTC from IEEE Xplore. Restrictions apply.

### B. Results

The TFN [1] classifier achieved a binary class accuracy of 77.1% and F1 score of 77.9% whereas a 5 class classification accuracy of 42% .

## IV. DATASET

For the purpose of this project, we have chosen the CMU-MOSI[8] Multimodal Sentiment Dataset, which is the second-largest dataset for multimodal sentiment analysis prepared by Carnegie Mellon University, USA. These datasets have been used in the state-of-the-art models in the field of Multimodal Sentiment Analysis.

The Multimodal Corpus of Sentiment Intensity (CMU-MOSI[8]) dataset consists of 2199 opinions/reviews, in text, audio, and video format, and each review is assigned an intensity from the range [-3,3]. The dataset is segmented on the basis of the intensity score, making it suitable for 3 class classification as well as binary classification and emotion intensity prediction as well.

## V. PROPOSED MODEL

Inspired by the ideas grasped during the literature review and exploring related works, we came up with the pipeline/architecture shown in the figure :

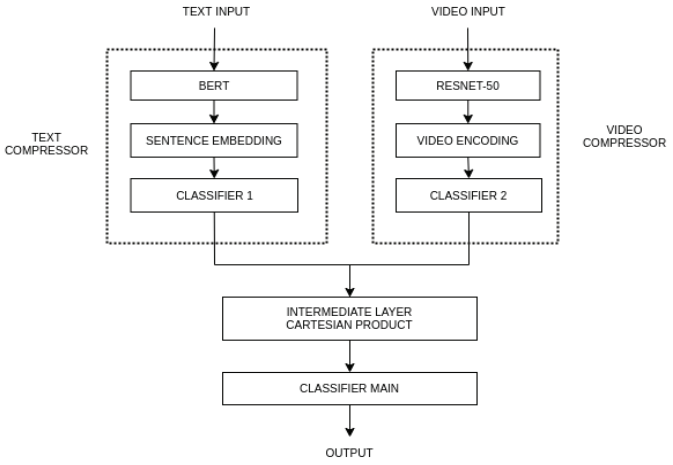



Fig. 6. Model Architecture

The model is somewhat inspired by the TFN [1] network discussed in the 'Related Works' section. The idea of sentence embedding using BERT [11] has been used here along with that of video encoding using ResNet-50 [14].

### A. Why Compressors?

The feature vector size of the BERT [11] encodings is (768,), while that of ResNet-50 [14] is (2048,). Upon calculations, we can show that the total size of variables and amount of memory that'll be used if a direct cartesian product is done without compressing is about 12.88 GB, which is more than what is allowed on open-source machine learning notebooks. Also, the training time would be 18 days per epoch. Hence, the requirement of a text and a video compressor was a must and has been successfully achieved by using sentence embedding and video encoding.

### B. Text Compressor Model

The Text Compressor used here is a simple multilayer perceptron classifier, that generates the first intermediate output for cartesian products. The input layer has shape (768,) and takes in the BERT [11] sentence embedding as input. Dense layers with 100 and 50 neurons have been added having ReLU [18] activation. The model contains 81,950 parameters.

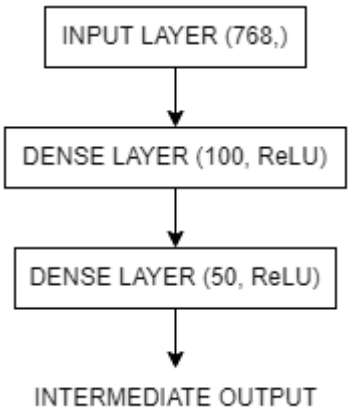



Fig. 7. Text Compressor Architecture

### C. Video Compressor Model

Proceeding on similar lines as that of text compressor, we developed the video compressor model by exploiting the ResNet-50 [14] model's feature extraction capability. However, some more concepts of video processing were also used which will be discussed here.

CMU-MOSI[8] dataset features videos at 32 frames-per-second (fps), which essentially means 32 images stacked in a second. To encode video into a sequence of numbers, ResNet-50 [14] was used for each frame(image) of the sliding window of size 16 (to account for videos less than 1 second). The combined output was calculated by taking the average of the sliding window because each video described a specific sentiment, with little or no variations in the visual features. Hence taking the average of the window results in a smoother feature map rather than taking maximum, which may combine different features present in different frames, resulting in a false feature map.

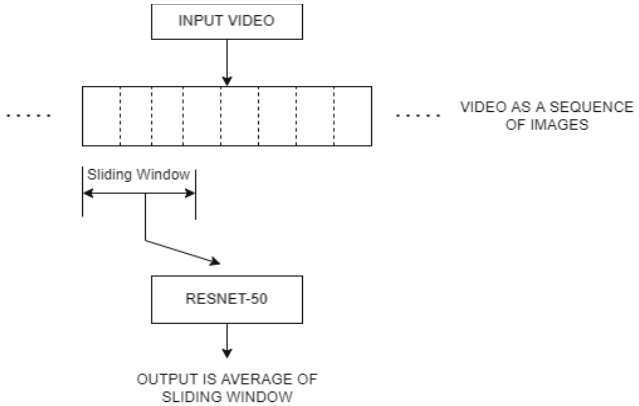



Fig. 8. Sliding Window Video Encoding



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:22:36 UTC from IEEE Xplore. Restrictions apply.

Following the above method, the video compressor model was created that takes in the averaged output of 16 frames and uses a multilayer classifier similar to the text compressor. The model contains an Input layer that accepts tensors of shape (2048,) and takes in the encoded video as input. Dense layers having 150 and 50 neurons respectively are added having ReLU [18] activation. The total trainable parameters are 314,900.

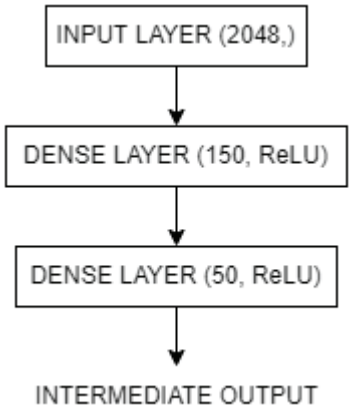



Fig. 9. Video Compressor Architecture

### *D. Main Classifier Model*

The main classifier is used to combine the intermediate outputs of the compressors by using a cartesian product and produces sigmoid output for binary classification. It accepts the flattened cartesian product of the compressor outputs, i.e. of shape (2500,), and subsequent dense layers having 200, 150, and 50 neurons, used with ReLU [18] activation. The total number of parameters for the entire pipeline becomes 922,251.

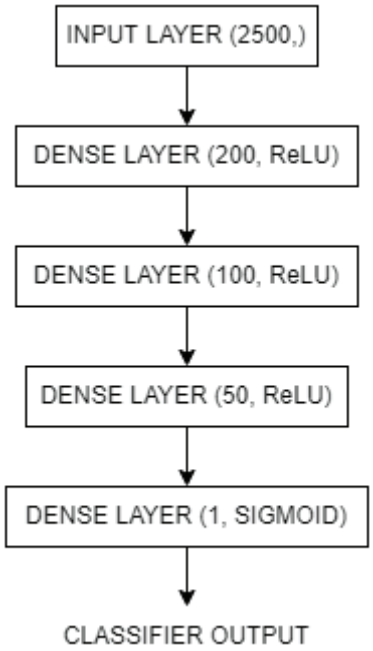



Fig. 10. Combined Model Architecture

## VI. EXPERIMENTS AND RESULTS

The CMU-MOSI[8] Dataset contains emotions in the form of real numbers from the range [-3,3] as discussed earlier, representing the intensity and type of emotion. As far as intensity is concerned, we can manually annotate the dataset into 2 class labels, using the logic that positive intensity can be labeled as a positive class (1) and negative labels can be labeled as a negative class (0), although setting different thresholds can result in different data suitable for different types of classification . Thus we focus only on binary classification results for this paper

The binary classification validation accuracy of BERT[11] on the CMU-MOSI[8] dataset was 44.5%, with the learning rate being 3e-5 and epsilon being 1e-8, while the accuracy of the text compressor model came out to be 64%. The standalone video compressor model achieved a validation accuracy of 57% with learning rate 3e-5 on CMU-MOSI[8] dataset. Adam optimizer [17] was used for all these experiments and the results correspond to binary classification only.

Followed by standalone model training, the combined classifier (proposed model) was trained with a learning rate of using the CMU-MOSI[8] dataset. Adam optimizer [17] was used for training the model with variable learning rate and model evaluation was done for 200 randomly selected examples of the dataset after every epoch. Confusion Matrix and Accuracy plots were plotted to calculate F1 score and track the progress of the training respectively.

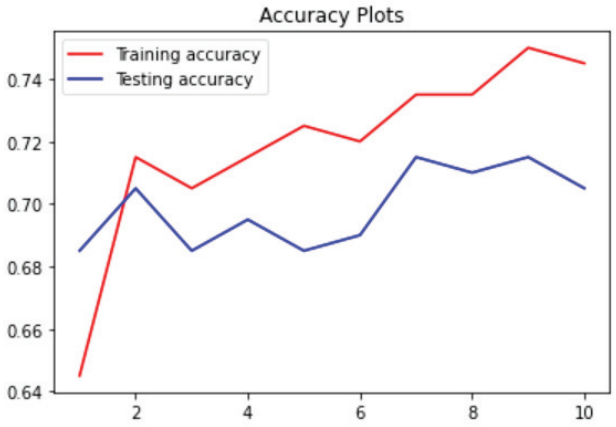



Fig. 11. Accuracy Plots for Training

The results are as follows in comparison to TFN[1] and other models :

TABLE I. COMPARISON OF TFN WITH ITS SUBTENSORVARIANTS AND OUR PROPOSED MODEL. ALL THE UNIMODAL, BIMODAL AND TRIMODAL SUBTENSORS ARE IMPORTANT.

| **Baseline** | **Binary Acc(%)** | **Binary F1** |
|---|---|---|
| $TFN_{language}$ | 74.8 | 75.6 |
| $TFN_{visual}$ | 66.8 | 70.4 |
| $TFN_{acoustic}$ | 65.1 | 67.3 |



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:22:36 UTC from IEEE Xplore. Restrictions apply.

| | | |
|---|---|---|
| $TFN_{bimodal}$ | 75.2 | 76.0 |
| $TFN_{trimodal}$ | 74.5 | 75.0 |
| $TFN_{notrimodal}$ | 75.3 | 76.2 |
| TFN | 77.1 | 77.9 |
| **Our Model** | **71.5** | **71.6** |

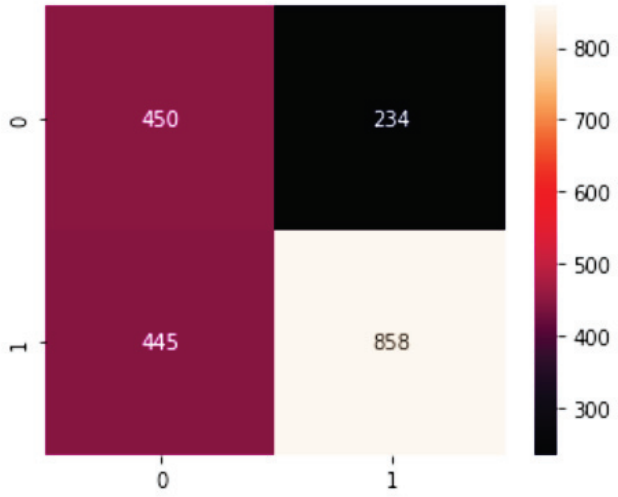



Fig. 11. Confusion Matrix for proposed model

## VII. QUALITATIVE ANALYSIS OF RESULTS

The results, although not state-of-the-art, offer great insight into how the feature extractors perform in case the size of the dataset is limited.

The reason for the low scores of our model and also for the state-of-the-art BERT[11] and ResNet-50[14] could be:

- Quality of dataset: The dataset is not rich in features when it comes to the Linguistic and Visual part, with some examples containing only one word and some videos of less than one second. The poor performance of BERT[11] and ResNet-50[14] supports this fact.
- HyperParameter Tuning: This is one of the most important steps when building a machine learning model. Learning rate, epsilon, number of layers, number of neurons, and respective activation functions need to be carefully chosen for better results
- Loss of Inference due to Compressors and Classifier: Due to improper hyper-parameter tuning, there is a possibility that the features of the input vector/ tensor may not be used to their full extent, and some features might be lost in the process of dimension reduction. Similarly, some information may have been lost during the process of video encoding, as averaging the feature vectors of each frame doesn't guarantee a perfectly encoded representation of the video.

## VIII. CONCLUSION AND FUTURE WORK

Our experiments have produced results that confirm the idea of cartesian-product as input produces better results than individual models, which was evident from TFN[1]. The qualitative analysis shows that results can be improved significantly by carefully adjusting the hyperparameters and using larger and more complex datasets. Future work/steps include hyperparameter tuning, using a different approach than cartesian product to transform the features before being fed to the classifier and treating the task of multimodal sentiment analysis as standalone and not an extension of sentiment analysis. Hence a different approach altogether may be devised specifically designed for multimodal sentiment analysis.

Our model consists of 922,251 parameters and occupies approximately 7MB of memory space, which is space-efficient and can be used in case a feature vector cartesian-product model has to be implemented with results nearly approaching TFN[1] levels.

## REFERENCES


[1] Amir Zadeh, Minghai Chen, Soujanya Poria, Erik Cambria, Louis-Philippe Morency Tensor Fusion Network for Multimodal Sentiment Analysis (EMNLP) 2017

[2] Taeyong Kim, Bowon Lee Multi-Attention Multimodal Sentiment Analysis ICMR '20: International Conference on Multimedia Retrieval

[3] Wenmeng Yu, Hua Xu, Ziqi Yuan, Jiele Wu Learning Modality-Specific Representations with Self-Supervised Multi-Task Learning for Multimodal Sentiment Analysis arXiv:2102.04830v1 [cs.CL] 9 Feb 2021

[4] Wasifur Rahman1, Md. Kamrul Hasan1*, Sangwu Lee1*, Amir Zadeh2, Chengfeng Mao2, Louis-Philippe Morency2, Ehsan Hoque Integrating Multimodal Information in Large Pretrained Transformers published in ACL 2020

[5] Amir Zadeh, Rowan Zellers, Eli Pincus, and Louis- Philippe Morency. 2016b. Multimodal sentiment in- tensity analysis in videos: Facial gestures and verbal messages. IEEE Intelligent Systems 31(6):82–88.

[6] María Teresa García-Ordás, Héctor Alaiz-Moretón, José Alberto Benítez-Andrades, Isaías García-Rodríguez, Oscar García-Olalla, Carmen Benavides,Sentiment analysis in non-fixed length audios using a Fully Convolutional Neural Network,Biomedical Signal Processing and Control,Volume 69,2021,102946,ISSN 1746-8094.

[7] Sherstinsky, Alex. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. Physica D: Nonlinear Phenomena. 404. 132306. 10.1016/j.physd.2019.132306.

[8] MOSI: Multimodal Corpus of Sentiment Intensity and Subjectivity Analysis in Online Opinion Videos Amir Zadeh, Rowan Zellers, Eli Pincus, Louis-Philippe Morency, arXiv:1606.06259v2, IEEE Intelligent Systems 31.6 (2016): 82-88

[9] TEASEL: A transformer-based Speech-Prefixed Language Model Mehdi Arjmand and Mohammad Javad Dousti and Hadi Moradi, arXiv:2109.05522, 2021

[10] Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013, October). Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 conference on empirical methods in natural language processing (pp. 1631-1642).

[11] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[12] Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).




Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:22:36 UTC from IEEE Xplore. Restrictions apply.

[13] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun (2015), Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385.

[15] S. Hochreiter, J. Schmidhuber, Long short-term memory. Neural Comput, 9 (1997), pp. 1735-1780

[16] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. (2011). Learning Word Vectors for Sentiment Analysis. The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011).

[17] Diederik P. Kingma, Jimmy Ba, Adam: A Method for Stochastic Optimization (2014), Published as a conference paper at ICLR 2015.

[18] Abien Fred Agarap,Deep Learning using Rectified Linear Units (ReLU),2019, arXiv:1803.08375

[19] Chung, Junyoung, et al. "Empirical evaluation of gated recurrent neural networks on sequence modeling." arXiv preprint arXiv:1412.3555 (2014).

[20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12). Curran Associates Inc., Red Hook, NY, USA, 1097–1105.

[21] M. Ahmed, M. Goel, R. Kumar and A. Bhat, "Sentiment Analysis on Twitter using Ordinal Regression," 2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), 2021, pp. 1-4, doi: 10.1109/SMARTGENCON51891.2021.9645751.

[22] Isaac, Akileng & Bhat, Aruna. (2022). A Conceptual Enhancement of LSTM Using Knowledge Distillation for Hate Speech Detection. 10.1007/978-981-16-4016-2_53.

[23] R. Kumar and A. Bhat, "An Analysis On Sarcasm Detection Over Twitter During COVID-19," 2021 2nd International Conference for Emerging Technology (INCET), 2021, pp. 1-6, doi: 10.1109/INCET51464.2021.9456392.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:22:36 UTC from IEEE Xplore. Restrictions apply.

Journal of Energy Storage 54 (2022) 105318

Contents lists available at ScienceDirect

Journal of Energy Storage

journal homepage: www.elsevier.com/locate/est

Research papers

# Fabrication and investigation on influence of metal oxide nanoparticles on thermal, flammability and UV characteristics of polyethylene glycol based phase change materials

Surya Tanwar, Raminder Kaur *

*Discipline of Polymer Science and Chemical Technology, Department of Applied Chemistry, Delhi Technological University (Formerly Delhi College of Engineering), Delhi 110042, India*

ARTICLE INFO

*Keywords:*
Phase change materials
Form-stable
Polyvinyl alcohol
Polyethylene glycol
Thermal energy storage

ABSTRACT

The current study is aimed at the fabrication and property evaluation of form-stable phase change material (PCMs) nanocomposites films made by solvent casting method using Polyethylene Glycol (PEG) as working PCM and Polyvinyl Alcohol (PVA) as a supporting matrix, reinforced with Titanium Dioxide Nanoparticles (NTO). The influence of the incorporated metal oxide nanoparticles in varying concentrations (0.25–1 %) on the chemical, thermal, microstructural, UV resistance, and flammability characteristics for the developed nanocomposites has been thoroughly investigated. The resultant form-stable PCMs are characterized by FTIR and FE-SEM. Phase change behaviours and thermal stability of the prepared PCMs are determined by DSC and TGA analysis. T-History and thermal cycling tests have been carried out to access the charging-discharging time as well as the thermal reliability of the form-stable PCMs. The addition of the varying loadings of NTO particles to the optimized PEG/PVA composition significantly improves the phase change parameters, such as latent heat storage capacities from 65.59 J/g to 73.02 J/g and degree of supercooling from 18.93 to 17.03 respectively; along with the thermal stability of the nanocomposites with increasing order of nanoparticles concentrations. The presence of NTO particles in nanocomposites has proved to be effective in successfully reducing the UV transmittance across the films. The metal oxide filler also enhances the flame-retardant behaviour of the form-stable PCMs.

## 1. Introduction

With the rise of the industrial revolution in the 20th century to the present day, the world has witnessed tremendous development in society. The significant uplift in the basic needs and living standards of the people leads to enormous energy demands, which have been derived from coal and petroleum resources. Unfortunately, these non-renewable resources are exhausting gradually and may not be able to serve the energy requirements for longer periods. Simultaneously, the excessive dependence on the use of these nature of resources has also resulted in global warming due to greenhouse gas emissions [1]. The research community across the globe, advocates for the utilization of the available renewable and sustainable energy sources such as solar and wind energy, etc. However, the intermittent and time-dependent nature of these energy sources pushes researchers to look for more efficient energy storage technologies [2].

To sustain day-to-day life, a variety of basic commodities products/goods are the vital obligations that include fresh food items, medicinal products, electrical or electronic goods, etc. A crucial factor that can strongly affect the quality or deteriorate the inherent product properties is the temperature variations during storage, transportation, and distribution chain until the commodities/goods are delivered to the end-

*Abbreviations:* BJH, Barrett–Joyner–Halenda; BET, Brunauett–Emmett–Teller; CA, Cellulose Acetate; DSC, Differential Scanning Calorimetry; EDX, Energy Dispersive X-ray Spectroscopy; EVA, Ethylene-Vinyl Acetate; EG, Expanded Graphite; FESEM, Field-Emission Scanning Electron Microscopy; PCCFMs, Form-Stable Phase Change Composite Fibrous Membranes; FTIR, Fourier Transform Infrared Spectroscopy; GNP, Graphene Nanoplatelets; HDPE, High-Density Polyethylene; HR-TEM, High-Resolution Transmission Electron Microscopy; LOI, Limiting Oxygen Index; LDPE, Low Density Polyethylene; LLDPE, Linear-Low Density Polyethylene; MF, Melamine-Formaldehyde; MWCNTs, Multi-Walled Carbon Nanotubes; NTO, Nano-Titanium Dioxide; PA, Palmitic Acid; PCMs, Phase Change Materials; PAN, Polyacrylonitrile; PA, Polyamide; PEGs, Polyethylene Glycols; PP, Polypropylene; PU, Polyurethane; PVA, Polyvinyl Alcohol; PVC, Polyvinyl Chloride; SAED, Selected Area Electron Diffraction; SAL, Stearyl Alcohol; TGA, Thermogravimetric Analysis; UV, Ultraviolet; WPCs, Wood Plastic Composites; XRD, X-Ray Diffraction.

* Corresponding author.
*E-mail address:* raminderkaur@dce.ac.in (R. Kaur).

https://doi.org/10.1016/j.est.2022.105318
Received 4 February 2022; Received in revised form 5 June 2022; Accepted 13 July 2022
Available online 22 July 2022
2352-152X/© 2022 Elsevier Ltd. All rights reserved.

user [3]. Many efforts and huge monetary value of manufacturers and logistic providers come into play to maintain the required temperature and preserve the product's temperature but an undesirable spike in product temperature leads to premature failure of many temperature-sensitive goods. One handsome approach in the direction to limit the undesirable temperature surge and ensure the quality of the goods is the incorporation of Phase Change Materials (PCMs) within the base packaging materials to control the temperature within acceptable limits [4].

Phase change materials (PCMs) are well known for temperature control and thermal energy storage in the form of latent heat energy. These materials absorb and release a good amount of latent heat energy during the phase change, within a narrow temperature range [5,6]. Potential applications for PCMs employment include energy storage building structures, thermo-regulating textiles and greenhouses, heat regulation of electronics, telecommunications, coatings, transportation packaging of temperature-sensitive products, etc. [3,7–9]. A wide range of PCMs which undergo solid-liquid phase change during operation namely, salt hydrates, paraffins, sugar alcohols, polyethylene glycol, fatty acids, and their ester derivatives have been explored in the past decades for thermal energy storage application [10]. However, these solid-liquid PCMs possess some drawbacks such as large volume change during phase transitions, leakage issues, requires containment during thermal energy storage, low thermal stability (except inorganic PCMs-salt hydrates), etc.

To resolve the handling and working issues of solid-liquid PCMs, Shape-Stabilized PCMs or Form-Stable PCMs technology has been introduced into the market. Form-Stable PCMs are the combination of active energy storage PCMs and a supporting matrix of inorganic or organic nature. Working methodology of the form-stable PCMs can be understood in a manner as in when the temperature reaches the phase change temperature of the in-housed PCM, the active latent heat energy storage component (PCMs) undergoes melting accompanied by a change of physical state from solid to liquid while the supporting matrix forming the form-stable PCMs can remain in solid state [11].

Much research has been done focussing on the development of variety of Form-Stable PCMs by employing different supporting matrices such as High-Density Polyethylene (HDPE) [12–14], Low- Density Polyethylene (LDPE) [15], Linear-Low Density Polyethylene (LLDPE) [16], Ethylene-Vinyl Acetate (EVA) [17], Polypropylene (PP) [18], Polyacrylonitrile (PAN) [19], Polyvinyl Chloride [20], Polyvinyl Alcohol (PVA) [21], Styrene Polymers [22], Polyamide (PA) [23], Polyurethane (PU) [24], Cellulose Acetate (CA) [25] etc.

Cao et al. [26] fabricated Wood Plastic Composites (WPCs) based form stable PCMs by mixing Dodecanol filled Melamine-Formaldehyde (MF) Microcapsules with wood flour/HDPE supporting matrix. Prepared WPCs exhibited phase change around 22.2–28.7 °C with associated enthalpies of fusion and crystallization as 20.3 J/g and 31.6 J/g respectively. The presence of MF microcapsules in the HDPE matrix of WPCs caused a decrease in the tensile and flexural strength of WPCs, on contrary; the PCM microcapsules boosted the impact strength and surface hardness of the final product. Fang et al. [27] studied the effect of the addition of Expanded Graphite (EG) into the form-stable PCMs composed of Stearyl Alcohol (SAL) as working PCMs and HDPE. It was found that the added EG turns out to be very effective in limiting the leakage of SAL during phase change, by acting as a co-matrix along with HDPE and, highly efficient in promoting the thermal conductivity of the final article, as only the 3 wt% EG dosage inside the HDPE matrix rose its thermal conductivity to 0.6698 W (m K)$^{-1}$from 0.1966 W (m K)$^{-1}$.

Fang et al. [28] prepared form-stable PCMs containing Palmitic Acid (PA) active PCM and HDPE supporting matrix filled with Graphene Nanoplatelets (GNP). It was found that the presence of GNP at 4 wt% led to 2.5-fold increase in the thermal conductivity as compared to unfilled HDPE matrix and showed phase change temperature around 62 °C and latent heat enthalpy of 155.8 J/g. Ke et al. [29] used the electrospinning technique to produce Form-Stable Phase Change Composite Fibrous Membranes (PCCFMs) based on capric–palmitic–stearic acid ternary eutectic fatty acid mixture PCM and polyacrylonitrile matrix filled with COOH-functionalized multi-walled carbon nanotubes (COOH-MWCNTs). The prepared fibrous membranes showed phase change around 7–32 °C with associated latent heat enthalpies 130–138 kJ/kg, respectively.

Alkan et.al [30] have chosen Polyvinyl Alcohol (PVA) to hold inorganic salt hydrates such as sodium sulphate decahydrate ($Na_2SO_4.10H_2O$) and calcium chloride hexahydrate ($CaCl_2.6H_2O$) inside the interstices of hydrophilic polymer, to be used for passive thermal energy storage systems. Tang et al. [31] prepared Form-Stable Composite PCM made by incorporating Hexadecanol inside dye linked-Polyurethane (Dye-PU). Dye-PU is well effective in visible light absorption and its subsequent conversion to thermal energy. Also, the composite PCM exhibited high latent heat enthalpy of 229.5 J/g.

Among solid-liquid PCMs, Polyethylene Glycols (PEGs) have gathered the attention of a lots of researchers active in the field of thermal energy storage, and are proved to be potential latent heat energy storage components, due to certain attributes such as availability in a series of molecular weights with diverse melting temperatures ranging from 4.2 to 68.7 °C, tunable high latent heat enthalpy depending upon molecular weights, moderate chemical and thermal stability, congruent melting behaviour, no phase separation, biodegradable in nature, low vapor pressure, non-corrosive, non-toxic, etc. However, PEGs suffer from some shortcomings being leak-prone, have low thermal conductivity, low flammability and slight supercooling [32,33].

In the present study, the form-stable PCMs nanocomposite films with thermal buffering ability have been made by using PEG 1500 as an active latent heat energy storage component and Polyvinyl Alcohol (PVA) as a supporting matrix to control the leakage and accommodate large volume change of PEG during phase change, which is duly filled with Titanium Dioxide Nanoparticles. The influence of the incorporation of titanium dioxide nanoparticles on the thermal energy storage characteristics, thermal stability, flammability, and UV-blocking characteristics of PVA/PEG PCM films was studied in detail.

## 2. Materials and method

### 2.1. Materials

Titanium (IV) Isopropoxide (Avra Synthesis Pvt. Ltd., India), Isopropyl Alcohol (Merck, India), Concentrated Nitric Acid (Qualigens, India), Polyvinyl Alcohol (M.W.: 14000) (CDH, India), Polyethylene Glycol 1500 (Loba Chemie, India) and Milli-Q Water are used in this study.

### 2.2. Synthesis of titanium nanoparticles (NTO)

The Nano-$TiO_2$ (NTO) particles were synthesized by Sol-Gel Approach as follows: Titanium (IV) Isopropoxide precursor was dropwise added to Isopropyl Alcohol in a beaker under constant stirring for 30 min. The hydrolysis step was carried out by slow addition of Milli-Q water in a molar ratio (Ti:$H_2O$ = 1:4) to the above reaction mixture, with constant stirring for 2 h. Afterwards, a few drops of nitric acid were added to maintain the pH range (pH: 2–4) and arrest the extent of the hydrolysis process of titanate solution. The reaction mixture was constantly stirred for another 1 h to allow the formation of sols. Then, the above mixture was naturally aged for 24 h to allow gel formation. The obtained gels were filtered, water-washed and finally oven-dried to evaporate water and other organic impurities at 60 °C for 48 h. The resultant product was calcinated at 500 °C for 5 h to achieve Titanium Dioxide Nanoparticles (NTO). The scheme for the adopted experimental procedure for the synthesis of Titanium Dioxide Nanoparticles is represented in Fig. 1.

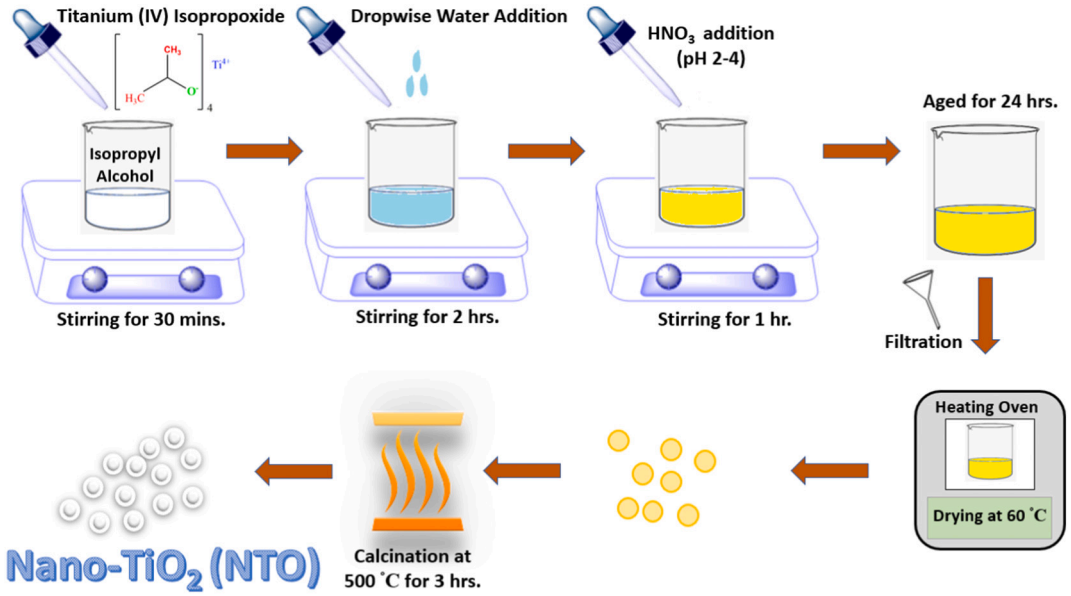



**Fig. 1.** Synthesis of Titanium Dioxide Nanoparticles via sol-gel route.

### 2.3. *Preparation of PVA/PEG/NTO form-stable nanocomposites films*

To achieve form-stable PCMs with high PCM content, it is important to deduce the suitable optimum concentration of PEG 1500 which the PVA matrix can be capable of easily holding and capturing the PCM into its macromolecular domain without any seepage. For this purpose, initially, PEG in different weight fractions (20 wt%, 30 wt%, 40 wt%, 50 wt%, 60 wt%, 70 wt%, 80 wt%) is added to the PVA matrix to form PVA-PEG blended films by the solvent-casting method. 5 % (w/v) of PVA aqueous solution was made by dissolving the desired amount of PVA powder in Milli-Q water under constant agitation (RPM: 1500) at 75 °C until a clear solution is obtained. To this solution, the PEGs in their respective weight fractions were added and further stirred for 5 h at 75 °C. After 5 h, the hot polymer solution was casted into the polypropylene petri-dish and left undisturbed for a few minutes for the removal of bubbles, and then subjected to drying in a hot-air oven at 30 °C for 24 h. The resultant film was peeled from the petri-dish and strongly pressed between the folds of filter paper, under load to squeeze out unreacted PEG present on the surface. To obtain nanocomposite films, a similar methodology was adopted as explained above, where the films were produced by adding the prepared Nano-$TiO_2$ (NTO) in different weight fractions (0.25 wt%, 0.50 wt%, 0.75 wt%, 1 wt%) to the

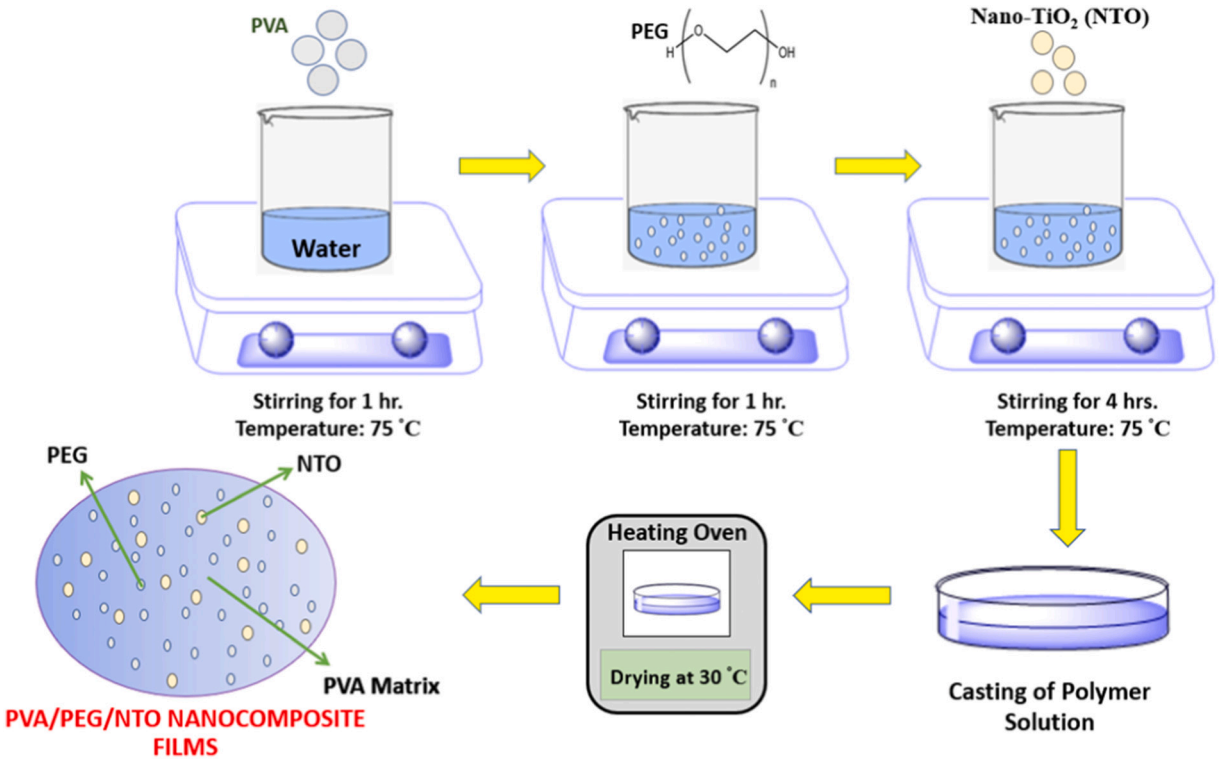



**Fig. 2.** A schematic representation of the experimental procedure to obtain form-stable NTO/PEG1500/PVA nanocomposites.

optimized PVA-PEG blend concentration, (which was deduced accordingly by leakage tests). A schematic representation of the adopted experimental procedure for the synthesis of PVA/PEG/NTO Nanocomposite Form-Stable Composite PCMs is shown in Fig. 2.

## 3. Characterization of PVA/PEG/NTO form-stable PCMs

The chemical structure and the inherent functional groups of form-stable PCMs and nanoparticles were investigated by Fourier Transform Infrared Spectrometer (Spectrum Two, Perkin Elmer, USA). The synthesized form-stable PCMs in a small quantity were mixed with KBr to form pellets and the spectrum was recorded at a resolution of 4 $cm^{-1}$, with 128 scans in the range of 450–4000 $cm^{-1}$. The analysis of the structural phases of the prepared Titanium Dioxide Nanoparticles was carried out using Powder X-Ray Diffractometer (PANalytical, X'Pert Pro, Netherlands) at a slow-scanning rate under 2θ range of 10–80° using fixed Ni-filtered Cu Ka radiation (λ = 1.54 Å). High-Resolution Transmission Electron Microscopy (HR-TEM) (Tecnai™ G2 20) was used to characterize the morphology and microstructure of the Titanium Dioxide Nanoparticles. Nitrogen adsorption-desorption isotherm at 77 K was carried out using MicrotracBEL (BELSORP, Japan). The surface area of the prepared nanoparticles was deduced by applying Brunauett–Emmett–Teller (BET) method. Barrett–Joyner–Halenda (BJH) model was used to evaluate the pore size distributions and total pore volume. Energy Dispersive Spectroscopy was used to determine the purity of the synthesized nanoparticles.

The morphology and distribution of nanoparticles within the formed nanocomposites were explored with the help of Field-Emission Scanning Electron Microscopy (Hitachi, SU8010, Japan), at an accelerated voltage of 5 kV, by smearing with gold coating and, examined under low vacuum pressure. For the evaluation of form stability, leakage test was conducted on films of PVA/PEG blends with varying weight fractions of PEGs (20–80 wt%), using two different methods, as described by different researchers [34,35]. In the first approach, a fixed weight of sample was placed on the filter papers and subjected to hot-air oven at 70 °C for 2 h. After removing from the oven, it was kept undisturbed to attain the ambient temperature. Further, it was transferred to a fresh filter paper and weighed again. The same procedure was repeated for all samples with varying weight fractions of PEGs (20–80 wt%). The leakage percent was calculated using the equation as follows:

$$L\% = (Mo - Mt)/Mo \times 100$$

where, Mo is the initial mass of the sample before heating; Mt. is the mass of the sample after oven heating.

In the alternative approach, to evaluate the leakage rate, the sample was placed on a Whatman filter paper, in the centre, within the diameter of 20 mm and subjected to heating inside a hot-air oven at 70 °C for 2 h. Afterwards, the PEGs traces on the filter paper were observed. The sample with less leakage percentage, when heated beyond PEGs phase change temperature, was considered as form-stable PCM. The leakage percent was deduced using the equation as follows:

$$L\% = (D - Do)/Do \times 100$$

where, Do is the diameter of the test area i.e. 20 mm; D is the diameter of the seepage area after the leakage test. The criteria of leakage are shown in the table given below (Table 1):

**Table 1**
Leakage criteria considered for as prepared PEG1500/PVA blends.

| Criteria of leakage | | |
|---|---|---|
| Leakage percentage (L%) | Phenomenon | Conclusion |
| L% ≤ 10 | Negligible or trace amounts | Highly stable |
| 10 < L% ≤ 30 | Very little leakage | Remained stable |
| 30 < L% ≤ 50 | Moderate leakage | Unstable |
| L% > 50 | Severe leakage | Highly unstable |

Differential Scanning Calorimetry (DSC 8000, Perkin Elmer, USA) was used to evaluate the potential of the form-stable PCMs as latent heat energy storage materials, by measuring their important thermal parameters, such as melting temperature (T*m*), crystallization temperature (Tc), and the enthalpies associated during both heating (ΔH*m*) and cooling (ΔHc) processes. Three samples of the prepared unfilled and nano-$TiO_2$ filled form-stable polymer films were analyzed for the determination of their heat storage attributes, and an average value of the obtained DSC results was reported. The analysis was carried out in a temperature sweep of 10 to 65 °C at a standard heating/cooling rate of 10 °C/min under a nitrogen blanket, at a constant flow rate of 40 ml/min. The thermal stability and the weight loss studies were conducted on the prepared titanium nanoparticles and form-stable PCMs films using a thermal analyzer (TGA 4000, Perkin Elmer, USA). About 4–5 mg of nanoparticles or form-stable PCM sample was placed in the silica crucible and was subjected to a temperature of 30 to 800 °C, at a standard heating rate of 10 °C/min, under nitrogen environment, at a constant flow rate of 20 ml/min.

Thermal Energy Storage Rate Test (T-History) was performed for the estimation of the charging and discharging time of the PEG and the form-stable PCMs. Each sample has been tested three times and the average value was reported. In this method, two temperature-controlled water baths were maintained at 10 °C (cold-bath) and 60 °C (hot-bath) respectively. Both the raw PCM and the form-stable PCMs sample films were placed inside cylindrical glass tubing and a K-type thermocouple was placed in the centre of the sample. The tubes were subjected to cold-bath maintained at 10 °C and after complete temperature saturation; the tubes were quickly transferred to a hot water bath operating at 60 °C, to study the heat-storage (charging rate) ability. Thereafter, achieving a constant equilibrating temperature, the tubes were again returned to the cold-bath to study the heat-release (discharging rate) ability. The measurements during this test were automatically recorded by a data acquisition system with an accuracy of ±0.2 °C at a regular time interval of 2 s. The thermal reliability of the form-stable PCMs was determined by the experimental procedure described elsewhere [36]. The form-stable PCMs sample was exposed to 100-thermal cycles followed by DSC analysis to investigate the changes in the thermal properties, post thermal cycling test.

UV–vis transmittance spectra of the nanocomposites were recorded within the wavelength range of 200–800 nm using a UV–vis spectrophotometer (Agilent Technologies, Carry Series). The flame retardant characteristics of the fabricated form-stable PCM nanocomposites were determined by the measured LOI (%) values using Oxygen Index Apparatus. (Fire Testing Technology, UK). Minimum 5 specimens of each nanocomposite composition were taken and clamped inside the sample holder. A concentration span of nitrogen gas ($N_2$) was fixed as 99 % whereas oxygen gas ($O_2$) was kept as 17.4 %. The gas flow rates of both $N_2$ and $O_2$ gases were maintained at 2 bars.

## 4. Results and discussions

I. Evaluation of synthesized Titanium Dioxide Nanoparticles (NTO)

### 4.1. Morphology of Nano-$TiO_2$ (NTO)

Fig. 3 shows HR-TEM images of Nano-$TiO_2$ (NTO), which revealed that the prepared nanoparticles possess almost spherical geometry and are present in a fair particle size range of 16–40 nm, with the mean particle size found to be 22 nm. Also, the Selected Area Electron Diffraction (SAED) pattern further supports the presence of Titanium Dioxide Nanoparticles having anatase type nature.

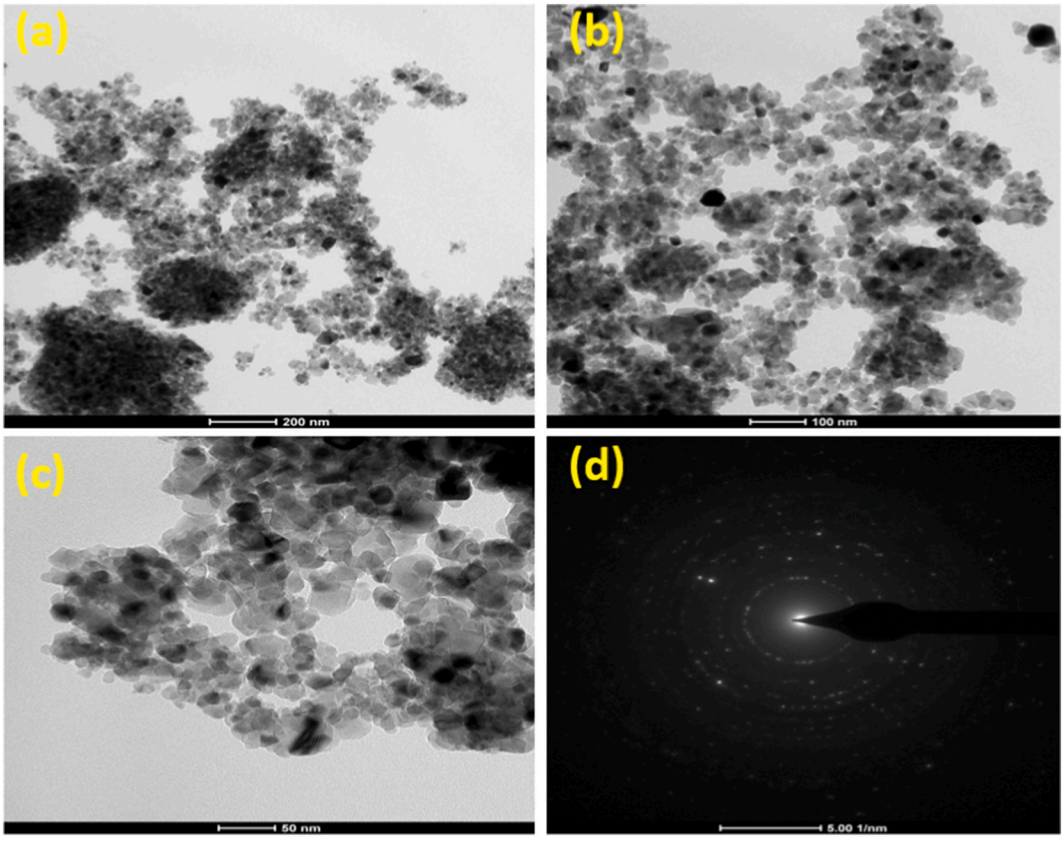



**Fig. 3.** High-resolution transmission electron microscope (HR-TEM) and Selected Area Electron Diffraction (SAED) images of the synthesized NTO.

### 4.2. X-Ray Diffraction (XRD) study of Nano-$TiO_2$

The diffraction pattern of prepared NTO using sol-gel method represented in Fig. 4 shows diffraction peaks at 2θ values of 25.3, 37.9, 48.1, 54.1, 55.2, 62.8, 69.0, 70.4, and 75.1 corresponding to crystal planes (101), (004), (200), (105), (211), (204), (116), (220) and (107) respectively, indicating the formation of anatase type of titanium nanoparticles in accordance to JCPDS File 21-1272 [37].

### 4.3. Fourier Transform Infrared (FTIR) Spectroscopy of Nano-$TiO_2$ (NTO)

Fig. 5 shows the FTIR spectra of Nano-Titanium Dioxide (NTO) synthesized by the sol-gel route observed in the range of 450–4000 $cm^{-1}$. The peaks at 3375 and 1623 $cm^{-1}$ correspond to the stretching and bending vibration of the —OH group because of the absorption of molecular water in the form of moisture, from the atmosphere. Few low-intensity peaks arising at 2896 and 2872 $cm^{-1}$ indicate the presence of

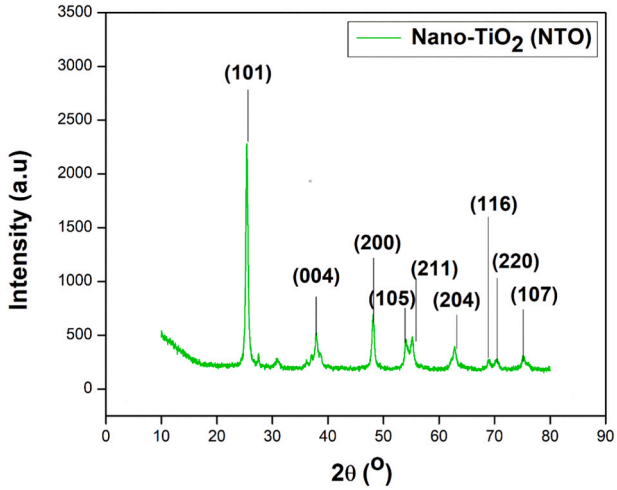



**Fig. 4.** X-Ray Diffraction (XRD) pattern of NTO particles.

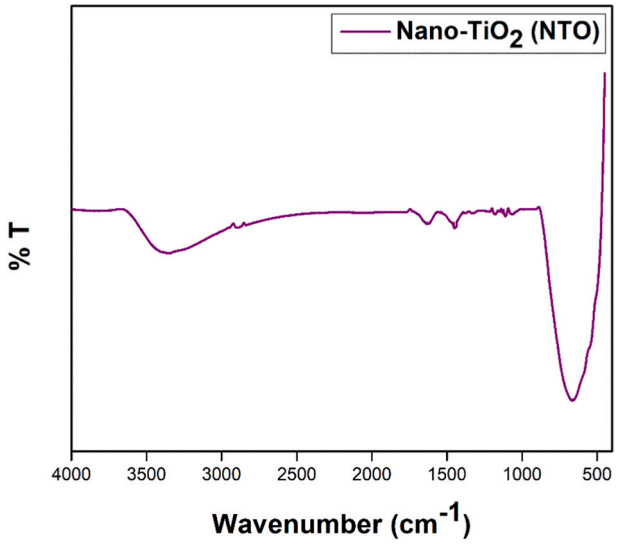



**Fig. 5.** FTIR spectra of Nano-Titanium Dioxide (NTO) derived by Sol-Gel Approach.

residual organic carbonaceous content of alkoxide precursors. The peaks occurring at 1448 and 664 $cm^{-1}$ are associated with the stretching vibrations of Ti—O—Ti and Ti—O planes of Titanium Dioxide Nanoparticles, respectively.

### 4.4. Thermal stability of Nano-$TiO_2$ (NTO)

Fig. 6 shows the thermal stability behaviour of prepared Nano-$TiO_2$ characterized by the TGA method carried out up to 800 °C. The Titanium Dioxide Nanoparticles showed an overall weight loss of 17.1 % when heated to 771.8 °C, thereby indicating the excellent thermal stability of the nanoparticles. The initial weight loss occurring at 130.0 °C is assumed to be the desorption of the water molecules which were absorbed from the ambient atmosphere over the nanoparticles surface. The second weight loss is observed between 174.6 and 369.9 °C, which can be considered as the degradation of the remnant carbonaceous impurity (arising from alkoxide moiety) present in trace concentration along with the Titanium Dioxide Nanoparticles, even after the water washing stage post-synthesis. This is also indicated by EDX analysis as shown in Fig. 8 in Section 4.6. The final weight loss is observed at 665.8 °C, which may correspond to the complete burning out of residual organic impurities from the interstices of the synthesized metal oxide nanoparticles.

### 4.5. Brunauer-Emmett-Teller (BET) studies of Nano-$TiO_2$ (NTO)

The nitrogen adsorption isotherm pattern of the synthesized Nano-$TiO_2$ (NTO) is shown in Fig. 7. The characterization of the samples including average surface area, average pore diameter, total pore volume is listed in Table 2. The BET measurements confirmed the absence of macropores in the prepared nanoparticles. The nanoparticles exhibited the Type IV isotherm pattern, (BDDT classification) with hysteresis loop of H1 nature inferring the presence of mesoporous structure. BJH (Barrete-Joynere-Halenda) method was utilized to measure the pore-size distribution (inset in Fig. 7). The BET surface area of the NTO was found to be 53.92 $m^2/g$. The pore diameter for NTO was observed as 10.236 nm. The mesoporous total pore volume was calculated and found to be 0.138 $cm^3/g$.

### 4.6. Energy-Dispersive X-ray (EDX) Spectra of Nano-$TiO_2$ (NTO)

Energy Dispersive X-ray (EDX) Spectra of the prepared Nano-$TiO_2$ (NTO), as shown in Fig. 8 indicates the peaks for Ti element and oxygen, which, suggests the chemical constituents of Ti and O present in the

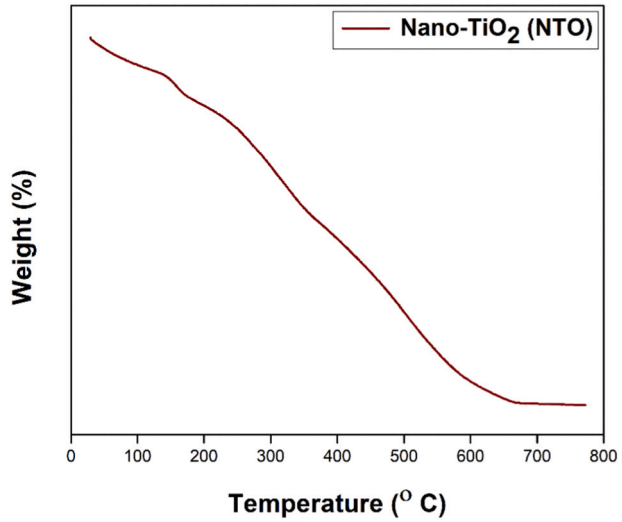



**Fig. 6.** TGA curve of NTO particles.

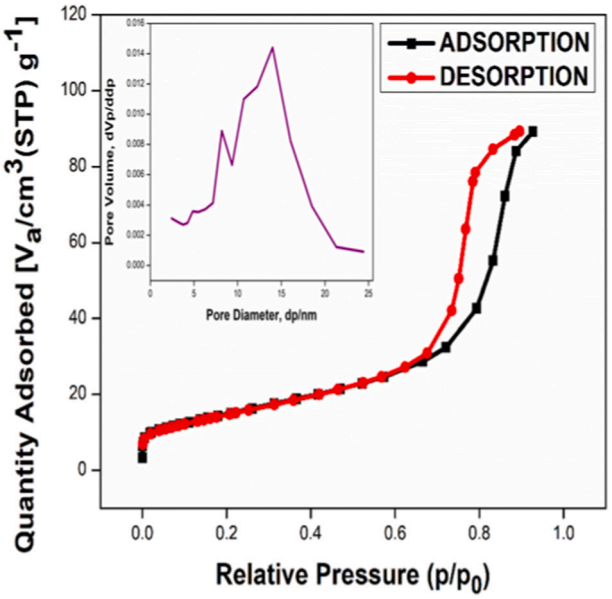
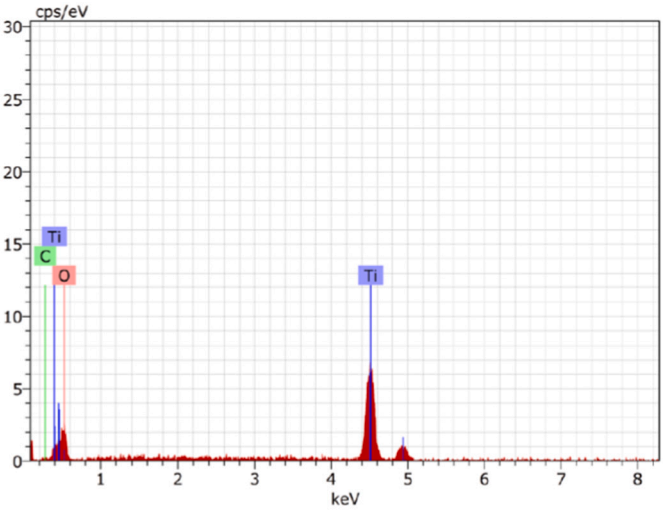



**Fig. 7.** Nitrogen adsorption isotherm of prepared Titanium Dioxide Nanoparticles with pore volume distribution (inset).

**Table 2**
BET characteristics of as synthesized NTO particles.

| | |
|---|---|
| Average surface area | 53.92 $m^2/g$ |
| Total pore volume | 0.138 $cm^3/g$ |
| Mean pore diameter | 10.236 nm |




| Element | Series | unn. C [wt.%] | norm. C [wt.%] | Atom. C [at.%] | Error (3 Sigma) [wt.%] |
|---|---|---|---|---|---|
| Oxygen | K-series | 42.15 | 39.48 | 65.45 | 38.25 |
| Carbon | K-series | 0.66 | 0.62 | 1.37 | 2.56 |
| Titanium | K-series | 63.96 | 59.90 | 33.18 | 7.21 |
| Total: | | 106.77 | 100.00 | 100.00 | |

**Fig. 8.** EDX Spectra of synthesized NTO particles.

sample. Apart from the little traces of C, no trace of any other undesirable impurities has been observed in the above spectrum.

II. Evaluation of fabricated PEG1500/PVA blends and Titanium Dioxide Nanoparticles filled form-stable PCMs nanocomposites (NTO/PEG1500/PVA)

### 4.7. Form-stability of PEG1500/PVA blends

The form-stable characteristics of the PEG1500/PVA blends have been examined based on their leakage rate (in %). Fig. 9 depicts the leakage stability of the hydrophilic polymer blend compositions. The experiment results obtained indicate that the leakage rates varied linearly with increasing PEG content, added to the PVA matrix. From the results obtained through mass loss test, it can be viewed that the blend compositions upto 60 wt% of PEG 1500 exhibited leakage rates <10 %, which, can be considered as minimum leakage, while assessing the efficiency of the form-stability offered by the PVA matrix. Furthermore, according to the filter paper test, the blend composition up to 40 wt% PEG exhibited a negligible trace of PEG on the filter paper surface. The two highest blend concentrations of 70 wt% and 80 wt% showed maximum leakage rate, thus, indicating the poor form-stability characteristics of PEG1500/PVA blends derived from these weight fractions. It has been found out that, the optimum and maximum concentration with decent form-stability behaviour that can be achieved successfully in above-prepared blend compositions turns out to be 60 wt% of PEG 1500 and is thus, labelled as form-stable PCM. Further, all the nanocomposites reinforced with Titanium Dioxide Nanoparticles are prepared using this particular optimized form-stable PCM composition and their intrinsic properties are evaluated thoroughly, for thermal energy storage applications.

### 4.8. Fourier Transform Infrared (FTIR) Spectroscopy of PEG1500/PVA blends and NTO/PEG1500/PVA form-stable PCMs nanocomposites

FTIR spectrum of pure PVA is represented in Fig. 10 (a). The most peculiar band of —OH stretching inferring the inherent feature of alcohol has been found at 3425 $cm^{-1}$. Also, the peaks arising respectively at 3238 & 1341 $cm^{-1}$ corresponds to the stretching and bending modes of H-OH and mixed —CH and —OH interactions associated with alcohols. The peaks observed at 2948 $cm^{-1}$ & 2881 $cm^{-1}$ are related to asymmetric stretching of —$CH_2$ groups and C—H stretching vibration modes. It has been observed that the peaks found at 1718 and 1534 $cm^{-1}$ are associated with the stretching vibrations of the carbonyl group, which hypothesizes the presence of residual acetate groups, at the manufacturing stage of PVA, involving the hydrolysis of the polyvinyl acetate precursor. The peak recorded at 1405 $cm^{-1}$ represents the bending vibrations of —$CH_2$ or —$CH_3$ groups. The characteristic peak at 1131 $cm^{-1}$ has been regarded as crystallization sensitive peak of PVA and it indicates the extent (degree) of crystallinity possessed by the hydrophilic polymer. The peak observed at 927 $cm^{-1}$ gives evidence about the syndiotactic structure and is attributed to —$CH_2$ rocking vibrational modes. The vibrational modes which represent the C—C stretching and —OH out of plane bending were observed at peaks 853 and 634 $cm^{-1}$ respectively. In case of raw PEG 1500 [Fig. 10 (b)], the peaks observed at 3427 and 1633 $cm^{-1}$ have been attributed to the vibrational modes of the —OH functional group. The C—H stretching band of PEG 1500 has been noted at 2887 $cm^{-1}$. The peak found at 1115 $cm^{-1}$ is assigned to the C—O stretching mode of ether moiety. The peaks observed at 964 and 843 $cm^{-1}$ correspond to the crystal peak of PEG and C—C—O stretching vibrations.

The FTIR spectrum of PVA blends with different PEG concentrations, as illustrated in Fig. 10 (c), exhibited nearly identical absorption bands, within the scanned wavelength range, therefore, confining the retention of the primary functional groups of both individual hydrophilic polymers, within the newly formed blend compositions. Though, there is an observable moderate shift in both band positions, as well as, the relative intensities of the stretching vibrations of the hydroxyl group in the resultant blends. This is accredited to the formation of H-bonds between PVA and PEG, at various PEG ratios. Other primary peaks observed in all the blend compositions (20–80 wt%) are found ranging between 2892-2896 $cm^{-1}$, 1710–1712 $cm^{-1}$, 1413–1423 $cm^{-1}$, 1088–1090 $cm^{-1}$, 961–964 $cm^{-1}$, 840–843 $cm^{-1}$, thus, correspond respectively to C—H, C═O (residual acetate groups of PVA), —$CH_2$, C—O—C, C—H (PEG), C—C—O (PEG) stretching and bending vibrational modes. These findings are concordant with the already reported literature [38].

Fig. 10 (d) depicts the FTIR spectrum of NTO/PVA/PEG1500 (60 wt %) nanocomposite films, with varying nanoparticle loadings. The films

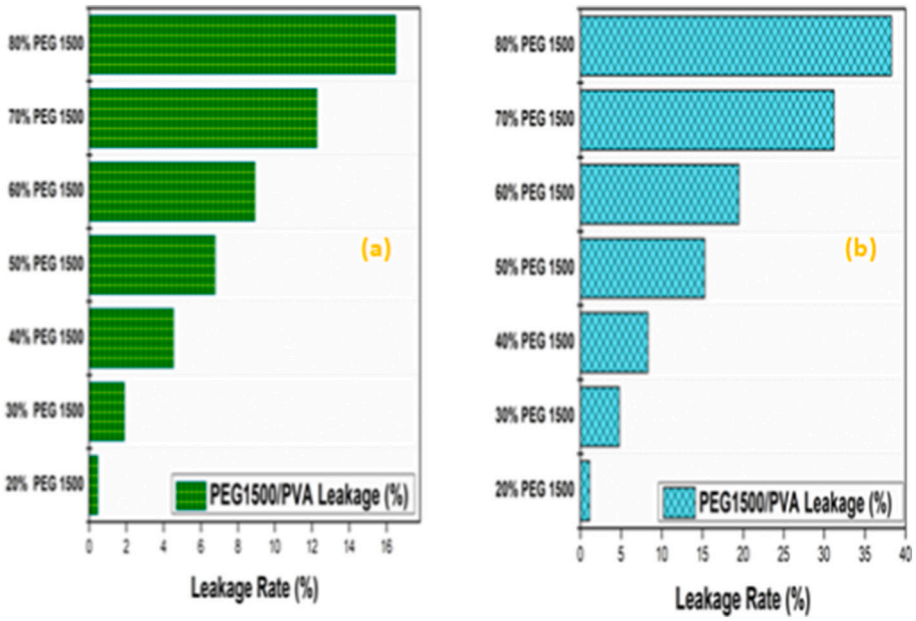



**Fig. 9.** Leakage rates observed for various prepared blend compositions based on (a) measuring the mass loss percentage on weight basis; (b) measuring the diameter of seepage area.

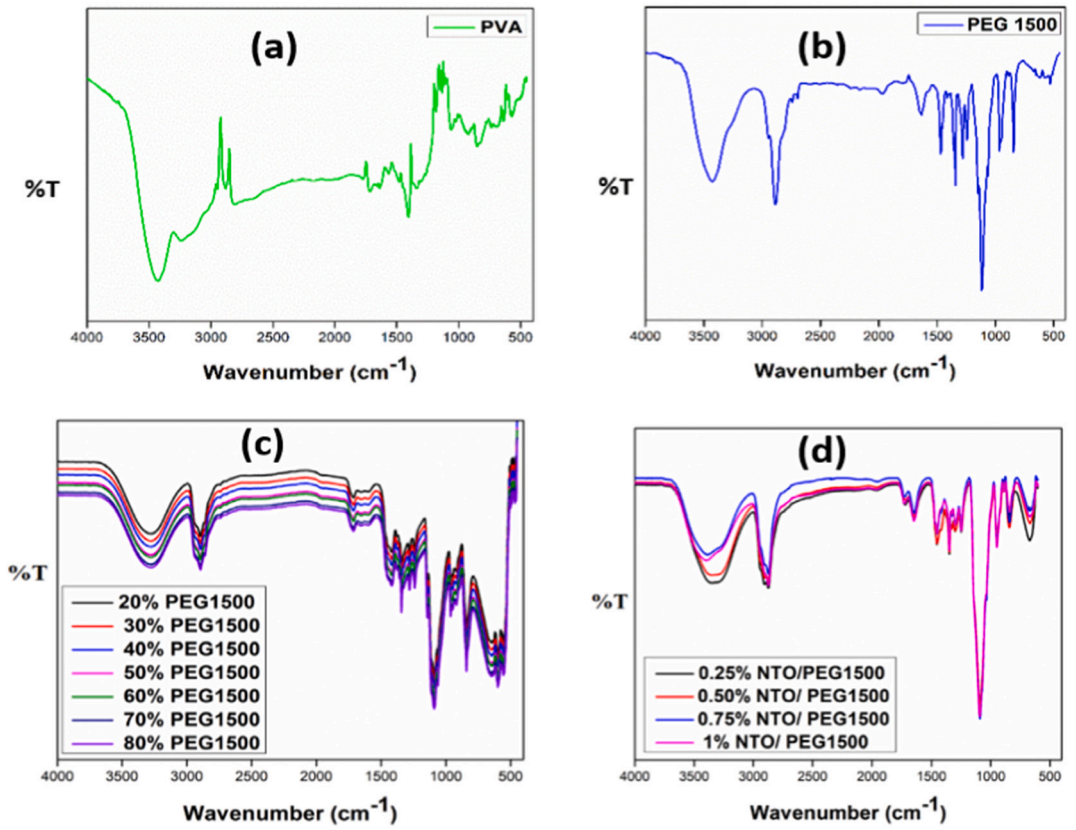



**Fig. 10.** FTIR spectra for (a) PVA; (b) PEG1500; (c) PEG1500/PVA blends with different PEG concentrations; (d) NTO/PEG1500/PVA nanocomposites at different NTO loadings.

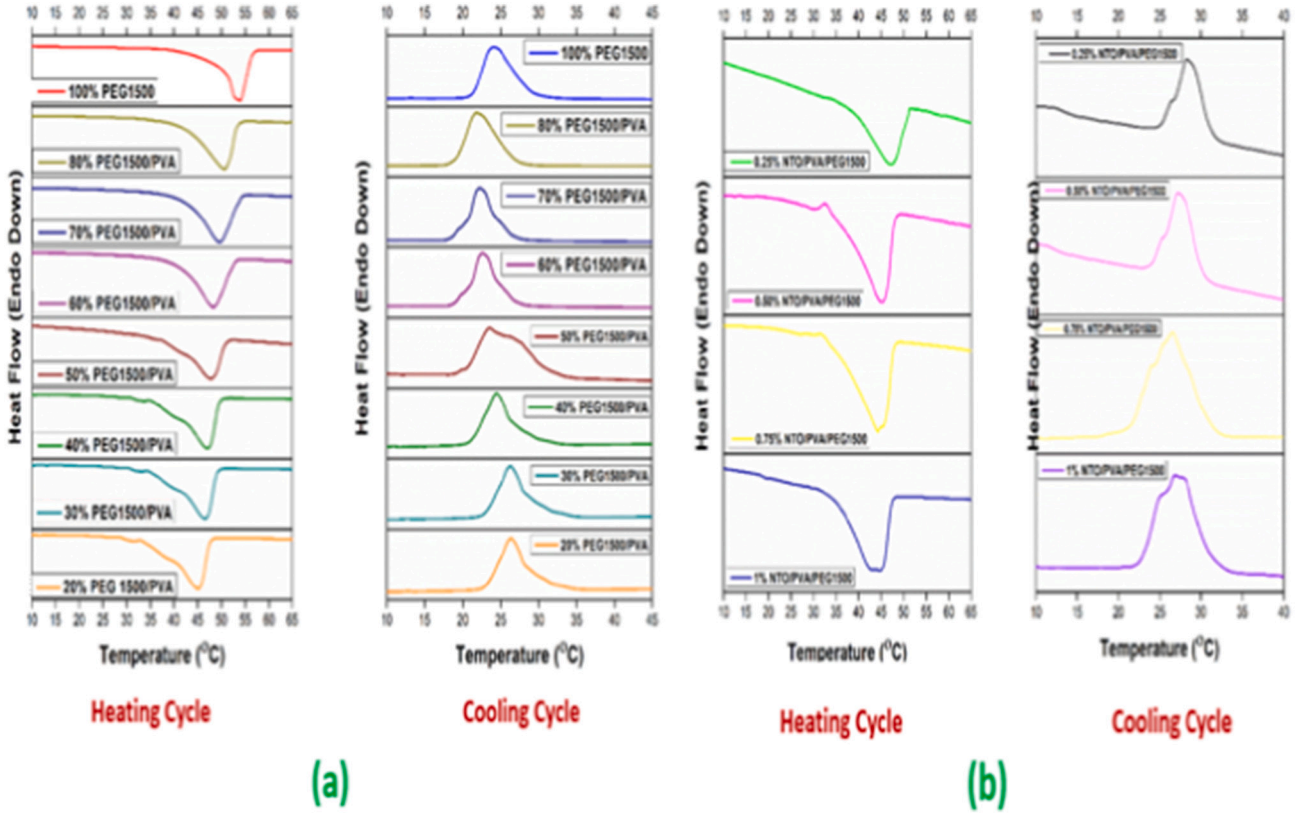



**Fig. 11.** Phase change characteristics for (a) PEG1500 and its associated blends; (b) form-stable PCMs.

showed indistinguishable peaks recorded for the particular blend composition, as also observed earlier in this study. However, the minor peak shifting of —OH bands has been observed, which occurred due to the H-bonding formation between the hydroxy group present in the blend, and the oxygen groups of Titanium Dioxide Nanoparticles. Also, the shifting in the peak intensity of —OH groups in the nanocomposite films with the increasing dosage of NTO has been observed, which might be due to the agglomeration of metal oxide particles, leading to the weakening in H-bond interactions [38].

### 4.9. Phase change characteristics of PEG1500/PVA blends and NTO/PEG1500/PVA form-stable PCM nanocomposites

Differential Scanning Calorimetry (DSC) has been utilized for the evaluation of the phase change behaviour of all the blend compositions (with 20–80 wt% of PEG1500) and the NTO filled-60 wt% PEG1500/PVA nanocomposites. Fig. 11 (a) and Table 3 (a & b) portray the melting and crystallization temperatures of PEG1500 and its corresponding blend compositions. Pure PEG used as the main latent heat storage PCM showed onset melting and crystallization temperatures at 38.13 °C and 34.18 °C, respectively. The enthalpies accompanying the phase change were estimated as 150.75 and 151.25 J/g, respectively. In case of blend compositions, it was found that the melting temperatures, and their associated enthalpies increases in a progressive manner, due to an increase in active PEG PCM concentrations, in the molecular domains of the PVA supporting matrix. The optimized blend composition of 60 wt% PEG1500/PVA undergoes melting and crystallization transitions at 36.14 °C and 31.02 °C, respectively. Also, it exhibited peak melting and crystallization temperatures at 48.27 °C, with the latent heat of fusion of 60.65 J/g, and 22.56 °C with the corresponding latent heat of crystallization 62.08 J/g, respectively.

The influence on the thermal energy storage parameters after the incorporation of metal oxide nanoparticles to the optimized blend composition (PEG~60 wt%) can be seen significantly from the obtained thermal performance data, demonstrated by the DSC analysis. The phase change curves representing the phase-change temperatures, and the latent heat enthalpy values for the varied loading of synthesized nanoparticles in the NTO/PEG1500/PVA nanocomposites are shown in Fig. 11 (b) and their results are summarized in Table 4 (a & b). The obtained results suggest that the addition of Titanium Dioxide Nanoparticles has led to enhanced heat transfer rates, which, tend to escalate with increasing order of NTO loading, owing to the high thermal conductivity of the metal oxide nanoparticles. This also accounts for the surge in the latent heat enthalpies of the nanocomposites against the gradual increase in the concentration of nanoparticles, in the form-stable PCM nanocomposites, as depicted in Fig. 12 (a). For the maximum NTO concentration of 1 %, the onset melting and crystallization temperatures are reported as 29.69 °C and 34.13 °C, respectively. The peak melting and crystallization temperatures have been recorded at 43.88 °C and 26.85 °C, along with corresponding estimated enthalpies of 73.02 J/g and 70.23 J/g, respectively.

An important concerning aspect associated with PEG PCMs, about their possible utilization as an effective latent heat thermal energy storage system, is the supercooling phenomenon. Typically, the degree of supercooling of a PCM is derived by calculating the difference between the peak melting and crystallization temperatures of the working PCM. It is desirable for a PCM used for varying TES application, to possess a narrow supercooling degree as possible. Fig. 12 (b) represents the variation in the degree of supercooling of the nanocomposite films subjected to different loading of NTO in the form-stable PCMs. It has been estimated that while increasing the NTO amount in form-stable composites, can remarkably refine the supercooling degree to a slightly better extent. This control in supercooling phenomenon arises due to the foremost action of Titanium Dioxide Nanoparticles, by acting as a potent nucleating agent during the heat-cool cycle of PEG-based form-stable nanocomposites, resulting in an improved solidification rate of the PEG macromolecules, during cooling process. The comparison of the form-stable PCM prepared in the current study with few existing counterparts, testified in the literature is reported in Table 5. Therefore, the resulted form-stable nanocomposites displayed a very

**Table 3**
DSC data of pure PCM and PEG blends representing their phase change characteristics.

(a)

| Samples | Onset melting temperature, $T_{Mo}$ (°C) | Peak melting temperature, $T_{Mp}$ (°C) | Offset melting temperature, $T_{Me}$ (°C) | Latent heat of fusion, $\Delta H_m$ (J/g) |
|---|---|---|---|---|
| PEG1500 | 38.13 | 53.89 | 58.18 | 150.75 |
| 20 % PEG1500/PVA | 28.59 | 45.08 | 49.21 | 8.81 |
| 30 % PEG1500/PVA | 30.19 | 46.61 | 50.78 | 19.36 |
| 40 % PEG1500/PVA | 30.62 | 47.10 | 51.23 | 30.57 |
| 50 % PEG1500/PVA | 32.07 | 47.78 | 52.20 | 57.56 |
| 60 % PEG1500/PVA | 36.14 | 48.27 | 54.15 | 60.65 |
| 70 % PEG1500/PVA | 37.66 | 49.65 | 55.15 | 81.45 |
| 80 % PEG1500/PVA | 37.88 | 50.63 | 55.38 | 131.32 |

(b)

| Samples | Onset crystallization temperature, $T_{Co}$ (°C) | Peak crystallization temperature, $T_{Cp}$ (°C) | Offset crystallization temperature, $T_{Ce}$ (°C) | Latent heat of crystallization, $\Delta H_f$ (J/g) |
|---|---|---|---|---|
| PEG1500 | 34.18 | 24.09 | 20.2 | 151.25 |
| 20 % PEG1500/PVA | 36.48 | 26.39 | 19.98 | 10.68 |
| 30 % PEG1500/PVA | 36.45 | 26.28 | 20.13 | 17.29 |
| 40 % PEG1500/PVA | 35.04 | 24.42 | 18.2 | 33.38 |
| 50 % PEG1500/PVA | 34.61 | 23.51 | 18.27 | 60.37 |
| 60 % PEG1500/PVA | 31.02 | 22.56 | 17.59 | 62.08 |
| 70 % PEG1500/PVA | 30.29 | 22.18 | 17.46 | 85.49 |
| 80 % PEG1500/PVA | 29.14 | 21.82 | 17.17 | 139.44 |

**Table 4**
DSC data of NTO/PEG1500/PVA form-stable PCMs representing their phase change characteristics.

(a)

| Samples | Onset melting temperature, $T_{Mo}$(°C) | Peak melting temperature, $T_{Mp}$ (°C) | Offset melting temperature, $T_{Me}$ (°C) | Latent heat of fusion, $\Delta H_m$ (J/g) |
|---|---|---|---|---|
| 0.25 % NTO/PEG1500/PVA | 33.48 | 47.27 | 51.3 | 65.59 |
| 0.50 % NTO/PEG1500/PVA | 32.49 | 45.24 | 49.43 | 68.43 |
| 0.75 % NTO/PEG1500/PVA | 31.48 | 44.25 | 48.8 | 71.70 |
| 1 % NTO/PEG1500/PVA | 29.69 | 43.88 | 48.52 | 73.02 |

(b)

| Samples | Onset crystallization temperature, $T_{Co}$(°C) | Peak crystallization temperature, $T_{Cp}$ (°C) | Offset crystallization temperature, $T_{Ce}$ (°C) | Latent heat of crystallization, $\Delta H_f$ (J/g) |
|---|---|---|---|---|
| 0.25 % NTO/PEG1500/PVA | 32.74 | 28.34 | 24.05 | 63.97 |
| 0.50 % NTO/PEG1500/PVA | 32.98 | 27.32 | 23.13 | 66.12 |
| 0.75 % NTO/PEG1500/PVA | 33.65 | 26.53 | 20.47 | 69.82 |
| 1 % NTO/PEG1500/PVA | 34.13 | 26.85 | 21.29 | 70.23 |

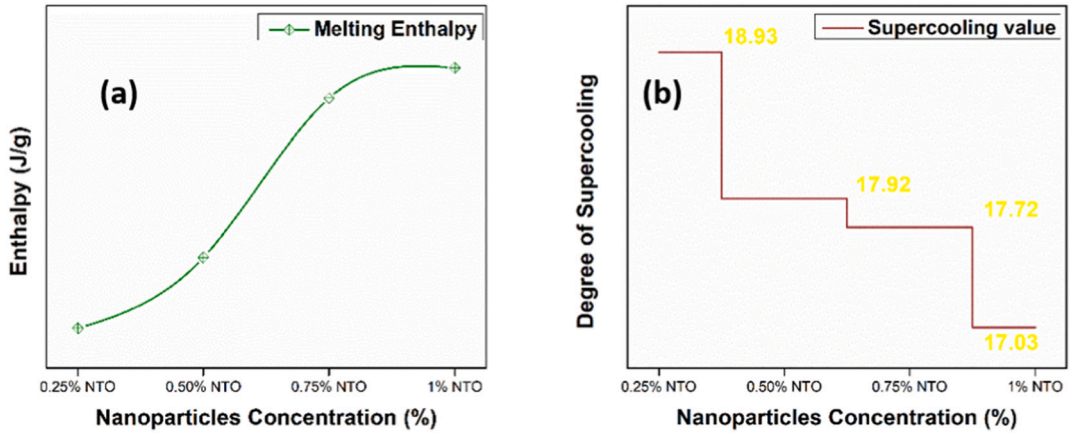



**Fig. 12.** Effect of nanoparticles concentrations on (a) latent heat enthalpy of fusion; (b) degree of supercooling.

**Table 5**
Comparison of some form-stable PCMs from literature and prepared form-stable PCMs.

| Core PCM | Matrix | Additive | $\Delta H_m$ (J/g) | References |
|---|---|---|---|---|
| PEG | PVA | Titanium Dioxide Nanoparticles | 73.02 | Present study |
| PEG | Graphitic Carbon Nitride | – | 45.8 | [39] |
| Dodecanol microcapsules | HDPE | Wood Flour | 31.6 | [26] |
| Paraffin microcapsules | PVC | Wood Flour | 15.92 | [40] |
| n-Octadecane | PVC | Titanium Dioxide Nanoparticles | 9.22 | [20] |
| Paraffin microcapsules | HDPE | Wood Flour | 42.8 | [14] |

good latent heat energy potential and can be applied for suitable thermal buffering applications.

### *4.10. Thermal stability of PVA/PEG1500 blends and NTO/PEG1500/PVA nanocomposites*

From a practical viewpoint, the thermal stability is a significant property to evaluate the polymer products, especially the thin plastic films. In commercial approach, generally, the temperature ($T_{50}$) corresponding to a weight loss equivalent to 50 %, is considered as an important parameter to access the thermal stability performance. Sometimes in many industrial applications from thermo-mechanical stability aspect, the 50 % weight loss is regarded as the initial failure of the material during service life, even before reaching the maximum degradation. Fig. 13 (a) represents the thermal stability characteristics of pure PVA, PEG1500, and PVA/PEG1500 blend compositions with varying PEG1500 content. The pure PVA showed $T_{50}$ and $T_{80}$ at 326.2 °C and 366.6 °C, respectively whereas in the case of working PCM 50 % and 80 % degradation have been observed to be occurring at 416.1 °C and 427.9 °C, respectively, which signifies the high thermal stability of the active PCM, even more than that of the supporting matrix. This is due to

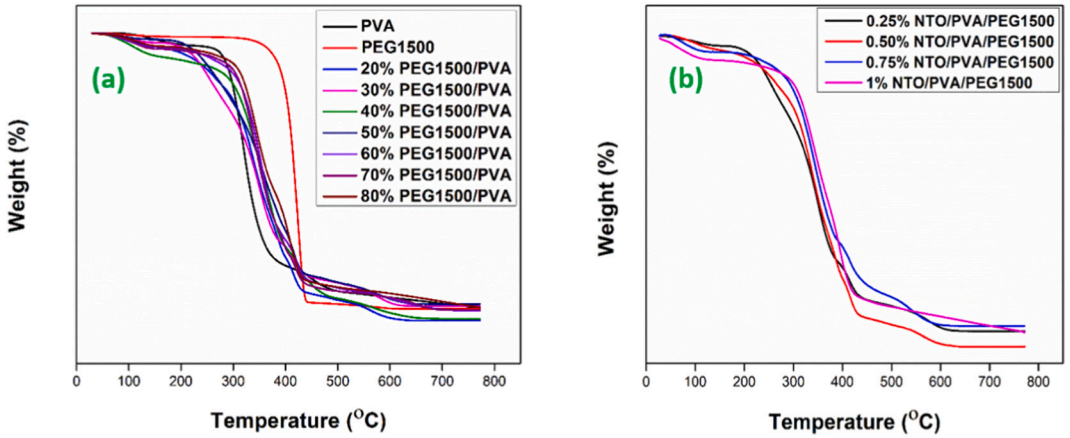



**Fig. 13.** TGA curves for (a) PEG1500/PVA blends with different PEG concentrations; (b) NTO/PEG1500/PVA nanocomposites at different NTO loadings.

the more crystalline nature of PEG compared to PVA, which arises because of more compact and linear chain structure of PEG macromolecule, which leads to an effective packing of the crystal lattice structure, in the crystalline region of the polymer. On the other hand, in PVA, the pendant hydroxy groups (-OH) attached to PVAs main backbone chain creates hindrance in the chain folding arrangement and disturbs the effective crystal packing phenomenon of the macromolecule chains, that enables the PVA to exhibit amorphous nature, and thus, a little less thermally stable, compared to PEG blend compositions, with varying PEG contents (20–80 wt%). It was found that the thermal stability behaviour of the PVA/PEG blends tends to improve, with increase in PEG concentration within the PVA matrix, owing to the high crystalline nature of polyethylene glycol. Therefore, it strongly can be considered as a highly advantageous approach for the incorporation of polymeric PCM into the PVA matrix, wherein the active PCM acts as dual roles including working as an efficient latent heat storage component, as well as, for enhancing the overall thermal stability, at elevated temperatures, in all blend compositions. The addition of 60 wt% PEG to a vinyl polymer resulting in PEG1500/PVA blend improves and shifts the $T_{50}$ of the PVA matrix to 354.8 °C and finally undergoes degradation at 410.6 °C.

The thermal stability of metal oxide filled nanocomposite films, (NTO/PEG1500/PVA) comprising of 60 wt% PEG loading, with varying concentrations of Titanium Dioxide Nanoparticles (NTO: 0.25 %–1 %) is shown in Fig. 13 (b). It has been observed that the addition of prepared nanoparticles to the fabricated 60 wt% PEG1500/PVA film, in increasing order dosage, has resulted in the enhancement of thermal stability of the nanocomposite films. This is attributed to the high thermally stable nature of nano-sized inorganic metal oxide filler particles. It is believed that the presence of NTO particles embedded inside the PEG1500/PVA blend boosts the thermal stability of the nanocomposites films by restricting the bond dissociation or the mobility of polymer chains due to the crowding of nanoparticles across the macromolecular network of blended polymers, at elevated temperatures (Table 6).

**Table 6**
Thermal stabilities of (a) PEG1500/PVA blends with different PEG concentrations; (b) NTO/PEG1500/PVA nanocomposites at different NTO loading.

(a)

| Samples | 50 % weight loss temperature, $T_{50}$ (°C) | 80 % weight loss temperature, $T_{80}$ (°C) |
|---|---|---|
| PVA | 326.2 | 366.6 |
| PEG 1500 | 416.1 | 427.9 |
| 20 % PEG 1500/ PVA | 340.3 | 393.6 |
| 30 % PEG 1500/ PVA | 342.4 | 402.5 |
| 40 % PEG 1500/ PVA | 351.0 | 405.1 |
| 50 % PEG 1500/ PVA | 353.9 | 405.6 |
| 60 % PEG 1500/ PVA | 354.8 | 410.6 |
| 70 % PEG 1500/ PVA | 355.1 | 413.7 |
| 80 % PEG 1500/ PVA | 364.6 | 414.6 |

(b)

| Samples | 50 % weight loss temperature, $T_{50}$ (°C) | 80 % weight loss temperature, $T_{80}$ (°C) |
|---|---|---|
| 0.25 % NTO/PVA/ PEG 1500 | 355.8 | 412.26 |
| 0.50 % NTO/PVA/ PEG 1500 | 358.4 | 416.32 |
| 0.75 % NTO/PVA/ PEG 1500 | 360.6 | 422.45 |
| 1 % NTO/PVA/PEG 1500 | 362.1 | 425.62 |

### *4.11. Thermal energy storage performance of NTO/PEG1500/PVA form-stable nanocomposites*

The thermal energy storage capability of the 1 % NTO/PEG1500/PVA nanocomposite form-stable PCM was determined by observing the charging-discharging times of the form-stable PCM relative to pure PEG. The melting and crystallization time were evaluated from the charging and discharging cycles of PEG and composite PCM. The representative temperature history curves of raw PCM and form-stable PCM nanocomposite are shown in Fig. 14. It has been observed that the PEG1500 reached its phase change temperature ($T_{mo}$: 38.13 °C) from 10 °C in 481 s whereas, in the case of form-stable PCMs, the time taken to hit its onset melting temperature ($T_{mo}$: 29.69 °C) was 472 s. Similarly, the discharging time for the pure PCM and metal oxide filled PCM nanocomposite were found as 519 and 747 s, respectively. The obtained results can be interpreted in such a manner that the introduction of Titanium Dioxide Nanoparticles to the form-stable PCM end product fastens the heat transfer mechanism within the composite structure, owing to their high thermal conductivity and large specific surface area. Secondly, in the case of nanocomposite films, the PCM content is only 60 wt% against the 100 wt% fraction of pure PEG PCM. Therefore, the presence of active ingredient components in lesser amounts tends to undergo phase transitions much faster, at shorter temperature profiles than that of the bulk material. Furthermore, since both samples were heated and cooled under the same temperature profile, as the water

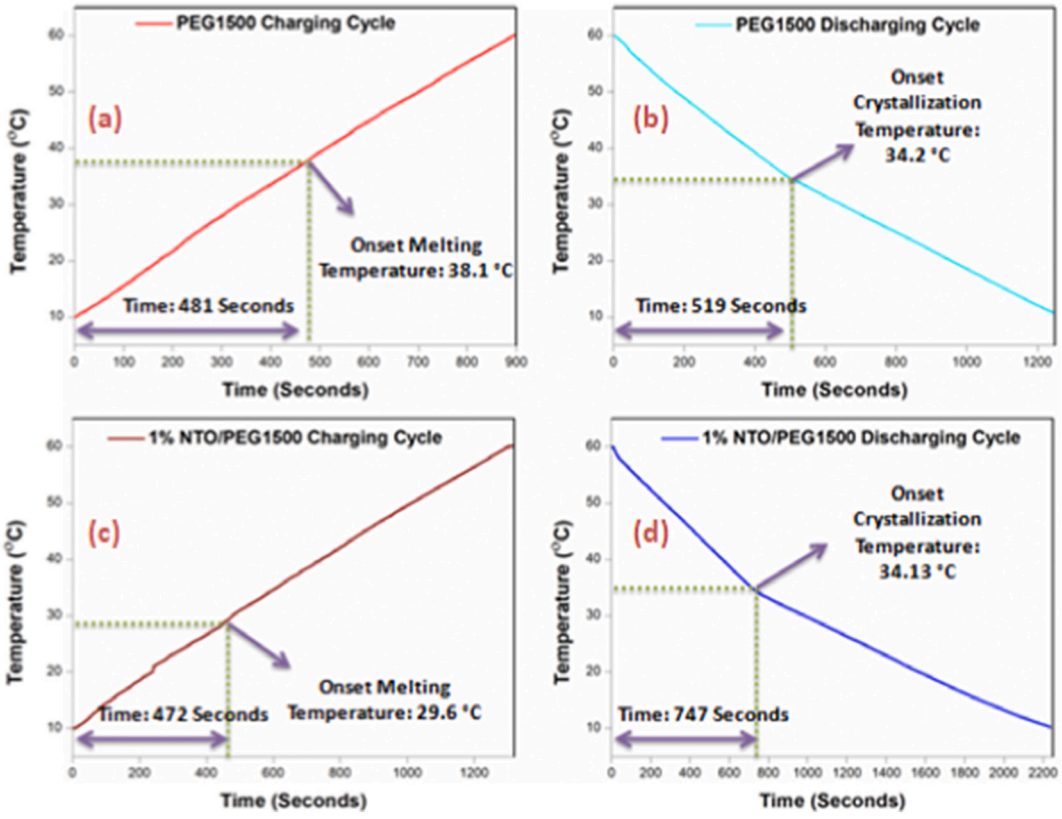



**Fig. 14.** Thermal energy storage curves for bulk PEG 1500 and form-stable PCM (T-History).

baths were thermostated constantly between 10 and 60 °C, it can be observed that the nanocomposite material took more time to attain this temperature limit during crystallization process as compared to raw PEG, which indicates the presence of polymer macromolecules in the neighbourhood of PEG molecules; giving evidence of successful fabrication of form-stable PCM. Thus, the insulating nature of the plastic supporting matrix along with the crowding of nanoparticles within the composite material hinders the mobility of PEG molecules at the time of crystallization period could be accounted for the delay in achieving the equilibrium temperature.

### *4.12. Thermal reliability performance of NTO/PEG1500/PVA form-stable nanocomposites*

DSC curves and changes in thermal characteristics of form-stable PCM after exposure to 100 thermal cycles are represented in Fig. 15 and Table 7 respectively. After 100 cycles, the metal oxide-filled PEG nanocomposite PCM exhibited latent heat enthalpies of melting and crystallization as 66.38 J/g and 63.14 J/g respectively. It is clear from the results obtained that the little variation in temperature or enthalpy values suggests the decent thermal reliability behaviour of the prepared form-stable PCMs, and these are much suitable for utility in the thermal energy storage applications.

**Table 7**
Phase change characteristics of 1 % NTO/PEG1500/PVA nanocomposite PCM post thermal cycling exposure.

| Number of cycles | Onset T*m* (°C) | Onset Tc (°C) | ΔHm (J/g) | ΔHc (J/g) |
|---|---|---|---|---|
| 0 | 29.69 | 34.13 | 73.02 | 70.23 |
| 100 | 31.9 | 32.63 | 66.38 | 63.14 |

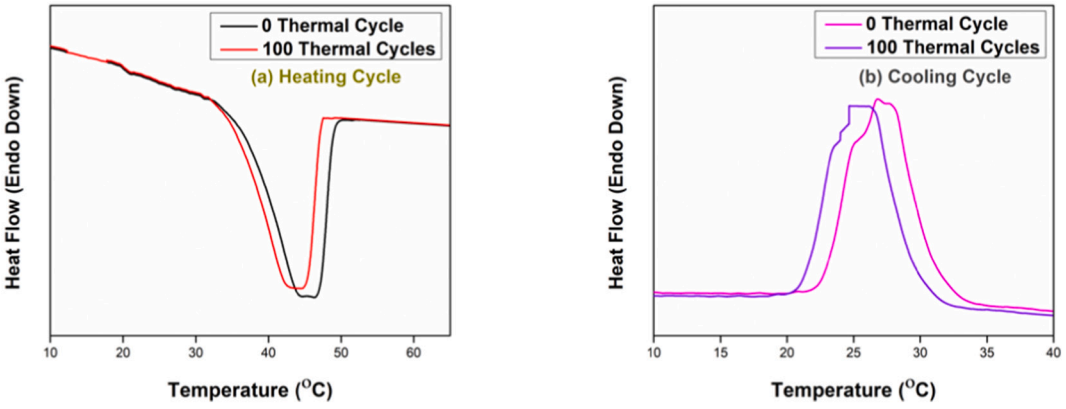



**Fig. 15.** Thermal reliability curves of form-stable PCMs (a) heating cycle; (b) cooling cycle.

### 4.13. Structural characterization of NTO/PEG1500/PVA form-stable PCMs

FE-SEM photographs of 1 % NTO/60 wt% PEG1500/PVA shown in Fig. 16 (a–d) provide an insight into the resulting microstructures of the prepared form-stable nanocomposites. From these obtained figures, it can be inferred that the PEG crystals of uneven shape and size can be seen randomly distributed and held together, throughout within the surface of the supporting matrix. This affirms the role of polymers such as PVA as an effective supporting matrix and could be considered as a potential matrix material component for fabrication of form-stable PCMs. Although, it seems difficult to observe the individual filler particles dispersion in the examined nanocomposite film due to the small size of fine nanoparticles, however, a few aggregates of NTO (represented in yellow circles) can be found present over the macromolecular structural networks of the resultant hydrophilic matrix blend composition. Therefore, the existence of the nano-sized metal oxide particles in the nanocomposites eventually contributed toward the enhancement in the heat-transfer rates, thermal resistance, flame-retardant, and UV protection characteristics of the successfully fabricated final product.

### 4.14. UV-blocking characteristics of NTO/PEG1500/PVA form-stable PCMs

It is now a well-known fact governing the ultraviolet rays (UV) incident on earth's surface and the adverse issues associated with them prolong unsafe exposure, which is considered critical both to human's safety and efficient materials performance. UV radiation is generally sub-categorized as: (i) UV-A ($\lambda = 315–380$ nm), (ii) UV-B ($\lambda = 280–315$ nm) and (iii) UV-C ($\lambda = 100–280$ nm), respectively. The sunlight elapses across the earth's atmosphere, most of the UV-B and UV-C get absorbed by the ozone layer, oxygen, carbon dioxide, water vapor, etc. However, unfortunately, the natural climatic ambiance is unable to filter out UV-A much significantly. Prolong UV irradiation has not been ever fruitful to human health. UV-A exposure causes intense aging, by diminishing the immunological response of skin cells. In addition, erythema and skin cancer are the harmful effects associated with unsafe exposure to UV-B [20]. These fabricated nanocomposites, depending upon the requirement, may be intended to serve outdoor applications. Also, it is well-known fact that the titanium dioxide profoundly absorbs the UV rays, owing to its semiconducting properties, which are well supported and elaborated by the electronic band theory of solids. Keeping all these points in mind, the Titanium Dioxide Nanoparticles (NTO) have been added to the optimized blend composition films to impart UV-blocking features to the end product. Fig. 17 shows the UV–vis transmittance spectra describing the UV protection characteristics of the NTO filled polymer nanocomposites. It is evident from the obtained results, that a progressive significant reduction in UV transmittance percentage (%) in the UV-A region has been observed in the metal oxide reinforced nanocomposites, with the gradual increase in the NTO concentrations (0.25–1 %). It is accounted in general, that there is a sort of ion transfer mechanism between the (PVA-PEG) polymer blend and NTO particles. Also it is assumed, for titanium dioxide, or its derived products, such as, nanocomposites filled with Titanium Dioxide Nanoparticles, the UV-blocking properties of such materials are governed by the absorption and reflection of UV rays, because of its high refractive index [41,42]. Hence, these findings signify the key role of the dispersed Titanium Dioxide Nanoparticles in the produced nanocomposites, bestowing them with excellent UV-resistant nature.

### 4.15. Flammability characteristics of NTO/PEG nanocomposite films

Limiting Oxygen Index (LOI) is a widely performed qualitative test method to access the flame retardant (FR) behaviour of the polymers. It

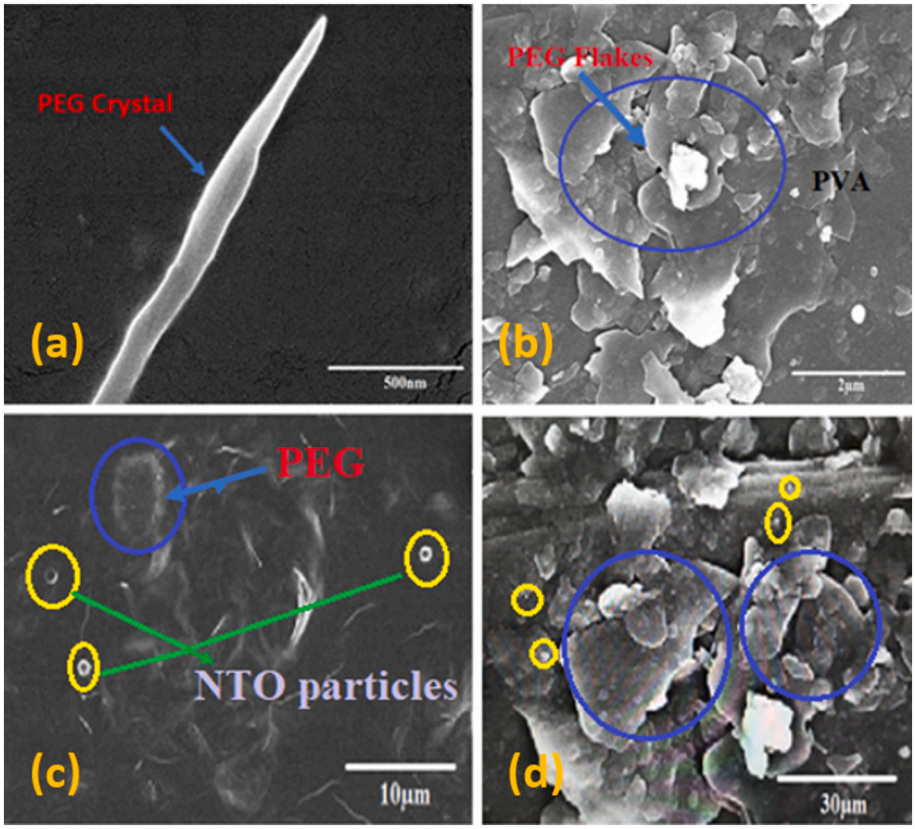



**Fig. 16.** FE-SEM micrographs of the prepared form-stable PCMs (a) magnification-7.6 mm × 70.0 K X (b) magnification-7.6 mm × 25.0 K X (c) magnification-7.6 mm × 5.0 K X and (d) magnification-7.6 mm × 1.50 K X.

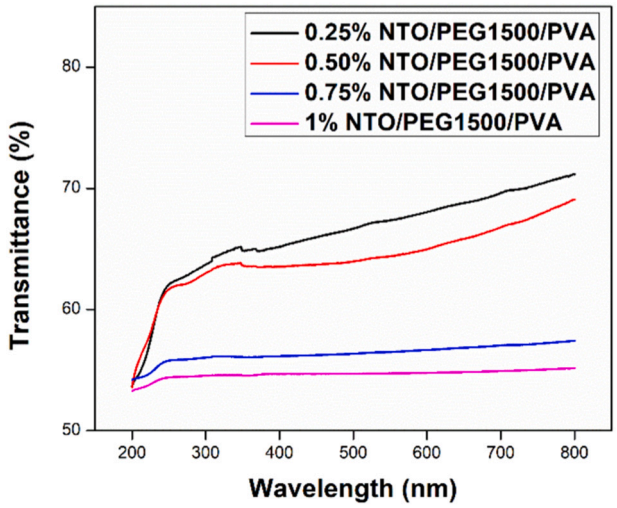



**Fig. 17.** UV–vis transmittance spectra of NTO/PEG1500/PVA with varying loading of Titanium Dioxide Nanoparticles.

is expressed in number percentage (%). The significance of the LOI value obtained during the performance of the test provides the information about the minimum concentration of oxygen (in a mixture of oxygen and nitrogen) required to support continuous combustion of the polymer. There exists an inverse relationship between the LOI and flammability nature. Table 8 represents the LOI values of 60 wt% PEG1500/PVA blend and their metal oxide-filled nanocomposites, with varying NTO dosage. Pure PVA has LOI value of 22.7 %. After blending 60 wt% PEG1500 into the PVA matrix has led to a sudden decline in LOI value of the blend composition to 19.8 %. It is possibly due to the colossal molecular size difference between the macromolecules used for the blend preparation. In comparison to the PVA molecular chain size, having molecular weight; M.W.: 14000, the PEG 1500 can be considered to be acting as a low molecular weight plasticizer molecule, with respect to the vinyl polymer, which offers more flexibility to the resultant blend thus, led to the reduction in LOI value in the examined blend composition. However, the incorporation of the metal oxide nanoparticles to the above-formed blend is significant for their flame retardancy characteristics, as they tend to minimize the production of toxic burning products, by allowing the formation of the protective shield layer against the flame as compared to the production of low molecular weight fragments or by forming complex products, as found with other conventional FR additives such as antimony, phosphorous, or halogenated compounds, etc. It has been observed that, the introduction of titanium nanoparticles has led to a jump in the LOI value of the NTO/PEG1500/PVA nanocomposites, and it further increases linearly with the increasing NTO weight fraction in the final products. This is credited to the basic flame retardant working mechanism of metal oxides nanoparticles, as the Titanium Dioxide is often considered as an inert mineral FR agent, which follows the physical pathway mechanism for flame retardancy action. Because of the high specific surface area of the NTO particles, they tend to fill more and more localized spaces between the macromolecular chains, and, such effective crowding of nanoparticles restricts the chain mobility at elevated temperature, and thus, demands more thermal energy for bond dissociation. During the burning process, uniformly distributed NTO particles inside the nanocomposites are supposed to accumulate over the surface of the composite sample, forming a barrier layer. This newly formed inert protective thermal layer, reduces the heat transfer to the remaining unexposed surface layer of the composite material, and hinders the passage of burning combustible gases and cut-off the oxygen supply, along with suppressing the smoke generation during combustion [43–45].

**Table 8**
Flammability characteristics of NTO/PEG1500/PVA form-stable PCMs.

| Samples | LOI value (%) |
|---|---|
| PVA | 22.7 ± 0.2 |
| PVA/PEG 1500 | 19.8 ± 0.3 |
| 0.25 % NTO/PEG1500 | 20.3 ± 0.2 |
| 0.50 % NTO/PEG1500 | 21.2 ± 0.3 |
| 0.75 % NTO/PEG1500 | 21.8 ± 0.4 |
| 1 % NTO/PEG1500 | 22.4 ± 0.3 |

## 5. Conclusions

A series of form-stable PCM nanocomposites have been prepared successfully, comprising of PEG based phase change material confined inside PVA supporting matrix reinforced with Titanium Dioxide Nanoparticles. FTIR studies have provided the information about the chemical composition of nanocomposites. The FE-SEM images inferred that the mixing of both hydrophilic polymers has resulted into the formation of immiscible blend. Also, the crystal of PEG and NTO particles can been well distributed within the PVA domains, in the microphotograph. The TGA analysis has portrayed the good thermal stability of the prepared form-stable PCMs, which tends to increase with increasing NTO loading. DSC analysis has shown the potential of the prepared form-stable PCMs as efficient thermal energy storage component. Besides this, the latent heat capacities of the nanocomposite PCMs progresses linearly with increased loading of the metal oxide nanoparticles. The PEG-based form-stable PCMs exhibited a valuable thermal cycling stability post 100-thermal cycle exposure. In UV–vis transmittance spectra, the NTO particles prove effective in decreasing the UV transmittance rates, even though the nanocomposite films, by absorbing/scattering the large amount of part of solar radiation. Furthermore, the incorporation of NTO contributed toward the significant improvement in the flame-retardancy of the form-stable PCMs. Insight of the results obtained in the present study, the fabricated form-stable PCMs can play a vital role in the field of thermal energy storage applications, wherein, their thermal buffering capacity can be utilized in different sectors such as packaging, building, transportation and logistic support etc.

## Declaration of competing interest

The work submitted in this manuscript has been conducted for the research purpose only, and there is no conflict of interest involved.

## Data availability

Data will be made available on request.

## Acknowledgements

We are highly thankful to Delhi Technological University for providing us the necessary facilities to conduct this study.

## References

[1] F. Jiang, X. Wang, D. Wu, Design and synthesis of magnetic microcapsules based on n-eicosane core and Fe3O4/SiO2 hybrid shell for dual-functional phase change materials, Appl. Energy 134 (2014) 456–468, https://doi.org/10.1016/j.apenergy.2014.08.061.
[2] M. Nourani, N. Hamdami, J. Keramat, A. Moheb, M. Shahedi, Preparation of a stable nanocomposite phase change material (NCPCM) using sodium stearoyl lactylate (SSL) as the surfactant and evaluation of its stability using image analysis, Renew. Energy 93 (2016) 404–411, https://doi.org/10.1016/j.renene.2016.02.073.
[3] M. Ünal, Y. Konuklu, H. Paksoy, Thermal buffering effect of a packaging design with microencapsulated phase change material, Int. J. Energy Res. 43 (2019) 4495–4505, https://doi.org/10.1002/er.4578.
[4] L. Melone, L. Altomare, A. Cigada, L. De Nardo, Phase change material cellulosic composites for the cold storage of perishable products: from material preparation

to computational evaluation, Appl. Energy 89 (2012) 339–346, https://doi.org/10.1016/j.apenergy.2011.07.039.
[5] Y. Konuklu, H. Paksoy, Polystyrene-based caprylic acid microencapsulation for thermal energy storage, Sol. Energy Mater. Sol. Cells 159 (2017) 235–242, https://doi.org/10.1016/j.solmat.2016.09.016.
[6] R. Kaur, S. Tanwar, Development and investigation of microencapsulated caprylic acid-based phase change materials for thermal energy storage application, Int. J. Energy Res. 45 (2021) 17302–17314, https://doi.org/10.1002/er.6611.
[7] A. Bastani, F. Haghighat, J. Kozinski, Designing building envelope with PCM wallboards: design tool development, Renew. Sustain. Energy Rev. 31 (2014) 554–562, https://doi.org/10.1016/j.rser.2013.12.031.
[8] W. Chen, W. Weng, Ultrafine lauric – myristic acid eutectic/poly (meta-phenylene isophthalamide) form-stable phase change fibers for thermal energy storage by electrospinning, Appl. Energy 173 (2016) 168–176, https://doi.org/10.1016/j.apenergy.2016.04.061.
[9] X. Yuan, K. Zhu, X. Li, J. Su, H. Li, Y. Zhao, Improvement of anti-icing properties of low surface energy coatings by introducing phase-change microcapsules, Polym. Eng. Sci. 58 (2017) 973–979, https://doi.org/10.1002/pen.24654.
[10] D. Su, Y. Jia, G. Alva, F. Tang, G. Fang, Preparation and thermal properties of n–octadecane/stearic acid eutectic mixtures with hexagonal boron nitride as phase change materials for thermal energy storage, Energy Build. 131 (2016) 35–41, https://doi.org/10.1016/j.enbuild.2016.09.022.
[11] L. Feng, J. Zheng, H. Yang, Y. Guo, W. Li, X. Li, Preparation and characterization of polyethylene glycol/active carbon composites as shape-stabilized phase change materials, Sol. Energy Mater. Sol. Cells 95 (2011) 644–650, https://doi.org/10.1016/j.solmat.2010.09.033.
[12] M. Mu, P.A.M. Basheer, W. Sha, Y. Bai, T. Mcnally, Shape stabilised phase change materials based on a high melt viscosity HDPE and paraffin waxes, Appl. Energy 162 (2016) 68–82, https://doi.org/10.1016/j.apenergy.2015.10.030.
[13] L. Xu, J. Wang, R. Yang, A new flame retardance strategy for shape stabilized phase change materials by surface coating, Sol. Energy Mater. Sol. Cells 170 (2017) 87–94, https://doi.org/10.1016/j.solmat.2017.05.037.
[14] A. Jamekhorshid, S.M. Sadrameli, R. Barzin, M.M. Farid, Composite of wood-plastic and micro-encapsulated phase change material (MEPCM) used for thermal energy storage, Appl. Therm. Eng. 112 (2017) 82–88, https://doi.org/10.1016/j.applthermaleng.2016.10.037.
[15] A. Trigui, M. Karkri, I. Krupa, Thermal conductivity and latent heat thermal energy storage properties of LDPE/wax as a shape-stabilized composite phase change material, Energy Convers. Manag. 77 (2014) 586–596, https://doi.org/10.1016/j.enconman.2013.09.034.
[16] P. Sobolciak, M. Karkri, M.A. Al-Maadeed, I. Krupa, Thermal characterization of phase change materials based on linear low-density polyethylene, paraffin wax and expanded graphite, Renew. Energy 88 (2016) 372–382, https://doi.org/10.1016/j.renene.2015.11.056.
[17] B. Tian, W. Yang, L. Luo, J. Wang, K. Zhang, J. Fan, J. Wu, T. Xing, Synergistic enhancement of thermal conductivity for expanded graphite and carbon fiber in paraffin/EVA form-stable phase change materials, Sol. Energy 127 (2016) 48–55, https://doi.org/10.1016/j.solener.2016.01.011.
[18] L. Li, G. Wang, C. Guo, Influence of intumescent flame retardant on thermal and flame retardancy of eutectic mixed paraffin/polypropylene form-stable phase change materials, Appl. Energy 162 (2016) 428–434, https://doi.org/10.1016/j.apenergy.2015.10.103.
[19] S. Mu, J. Guo, Y. Yu, Q. An, S. Zhang, D. Wang, S. Chen, X. Huang, S. Li, Synthesis and thermal properties of cross-linked poly(acrylonitrile-co-itaconate)/polyethylene glycol as novel form-stable change material, Energy Convers. Manag. 110 (2016) 176–183, https://doi.org/10.1016/j.enconman.2015.12.004.
[20] H. Zhang, Y. Zou, Y. Sun, L. Sun, F. Xu, J. Zhang, A novel thermal-insulating film incorporating microencapsulated phase-change materials for temperature regulation and nano-TiO 2 for UV-blocking, Sol. Energy Mater. Sol. Cells 137 (2015) 210–218, https://doi.org/10.1016/j.solmat.2015.02.018.
[21] I. Bonadies, A.I. Renzi, M. Cocca, M. Avella, C. Carfagna, P. Persico, in: Heat storage and dimensional stability of poly (vinyl alcohol) based foams containing microencapsulated phase change materials 54(38), 2015, pp. 9342–9350, https://doi.org/10.1021/acs.iecr.5b02187.
[22] P. Chen, X. Gao, Y. Wang, T. Xu, Y. Fang, Z. Zhang, Metal foam embedded in SEBS/paraffin/HDPE form-stable PCMs for thermal energy storage, Sol. Energy Mater. Sol. Cells 149 (2016) 60–65, https://doi.org/10.1016/j.solmat.2015.12.041.
[23] A. Babapoor, G. Karimi, M. Khorram, Fabrication and characterization of nanofiber-nanoparticle-composites with phase change materials by electrospinning, Appl. Therm. Eng. 99 (2016) 1225–1235, https://doi.org/10.1016/j.applthermaleng.2016.02.026.
[24] Y. Zhou, D. Sheng, X. Liu, C. Lin, F. Ji, L. Dong, Synthesis and properties of crosslinking halloysite nanotubes/polyurethane- based solid-solid phase change materials, Sol. Energy Mater. Sol. Cells 174 (2018) 84–93, https://doi.org/10.1016/j.solmat.2017.08.031.
[25] S. Sundararajan, A.B. Samui, P.S. Kulkarni, Shape-stabilized poly(ethylene glycol) (PEG)-cellulose acetate blend preparation with superior PEG loading via microwave-assisted blending, Sol. Energy 144 (2017) 32–39, https://doi.org/10.1016/j.solener.2016.12.056.
[26] X. Guo, J. Cao, Y. Peng, R. Liu, Incorporation of microencapsulated dodecanol into wood flour/high-density polyethylene composite as a phase change material for thermal energy storage, JMADE 89 (2016) 1325–1334, https://doi.org/10.1016/j.matdes.2015.10.068.
[27] Y. Tang, Y. Lin, Y. Jia, G. Fang, Improved thermal properties of stearyl alcohol/high density polyethylene/expanded graphite composite phase change materials for building thermal energy storage, Energy Build. 153 (2017) 41–49, https://doi.org/10.1016/j.enbuild.2017.08.005.
[28] Y. Tang, Y. Jia, G. Alva, X. Huang, G. Fang, Synthesis, characterization and properties of palmitic acid/high density polyethylene/graphene nanoplatelets composites as form-stable phase change materials, Sol. Energy Mater. Sol. Cells 155 (2016) 421–429, https://doi.org/10.1016/j.solmat.2016.06.049.
[29] H. Ke, Q. Wei, Use of MWNTs-COOH to improve thermal energy storage and release rates of capric–palmitic–stearic acid ternary eutectic/polyacrylonitrile form-stable phase change composite fibrous membranes, Polym. Eng. Sci. 59 (2019) E403–E411, https://doi.org/10.1002/pen.25001.
[30] C. Alkan, D. Kahraman, A. Gottschalk, U. Ramamoorthi, Ö. Alt, Y. Daml, A. Çetin, in: Polyvinyl alcohol-salt hydrate mixtures as passive thermal energy storage systems 91, 2016, pp. 1012–1017, https://doi.org/10.1016/j.egypro.2016.06.269.
[31] B. Tang, L. Wang, Y. Xu, J. Xiu, S. Zhang, Hexadecanol/phase change polyurethane composite as form-stable phase change material for thermal energy storage, Sol. Energy Mater. Sol. Cells 144 (2016) 1–6, https://doi.org/10.1016/j.solmat.2015.08.012.
[32] F. Tian, S. Zhang, M. Zhai, J. Sui, X. Lan, J. Gao, Thermal properties of nano-sized polyethylene glycol confined in silica gels for latent heat storage, Thermochim. Acta 655 (2017) 211–218, https://doi.org/10.1016/j.tca.2017.05.006.
[33] Y. Wang, D. Liang, F. Liu, W. Zhang, X. Di, C. Wang, A polyethylene glycol/hydroxyapatite composite phase change material for thermal energy storage, Appl. Therm. Eng. 113 (2017) 1475–1482, https://doi.org/10.1016/j.applthermaleng.2016.11.159.
[34] L. Jiesheng, Y. Yuanyuan, H. Xiang, Research on the preparation and properties of lauric acid/expanded perlite phase change materials, Energy Build. 110 (2016) 108–111, https://doi.org/10.1016/j.enbuild.2015.10.043.
[35] L. Zhang, W. Yang, Z. Jiang, F. He, K. Zhang, J. Fan, J. Wu, Graphene oxide-modified microencapsulated phase change materials with high encapsulation capacity and enhanced leakage-prevention performance, Appl. Energy 197 (2017) 354–363, https://doi.org/10.1016/j.apenergy.2017.04.041.
[36] J. Liu, L. Chen, X. Fang, Z. Zhang, Preparation of graphite nanoparticles-modified phase change microcapsules and their dispersed slurry for direct absorption solar collectors, Sol. Energy Mater. Sol. Cells 159 (2017) 159–166, https://doi.org/10.1016/j.solmat.2016.09.020.
[37] S. Phromma, T. Wutikhun, P. Kasamechonchung, T. Eksangsri, C. Sapcharoenkun, Effect of calcination temperature on photocatalytic activity of synthesized TiO2 nanoparticles via wet ball milling sol-gel method, Appl. Sci. 10 (2020) 1–13, https://doi.org/10.3390/app10030993.
[38] F.H. Falqi, O.A. Bin-Dahman, M. Hussain, M.A. Al-Harthi, Preparation of miscible PVA/PEG blends and effect of graphene concentration on thermal, crystallization, morphological, and mechanical properties of PVA/PEG (10wt.%) blend, Int. J. Polym. Sci. (2018), https://doi.org/10.1155/2018/8527693.
[39] L. Feng, P. Song, S. Yan, H. Wang, J. Wang, The shape-stabilized phase change materials composed of polyethylene glycol and graphitic carbon nitride matrices, Thermochim. Acta 612 (2015) 19–24, https://doi.org/10.1016/j.tca.2015.05.001.
[40] X. Jin, J. Li, P. Xue, M. Jia, Preparation and characterization of PVC-based form-stable phase change materials, Sol. Energy Mater. Sol. Cells 130 (2014) 435–441, https://doi.org/10.1016/j.solmat.2014.07.013.
[41] A. Alsaad, A.R.Al Dairy, A. Ahmad, I.A. Qattan, S.Al Fawares, Q. Al-Bataineh, Synthesis and characterization of polymeric (PMMA-PVA) hybrid thin films doped with TiO2 nanoparticles using dip-coating technique, Crystals 11 (2021) 1–21, https://doi.org/10.3390/cryst11020099.
[42] M.M. Rashid, B. Simončič, B. Tomšič, Recent advances in TiO2-functionalized textile surfaces, Surf.Interfaces 22 (2021), https://doi.org/10.1016/j.surfin.2020.100890.
[43] K. Suhailath, M.T. Ramesan, B. Naufal, P. Periyat, V.C. Jasna, P. Jayakrishnan, Synthesis, characterisation and flame, thermal and electrical properties of poly (n-butyl methacrylate)/titanium dioxide nanocomposites, Polym. Bull. 74 (2017) 671–688, https://doi.org/10.1007/s00289-016-1737-9.
[44] A. Afzal, M. Usama, I.Abdul Rashid, Z. Khalid, M. Mohsin, M.F. Shakir, A. Tariq, Effect of MgOH/TiO2 on flame retardancy and mechanical behavior of composite, Mater. Res. Express 6 (2019) 1–7, https://doi.org/10.1088/2053-1591/ab624b.
[45] C. Jiao, X. Chen, Synergistic effects of titanium dioxide with layered double hydroxides in EVA/LDH composites, Polym. Eng. Sci. 51 (2011) 2166–2170, https://doi.org/10.1002/pen.21984.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/361459321

# Flower Pollination Algorithm for Scheduling Optimization of Flexible Manufacturing System

**Article** · June 2022

CITATIONS
0

READS
19

**3 authors**, including:

Gayathri Devi Ka
Delhi Technological University
**7** PUBLICATIONS **1** CITATION

SEE PROFILE

**Some of the authors of this publication are also working on these related projects:**




optimisation for scheduling in FMS View project

All content following this page was uploaded by Gayathri Devi Ka on 22 June 2022.

The user has requested enhancement of the downloaded file.

RESEARCH ARTICLE OPEN ACCESS

# Flower Pollination Algorithm for Scheduling Optimization of Flexible Manufacturing System

Gayathri Devi K [1], RS Mishra [2], AK Madan [3]

Department of Mechanical and Production Industrial and Automobiles Engineering, Delhi Technological University, Delhi, India

[1] Research Scholar itzgaya@gmail.com [2,3] Professor

## Abstract:

Flexible manufacturing system (FMS) is a part of smart manufacturing and their implementation is on the rise in manufacturing industrial scenario. The need of FMS arises in order to meet the customized demands of consumers. It is highly complicated due to its flexibility, tools and AGVs which makes it unpredictable dynamic manufacturing environment. Finding an optimal schedule is relatively easy for small instance problems, whereas large realistic problems are difficult to solve. Recently, many swarm intelligence techniques have been developed and implemented to achieve an optimal schedule for FMS. Flower Pollination Algorithm (FPA) is one of the swarm intelligence techniques which mimics the behaviour of pollination of flowering plants. This paper proposes a Hybrid Flower Pollination Algorithm (HFPA) which is an amalgamation of classic FPA with Simulated Annealing (SA) to solve two objectives-1) to minimise the machine idle time and 2) to minimise the total penalty cost. The algorithm's efficiency is demonstrated through comparison of HFPA with other meta heuristics found in literature and it is concluded that HFPA gives better results in less computational time.

***Keywords* — *Flexible manufacturing system, Hybrid Flower Pollination algorithm, scheduling, multi-objective.***

## 1. Introduction

Flexibility is a major consideration in the design of manufacturing systems. Flexible manufacturing systems (FMSs) have been developed over the last two decades to help manufacturing industry move towards the goal of flexibility. An FMS comprises three principal elements: computer controlled machine tools; an automated transport system and a computer control system. An FMS combines high levels of flexibility with high productivity and low levels of work-in-process inventory. It may also allow unsupervised production. In order to achieve these desirable benefits the control system must be capable of exercising intelligent supervisory management. Scheduling is at the heart of the control system of an FMS. The development of effective and efficient FMS scheduling strategies remains an important and active research area. Scheduling problems of FMS are known to be NP-hard, which means that they are intractable and the computation time increases exponentially as the problem size goes up. Some assumptions have been made about scheduling problems in order to simplify the model and solve it. One of these assumptions is that the transferring times of jobs between different machines are negligible.

However, it is highly impractical not to model the material handling activities in scheduling problems in reality, especially when the movements of jobs are completely relying on the material handling equipment and their transferring times are comparable to the production times. Omission of transferring times will make the result of scheduling impossible for implementation.

### *1.1 Earlier Research*

The optimization of scheduling are usually done by different approaches spanning from mathematical approaches and branch and bound techniques to bottleneck based heuristics, artificial intelligence and by local search methods. This section of literature helped the research scholar in exploring the FMS and to understand the need for better scheduling. The summary of literature surveyed in this regard is given below.

Liu et al [1] gave a detailed review on FMS, its variations, its components, scheduling problems and different methodologies used to solve for FMS. The study made by Sharma [2] helps the practitioners to study about various issues encountered in FMS setup and it also helps in bridging the gaps between the various crucial aspects required for its implementation.

Wu et al [3] addressed a single-machine total completion time problem with learning effect and release times based on the sum of processing times. A simulated-annealing algorithm was also proposed to obtain a near optimal solution. They proposed for Future research may consider other criterion such as the tardiness or lateness or studying the problem in the multi-machine setting. Noorul Haq et al. [4] employed a Giffler and Thompson heuristic with six priority-dispatching rules for determining the best routing of the jobs and this optimum part routing is considered as an input to schedule the material handling system. In order to minimize the distance travelled by the AGV and transportation cost of the AGV, a hybrid approach of GA and simulated annealing is employed for obtaining quality solution. Jerald et al [5] implemented optimization approaches such as GA, SA, Mematic algorithm and PSOA for a multi objective function. The results show that PSOA is found to be better than other three algorithms and provides the minimum combined objective function value. Chawla et al. [6] evaluated the performance of dynamic job selection dispatching rules for simultaneous dispatching and scheduling of multi-load automated guided vehicles (AGVs) cruising in two different sizes of FMS. Zhao et al [7] proposed genetic algorithm (GA) to solve the job-sequencing problem for a production shop that is characterized by flexible routing and flexible machines. Saravanan [8] proposed a modified NSGA II to solve FMS scheduling and he demonstrated that his method is very effective. Pandey et al [9] presents a review on design and control of automated guided vehicle systems. Various types of scheduling problems are solved in different job shop environments, vehicle routing, guide-path design, vehicle dispatching. The prior objective of this paper is to extend previous research by examining the effects of scheduling rules and routing flexibility on the performance of a constrained and utilization of AGVs and machines. Prakash Babu et al [10] made an attempt to study the machine and vehicle planning features simultaneously in an FMS for minimization of the make-span.

Swarm intelligence techniques are gaining importance recently to solve large real life manufacturing problems. After a detailed study of literature, we aim to solve the FMS scheduling problem by Hybrid Flower pollination algorithm. The study made by Abdel-Baset and Hezam [11] provides a detailed working methodology of Flower pollination Algorithm (FPA) . The paper also discusses about various hybridizations used for FPA so far. Khursheed et al. [12] is another review article that details about the variations prosed for FPA and its applications. This article also serves as good base paper. [13] solved a scheduling problem by disparity count process with FPA. [14] solved flexible job shop problem (FJSP) by

FPA which includes fuzzy processing time. Sivarami Reddy et al [15] solved the simultaneous scheduling of machines and AGVs using FPA. And demonstrated the efficiency of the proposed algorithm. Therefore after a thorough analysis of recent literature from last decade, this paper aims to solve a FMS scheduling problem for 80 jobs and 16 machines with multi objectives of minimizing the machine idle time and minimizing the total penalty cost. The results have been compared with other known metaheuristics and conclusions are drawn.

## 2. Problem Description

The problem environment, assumption and aim of the present work are as follows:

1. The FMS considered in this work has a configuration as shown in Fig. 1. There are five flexible machining cells (FMCs), each with two to six computer numerical machines (CNCs), an independent and a self-sufficient tool magazine, one automatic tool changer (ATC) and one automatic pallet changer (APC). Each cell is supported by one to three dedicated robots for intra-cell movement of materials between operations. There is a loading station from which parts are released in batches for manufacturing in the FMS. There is an unloading station where the finished parts are collected and conveyed to the finished storage. There is one automatic storage and retrieval system (AS/RS) to store the work in progress. The five FMCs are connected by two identical automated guided vehicles (AGVs). These AGVs perform the inter cell movements between the FMCs, the movement of finished product from any of the FMCs to the unloading station and the movement of semi-finished products between the AS/RS and the FMCs.

2. The assumptions made in this work are as follows: There are 40 to 80 varieties of products for a particular combination of tools in the tool magazines. Each type/variety has a particular processing sequence batch size, deadline and penalty cost for not meeting the deadline. Each processing step has a processing time with a specific machine. The problem matrix taken is given in table 1.

3. The objective of the schedule is the combination of minimizing the machine idle time and minimizing the total penalty cost.

$$COF = (w1)^* \frac{Total\ penalty\ cost}{\text{Maximum permissible penalty}} + (w2)^* \frac{\text{Total machine idle time}}{\text{Maximum total machine elapsed time}} \quad (1)$$

Where w1and w2 are the weights assigned to each objective function.

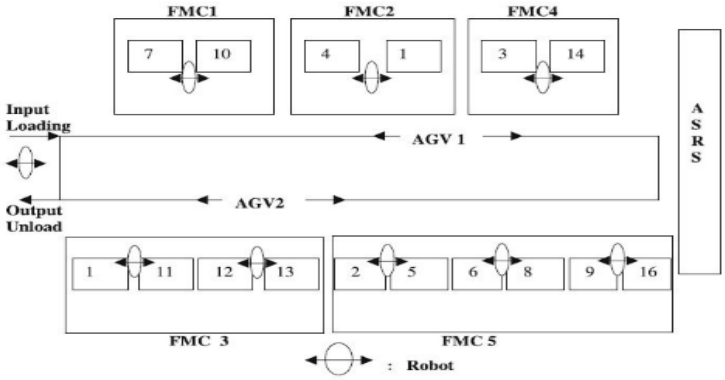


Fig 1. FMS Structure

## 3. Proposed Methodology

### 3.1 Classic FPA

Flower pollination is a process related to transferring flowers' pollens. Transferring flowers' pollens are carried by birds, bats, insects and other animals. There are two categories of pollination i.e., abiotic and biotic. In biotic pollination, pollinators transfer pollen and abiotic pollination does not require pollinators. Some pollinators tend to visit some specific type of flowers and at the same time, other species of flowers will be bypassed. This phenomenon is known as flower constancy. All the flowers with flower constancy property guarantee the reproduction maximisation.
Pollination can be achieved through cross- pollination or self-pollination. In cross-pollination, pollens are transferred from a different plant (Yang, 2012).
The cross pollination takes place at long distances and is usually carried by pollinators like bats, bees, butterflies and they are considered as discrete jumps that obey Levi distribution.

The following rules idealizes the FPA

- biotic and cross-pollination is interpreted as global pollination where the pollinators carrying pollens performs levy flights
- the abiotic and self-pollination can be recognized as local pollination
- flower constancy can be considered as a reproduction capability and is proportional to the similarity of the two flowers involved
- due to the wind and physical proximity, local pollination has a little advantage over global pollination.

Rule 1 can be mathematically represented as shown in Eq 2

$$X_i^{t+1} = X_i^t + L\,(X_i^t - g*) \qquad (2)$$

where $X_i^t$ is the individual flower at generation t, L denotes step size from the levy distribution, and g* is the best solution of all.

Rule 2 and 3 can be represented as shown in Eq 3

$$X_i^{t+1} = X + U\,(X_j^t - X_k^t) \qquad (3)$$

Where $X_j^t$ and $X_k^t$ are pollen from different flowers of the same plant species. This essentially imitates the flower constancy in a limited neighborhood. Mathematically, if $X_j^t$ and $X_k^t$ comes from the same species or selected from the same population, this equivalently becomes a local random walk if we draw U from a uniform distribution in [0, 1].

Switch probability p € [0, 1] controls local pollination and global pollination.

If a randomly generated number is less than Switch probability p, global pollination takes place otherwise local pollination takes place. From the literature study done, it is proven in various studies that p=0.8 provides a better solution. The pseudo code of

classic FPA is shown in Fig 2.

Objective $f(x)$, $x=(x_1,...,x_d)^T$
Initialize a population of $n$ flowers at random positions as cluster centres
Compute best solution $\boldsymbol{g}_*$ from the initial population
Define a switching probability $p \in [0,1]$
$t = 0$
while $t<Maximum_Iteration$
    for $i = 1: f$ (all flowers in population)
        if $rand < p_s$
    Form a $d$-dimensional step vector $L$ [by drawing from Lévy distribution]
    Do global pollination $x_i^{t+1} = x_i^t + L.(x_i^t - g_*)$
  else
    Draw $\in$ from random distribution [0,1]
    Randomly select $x_j$ and $x_k$ from the population
    Do local pollination via (2)
  end if
    Evaluate latest generated solution $x_i^{t+1}$
    Replace $x_i^t$ by $x_i^{t+1}$ if the new solution is better
 end for
Update present best solution $g_*$
$t = t + 1$
end while

Fig 2 Pseudo code for classic FPA

### *3.2 Hybrid Flower Pollination Algorithm Methodology (HFPA)*

In the proposed algorithm, we have incorporated Simulated Annealing with classic FPA to excel its performance and also to avoid local optima entrapment. The Simulated annealing (SA) is one of the simplest and most popular heuristic algorithms. SA is a global search algorithm, based on annealing process of metal processing. It has been proved that it can have global convergence, though the convergence rate can be very slow. It is based on Monte Carlo iterative solution strategy. Its main advantage is that it not only accepts better solution than the current state, but also it can jump out of local minimum.

*3.2.1 Working of HFPA*

The initial solution is randomly generated in the standard simulated annealing algorithm. Therefore, the size of the solution is uneven. This feature will affect the effectiveness of algorithm. The defect can be avoided by using flower pollination algorithm to create the initial solution. In order to enhance the searching efficiency, the new solution is generated by FP algorithm. The worst solution is replaced by the new solution. The advantage of the searching process lies in retaining the intermediate optimal solution and updating on time. Finally, annealing process is executed once again based on the final optimal solution. Since FPA works on continuous problems it has to be discretized before it is implemented for scheduling FMS. This paper adopted the discretization method proposed by [14].

The parameters used in this study are; i) the population size (number of flowers) = 50, ii) selection probability of pollination method p =0.8, and iii) maximum iterations = 5000

## 4. Results and Discussions

The FMS configuration considered in this work is taken from existing literature [16]. The proposed algorithm has been

implemented for 80 jobs for 16 machines using combined objective optimization method. The problem matrix taken for this study can be found in [16]. In the experiments conducted, we have done simulations for different weights. (i) Simulation 1: w1 = 0.5 w2 = 0.5 (ii) Simulation 2: w1= 0.6 w2= 0.4 (iii) Simulation 3: w1= 0.4 w2= 0.6. HFPA has been run for different weights in the COF formula and the results are tabulated in Table 1. The optimal solution found by varying weights are highlighted in bold letters. Matlab software version R2016b and above is used for programming the proposed approach. The use of matlab enables us to solve complex scheduling problems involving different job types and multiple machines. TOMLAB/CPLEX has also been used in our program as it's a subproblem solver and handles quicker solving of mixed-integer linear and quadratic programming (MILP,MIQP), and linear and quadratic programming (LP,QP). Different optimal schedules are obtained for the FMS using the above hybrid approach, and the performances are compared and analysed. A comparison between the proposed HFPA and other algorithms, namely GA [5], SA [16] , Memetic Algorithm, GAPSOTS [17] (found in literature) have been shown in Table 2. The proposed algorithm has been run for different iterations. When the iterations number are less the results were not good, but as iterations increased, we got better results. We stopped at 5000 iterations as there was no significant change in optimal value later.

Table 1: COF for different weights

| PROBLEM | W1 0.5, W2 0.5 | W1 0.6 W2 0.4 | W1 0.4, W2 0.6 |
|---|---|---|---|
| 80 J X 16 M | **0.075676** | 0.08132 | 0.12037 |

Table 2 Comparison of results for various techniques with GAPSOTS

| Problem Size | Objectives | Genetic Algorithm | Simulated Annealing | Memetic Algorithm | GAPSOTS | HFPA |
|---|---|---|---|---|---|---|
| 80 jobs x 16 machines | COF | 0.0959237 | 0.14738715 | 0.1067266 | 0.08537 | 0.075676 |
| | Sequence | 34,50,59,76, 48,69,31,11, 1,26,56,23,4 7,25,62,20,3, 78,55,53,24, 51,70,71,75, 46,58,74,4,2, 28,18,38,22, 54,5,72,12,1 3,40,10,60,6 5,79,43,37,6 6,27,42,7,64, 61,29,35,36, 73,39,49,67, 2,41,15,17,4 4,45,63,21,6, 80,9,19,57,1 4,33,16,68,5 2,77 ,8,32 | 41,53,51,70,1,7 1,47,58,36,61,3 0,4,23,54,75,77, 76,52,79,31,15, 78,35,59,66,72, 55,56,64,28,40, 13,45,17,14,48, 80,49,34,25,10, 20,44,29,11,60, 67,38,27,9,12,3 2,46,5,18,43,65, 37,24,21,42,2,6 3,19,7,22,33,74, 26,57,16,62,3,3 9,69,6,50,73,68, 8 | 69,17,36,2 4,74,61,20, 35,75,49,2 5,18,58,47, 3,52,28,45, 51,59,46,6, 16,34,31,5 3,70,76,22, 2,40,79,15, 30,62,54,3 8,10,43,13, 80,37,71,2 6,66,73,72, 65,56,67,1 2,1,4,5,21, 32,77,27,2 9,63,33,60, 64,57,41,9, 14,48,55,7 8,7,42,11,2 3,39,8,50,6 | 39,26,8,32, 35, 78,59,27,51, 18,19,31 52,71,57 5,1,74,22 13,63,58 33,34,65 29,48,42,70, 17,9,56,40,4 9,67,30, 3,66,7,45,11, 23,80,38,16, 46,44,21,60, 75,24,10,4,5 3,47,15,76,2 0,6,54,72,43 64,50,2,14,7 3,25,69,77,4 1,28,79,61,5 5,68,36,12,3 | 35,75,49, 69,17,36,24, 74,61,20, 25,18,58,47, 52,28,45,51, 59,46,6, 16,34,31,53, 70,76,22,2,4 0,79,15, 30,62,54,38, 10,43,13,80, 37,71,26,66, 73,72,65,56, 67,12,1,4,5,2 1,32,77,27,2 9,63,33,60,6 4,57,41,9,14, 48,55,78,7,4 2,11,23,39,8, 50,68 ,19,44 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | 8 ,19,44 | 7,62 | |

## 5. Conclusions

Optimization procedure has been developed in this work which is based on genetic algorithm and is implemented successfully for solving the scheduling optimization problem of FMS. Software has been written in Matlab. Results are obtained for the 80 jobs and 16 machines FMS system. With less computational effort it is possible to obtain the solution for such a large number of jobs (80) and machines (16).This work leads to the conclusion that the procedures developed in this work can be suitably modified to any kind of FMS with a large number of components and machines subject to multi objective functions. Future work will include availability and handling times of loading/unloading stations, AGVs.

References


[1] *J. Liu and B. L. Maccarth, "The classification of FMS scheduling problems," International Journal of Production Research, vol. 34, no. 3, pp. 647–656, 1996, doi: 10.1080/00207549608904925.*

[2] *R. Sharma, "Implementation Issues in FMS : A Literature Review," International Journal of Innovations in Engineering and Technology, vol. 2, no. 2, 2013.*

[3] *C. C. Wu, P. H. Hsu, and K. Lai, "Simulated-annealing heuristics for the single-machine scheduling problem with learning and unequal job release times," Journal of Manufacturing Systems, vol. 30, no. 1, pp. 54–62, 2011, doi: 10.1016/j.jmsy.2011.03.004.*

[4] *A. Noorul Haq, T. Karthikeyan, and M. Dinesh, "Scheduling decisions in FMS using a heuristic approach," International Journal of Advanced Manufacturing Technology, vol. 22, no. 5–6, pp. 374–379, 2003, doi: 10.1007/s00170-002-1474-0.*

[5] *J. Jerald, P. Asokan, G. Prabaharan, and R. Saravanan, "Scheduling optimisation of flexible manufacturing systems using particle swarm optimisation algorithm," International Journal of Advanced Manufacturing Technology, vol. 25, no. 9–10, pp. 964–971, 2005, doi: 10.1007/s00170-003-1933-2.*

[6] *V. K. Chawla, A. K. Chanda, and S. Angra, "Simultaneous Dispatching and Scheduling of Multi-Load AGVs in FMS-A Simulation Study," in Materials Today: Proceedings, 2018, vol. 5, no. 11, pp. 25358–25367. doi: 10.1016/j.matpr.2018.10.339.*

[7] *C. Zhao and Z. Wu, "A genetic algorithm approach to the scheduling of FMSs with multiple routes," International Journal of Flexible Manufacturing Systems, vol. 13, no. 1, pp. 71–88, 2001, doi: 10.1023/A:1008148313360.*

[8] *R. Saravanan, "Fms Scheduling Optimization Using Modified Nsga-Ii," International Journal of Mechanical and Production Engineering, vol. 2, no. 2, pp. 1–6, 2014.*

[9] *R. Pandey and A. Singh, "Utilization of AGVs and Machines in FMS Environment," Journal of Material Science & Engineering, vol. 5, no. 4, 2016, doi: 10.4172/2169-0022.1000263.*

[10] *K. Prakash Babu, V. Vijaya Babu, and N. R. Medikondu, "Implementation of heuristic algorithms to synchronized planning of machines and AGVs in FMS," Management Science Letters, vol. 8, pp. 543–554, 2018, doi: 10.5267/j.msl.2018.5.001.*

[11] *M. Abdel-Baset and I. Hezam, "A Hybrid Flower Pollination Algorithm for Engineering Optimization Problems," International Journal of Computer Applications, vol. 140, no. 12, pp. 10–23, 2016, doi: 10.5120/ijca2016909119.*

[12] *M. U. N. Khursheed et al., "Review of Flower Pollination Algorithm: Applications and Variants," 2020 International Conference on Engineering and Emerging Technologies, ICEET 2020, pp. 1–6, 2020, doi: 10.1109/ICEET48479.2020.9048215.*

[13] *A. Phuang, "The flower pollination algorithm with disparity count process for scheduling problem," 2017 9th International Conference on Information Technology and Electrical Engineering, ICITEE 2017, vol. 2018-Janua, pp. 1–5, 2017, doi: 10.1109/ICITEED.2017.8250497.*

[14] *W. Xu, Z. Ji, and Y. Wang, "A flower pollination algorithm for flexible job shop scheduling with fuzzy processing time," Modern Physics Letters B, vol. 32, no. 34–36, 2018, doi: 10.1142/S0217984918401139.*

[15] *Dr. R. D. and Dr. P. R. K. Sivarami Reddy N, "Simultaneous Scheduling of Machines and Agvs Using Flower Pollination Algorithm: a New Nature-Inspired Meta-Heuristic," Manufacturing Technology Today, vol. 17, no. 7, pp. 19–30, 2018.*

[16] *Dr. R. S. Nidhish Mathew Nidhiry, "A COMBINED OBJECTIVE GENETIC ALGORITHM FOR SCHEDULING OF MACHINES IN FMS," International Journal of Engineering Research and Applications (IJERA) ISSN:, vol. 3, no. 2, 2012.*

[17] *K. Gayathri Devi, R. S. Mishra, and A. K. Madan, "Combined Objective Optimization of FMS Scheduling by a Hybrid Genetic Algorithm," in Advances in Mechanical and Materials Technology, 2022, pp. 125–139.*




View publication stats

2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) | 978-1-6654-1120-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAECT54875.2022.9807907

# Food Review Analysis and Sentiment Prediction using Machine Learning Models

Dhruv Gupta
*Electronics and Communication Engineering Department, Delhi Technological University*,
Delhi, India

Ausho Roup
*Electronics and Communication Engineering Department, Delhi Technological University,*
Delhi, India

Diksha Gupta
*Electronics and Communication Engineering Department, Delhi Technological University*,
Delhi, India

Avinash Ratre
*Electronics and Communication Engineering Department, Delhi Technological University*,
Delhi, India

***Abstract*—In this era of the digital world, text, messages, comments, numbers and videos have become an essential source of information. The trend of people trading through e-commerce giants like Amazon and Flipkart is proliferating. It's necessary to have a model or tool that helps retrieve helpful information from the customers' online reviews quickly that can also help product manufacturers have a better idea of their product. This paper targets the food industry, and a model is proposed that analyzes the customer reviews based on NLP techniques- TF-IDF Vectorizer and Count Vectorizer. Based on these analyses, customer sentiments are predicted using different machine learning classification algorithms like Logistic regression, Dummy classifier and Random forest classifier.**

***Keywords—review analysis; sentiment analysis; data mining; natural language processing; machine learning***

## I. INTRODUCTION

The Internet has become a popular platform for genuine reviews, opinions, emotions, and conclusions. The trend of people trading through e-commerce websites is proliferating. In this digital world, people usually prefer online shopping as an easy, fast and time-consuming way. Various e-commerce websites have provided options to consumers to share their experiences and post their reviews or feedback on the products they bought. Before buying a product online, consumers prefer to check the product reviews from previous buyers'. So, online reviews play a significant role for all the consumers to make a final decision about the product. These online reviews are crucial for the manufacturers, which help them make effective and productive marketing decisions. Based on customers' reviews, the manufacturers can get an idea about the competitive market, and they can design a new product or modify the existing one based on the customer's needs and requirements.

This paper aims at making a model for analyzing the customer reviews using different techniques of Natural language processing (NLP), and then based on those analyses, customer sentiments are predicted using two different machine learning classification methods. This can help the manufacturer to understand the customer sentiments towards their products and make better and more effective decisions for their products.

This paper utilizes two NLP techniques instead of one to get more accurate results on a large scale dataset containing food reviews, which makes the food industry the target for applying this model. The Frequency-inverse document frequency (TF-IDF) vectorizer and count vectorizer both have their importance. Here, a Count vectorizer is used to get the most frequent words in the document, whereas the TF-IDF Vectorizer is used to get the overall insight of a word and ignore the familiar words.

## II. RELATED WORK

Various research papers have been published in this area related to sentiment analysis or opinion mining [1]. Most of them focused on doing sentiment analysis using various deep learning and machine learning techniques on e-commerce reviews like Amazon and Flipkart, and classifying the text as good or bad. This paper compares the existing research articles using the state-of-art method.

Pankaj *et al.* [2] performed the comparative sentiment analysis of retrieved reviews of various smartphone opinions on Amazon and divided them into positive, negative, and neutral behaviour sentiments. Similarly, D'souza *et al*. [3] applied the same approach as done in [2], but they used dual sentiment analysis to divide the text into positive, negative, and neutral sentiments.

The authors in [4] proposed sentiment analysis of amazon books' reviews using a powerful analysis language R and then applied visualization techniques to provide a relationship between the old and the latest reviews. Santhosh Kumar *et al.* [5] used algorithms such as Naïve Bayes classifier, Logistic regression, and SentiWordNet to classify the review as a positive and negative reviews and computed the respective performance. Xiao *et al.* [6] applied the TextBlob-NLP library in python for sentiment analysis of amazon reviews on beauty and baby products and later classified them into positive, negative, and neutral emotions. Singla *et al.* [7] did sentiment analysis of customers' reviews on mobile phones on amazon, and they not only classified the text into positive and negative sentiments but also included the sentiments of fear, anger, joy, sadness etc. Haque *et al.* [8] used supervised machine learning techniques on a large scale amazon dataset to centralize it and get satisfactory

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:45:04 UTC from IEEE Xplore. Restrictions apply.

accuracy. The authors in [9] discussed the performance accuracy of different classification algorithms like logistic regression, decision tree, and naive Bayes using principal component analysis (PCA) on the Amazon food review dataset and calculating the confusion matrix.

This paper differs from the rest as it performs review analysis using two NLP techniques, the TF-IDF vectorizer and Count vectorizer, respectively, and predicts the customer sentiments using three different classification algorithms.

## III. DATASET DESCRIPTION

This paper aims at predicting the score, which represents the sentimental prediction using the text column, which represents the reviewers' reviews. This model works on a large-scale dataset of online reviews, taken from e-commerce giant Amazon [10], comprising 5 68,455 rows of users' reviews on food and 10 attributes.

## IV. METHODOLOGY

This section discusses the details about the proposed model for predicting customer sentiments. Figure 1 shows the flow diagram for the proposed model.

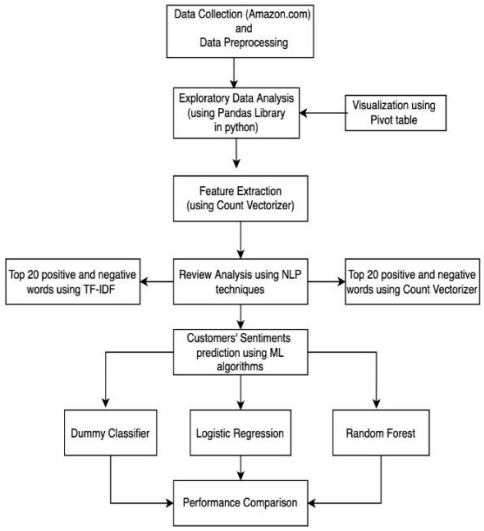



Fig. 1. Data flow of the proposed method.

The proposed prediction method is composed of four phases: data pre-processing, exploratory data analysis, feature extraction, and machine learning algorithm. These four phases are briefly introduced in the following subsections.

### A. Data Pre-processing

This process includes removing duplicate, corrupted or incomplete data and fixing the incorrectly formatted or preparing the data for the model in the dataset.

- The "Helpful%" column values are calculated from the "HelpfulNumerator" and "HelpfulDenominator" column values using the where() function from the NumPy library..
- The "Helpful%" column is then used to calculate the "%upvote" column values by using the cut() function from the panda's library.
- The score column has 5 unique values, {1,2,3,4,5}, where '3' corresponds to the neutral score. Since there is no prediction for the neutral score, all records with a score '3' are removed before the feature selection.

### B. Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an approach to analyzing datasets to discover patterns, test hypotheses, and check assumptions, often using statistical graphics and other data visualization methods. This section performs analysis over the dataset.

- This paper analyses the upvotes for different scores by grouping the data in the score and %upvote column using the groupby() function from the panda's library.
- The above analyses are further analyzed by finding the number of users associated with each score and upvote. This is visualized in a pivot table as shown in Table I.

TABLE I. Pivot Table

| Score | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| %Upvote | | | | | |
| Empty | 8060 | 4234 | 5062 | 4780 | 11638 |
| 0-20% | 2338 | 762 | 474 | 116 | 432 |
| 20-40% | 4649 | 1618 | 1506 | 909 | 2275 |
| 40-60% | 6586 | 3051 | 3384 | 3185 | 10312 |
| 60-80% | 5838 | 2486 | 2754 | 2941 | 11060 |
| 80-100% | 12531 | 7014 | 11037 | 26707 | 140661 |

### C. Feature Extraction

The dataset initially contains 10 features.

- Feature Extraction is done in the dataset using the CountVectorizer() function from sklearn.feature_extraction.text responsible for performing the bag of words.
- Then applying the fit_transform() function over the CountVectorizer() using the text column data creates 114969 features to train and test the data.

### D. Machine Learning Classification Algorithm

This section focuses on applying the NLP techniques, using CountVectorizer() function, over different machine learning algorithms, finding the accuracy for the same, and determining the top positive and negative words identified by the algorithm. This whole process is automated for each machine learning classification algorithm used. Different classification algorithms used in this model are as follows.

*1) NLP Techniques:* The NLP technique [11] helps machines interpret, analyze and manipulate human

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:45:04 UTC from IEEE Xplore. Restrictions apply.

languages. Its applications include speech recognition, sentiment analysis or opinion mining, text classification, etc. This paper uses TF-IDF Vectorizer and Count Vectorizer for the sentiment analysis [12]. The TF-IDF technique is popular in text mining or information retrieval. Here, the words are quantified from a given set of documents, and each word is assigned a score telling the importance of that word in the whole document. The term "Frequency" means determining the number of occurrences of a particular word in the given document. The frequency (TF) is calculated as

$$tf(t,d) = \text{count of t in d/number of words in d} \quad (1)$$

where the parameters $t$ and $d$ are the term (word) and document (set of words), respectively. The Document Frequency (DF) determines the number of documents in which a term t occurs. In other words, it counts the occurrences of a word from a given document set N and is calculated as

$$\text{df(t)} = \text{occurrence of t in N documents}, \quad (2)$$

where the parameter $N$ is the number of documents.

The Inverse document frequency (IDF) is the reverse of document frequency, and it calculates the informativeness of a word, and it is represented as

$$\text{idf(t)} = \text{N/df}. \quad (3)$$

The problem with the above formula is that when the value of N becomes too large, the value of IDF explodes. So to overcome that scenario, log of IDF is taken as

$$\text{idf(t)} = \log(\text{N}/(\text{df} + 1)). \quad (4)$$

The basic formula of TF-IDF vectorizer is represented as

$$\text{tf} - \text{idf(t, d)} = \text{tf(t, d)} \,.\, \log(\text{N}/(\text{df} + 1)) \quad (5)$$

The count vectorizer is a python tool given by Scikit-learn library. This method transforms the given text into a vector containing the frequency or count of each word present in the text. Top positive and negative words determined using CountVectorizer () function are shown in Tables II and III, respectively. The positive and negative coefficient generated corresponds to the weightage of the respective positive and negative words.

TABLE II. Top positive words using count vectorizer.

| Word | Coefficient |
|---|---|
| pleasantly | 3.97 |
| downside | 3.25 |
| addicting | 3.01 |
| skeptical | 2.87 |
| delish | 2.58 |
| resist | 2.3 |
| trainer | 2.29 |
| welcome | 2.28 |
| hooked | 2.26 |
| intend | 2.25 |
| drawback | 2.23 |
| terrific | 2.17 |
| worries | 2.16 |
| hesitant | 2.13 |
| easiest | 2.12 |
| worried | 2.08 |
| tastiest | 2.05 |
| whim | 2.04 |
| execellent | 2.03 |
| saves | 1.99 |

TABLE III. Top negative words using count vectorizer.

| Word | Coefficient |
|---|---|
| ruined | -2.33 |
| canceled | -2.35 |
| unappealing | -2.38 |
| vomiting | -2.47 |
| glowing | -2.51 |
| lousy | -2.55 |
| embarrassed | -2.57 |
| defeats | -2.61 |
| overpowers | -2.69 |
| ripoff | -2.74 |
| disapoint | -2.75 |
| lacked | -2.81 |
| disappointing | -2.85 |
| mdiocre | -3.01 |
| returnable | -3.09 |
| unacceptable | -3.17 |
| cancelled | -3.18 |
| worst | -3.23 |
| undrinkable | -3.71 |
| deceptive | -3.9 |

Top positive and negative words determined using the TF-IDF function are shown in Tables IV and V, respectively. The positive and negative coefficient corresponds to the respective positive and negative words.

TABLE IV. Top positive words using TF-IDF vectorizer.

| Word | Coefficient |
|---|---|
| great | 14.15 |
| delicious | 12.58 |
| best | 11.89 |
| perfect | 10.72 |
| excellent | 9.88 |

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:45:04 UTC from IEEE Xplore. Restrictions apply.

| | |
|---|---|
| loves | 9. 73 |
| highly | 8.85 |
| wonderful | 8. 24 |
| love | 8. 21 |
| amazing | 7.58 |
| hooked | 7.48 |
| good | 7.47 |
| awesome | 7.31 |
| pleasantly | 7. 25 |
| nice | 7.07 |
| favorite | 7 |
| pleased | 6.96 |
| yummy | 6.82 |
| glad | 6.57 |
| smooth | 6.54 |

TABLE V. Top negative words using TF-IDF vectorizer.

| **Word** | **Coefficient** |
|---|---|
| poor | -5.72 |
| yuck | -5.79 |
| hopes | -5.86 |
| worse | -6.33 |
| stale | -6.34 |
| undrinkable | -6.39 |
| bland | -6.42 |
| tasteless | -6.47 |
| disgusting | -6.59 |
| weak | -6.83 |
| return | -6.89 |
| threw | -6.96 |
| unfortunately | -8.01 |
| horrible | -8.02 |
| disappointment | -8.13 |
| disappointed | -8.37 |
| awful | -8.78 |
| terrible | -9.24 |
| disapointing | -9.95 |
| worst | -11.86 |

*2) Logistic Regression:* Logistic regression [13] is a supervised learning technique that solves classification problems. With a given set of independent variables, this algorithm predicts the output of a categorical dependent variable. The output is a discrete value that means True/False, Yes/No, 0 or 1, etc. It gives the probabilistic values between 0 and 1 instead of giving exact 0 and 1. Unlike linear regression, this model don't fit a straight line to our data. Instead, it fit an "S" shaped logistic function, called sigmoid, to our observations. The sigmoid function represents as $1/(1+e^{-x})$ and it predicts two maximum values (0 or 1). The curve shows the likelihood of whether the cells are cancerous or not.

Logistic regression is used to predict the scores or sentiments after analyzing the customers' reviews. The text is classified as a true negative or true positive and false negative or false positive.

Hypothesis Representation: When using linear regression, formula used for the hypothesis is given as

$$h\Theta(x) = \beta_0 + \beta_1 X\,, \qquad (6)$$

where $h\Theta$ is the hypothesis function, $\beta_0$ is the constant, and $\beta_1$ represents the slope.

For the logistic regression, the above equation is slightly modified as

$$\sigma(Z) = \sigma(\beta_0 + \beta_1 X)\ , \qquad (7)$$

where the parameter $\sigma(Z)$ represents the hypothesis function using the sigmoid function

It is expected that this hypothesis varies between 0 and 1, and represented as

$$Z = \beta_0 + \beta_1 X\,, \text{and} \qquad (8)$$

$$h\Theta(x) = \text{sigmoid}(Z) = \frac{1}{(1+e^{-((\beta_0+\beta_1 X))})}\ . \qquad (9)$$

OUTPUT: Figures 2 and 3 show the confusion matrix for Logistic regression using TF-IDF vectorizer and count vectorizer, respectively.

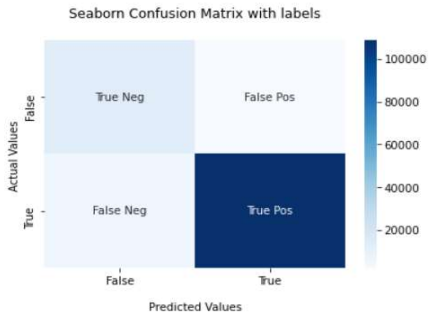



Fig. 2. Confusion Matrix using TF-IDF Vectorizer (using Logistic regression).

This shows that this algorithm computes a total of 14342 and 108595 correct predictions.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:45:04 UTC from IEEE Xplore. Restrictions apply.

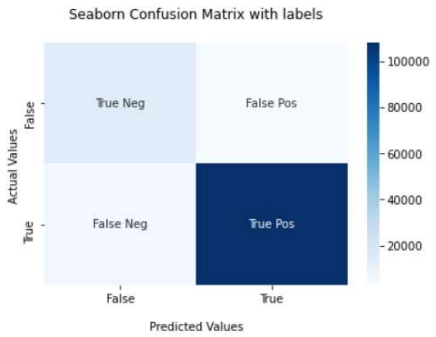



Fig. 3. Confusion matrix using Count Vectorizer (using Logistic regression).

This shows that this algorithm computes a total of 14852and 108239 correct predictions.

*3) Dummy Classification:* The Dummy classifier [14] classifies the given dataset using straightforward and easy rules. It does not give any information about the data, and it doesn't depend on training data because it ignores the trends in the training dataset. This classifier is useful for only those datasets where there is an imbalance of classes. Other classifiers use it as a baseline. This classifier is used to establish a baseline model to check and compare the performance of other models. It makes the predictions based on the most frequent label in the training dataset. This classifier uses one of the following strategies to predict the class label.

- Most frequent: The classifier predicts the class label that occurs most frequently in the training dataset.
- Uniform: The classifier makes predictions uniformly.
- Stratified: Each data point is associated with a probability of being a frequently occurred class label.
- Constant: A constant label is always predicted by the classifier. This strategy is mainly used for non-majority class labels.

The basic syntax is as under: class sklearn.dummy.DummyClassifier(*, strategy='most frequent', random_state=None, constant=None).

OUTPUT: Figures 4 and 5 show the confusion matrix for the Dummy classifier using TF-IDF vectorizer and count vectorizer, respectively.

Fig. 4. Confusion matrix using TF-IDF Vectorizer (using Dummy classifier).

This shows that this algorithm makes a total of 110878 correct predictions.

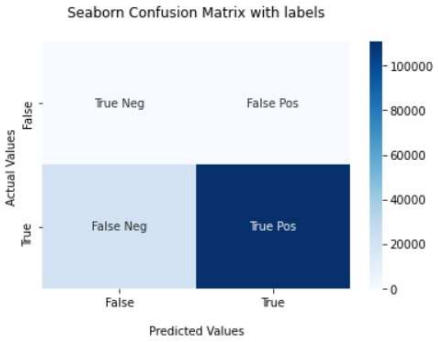



Fig. 5. Confusion matrix using count vectorizer (using Dummy classifier).

This shows that this algorithm computes a total of 110943 correct predictions.

4) *Random Forest Classification:* Random Forest [15] is a supervised learning technique used for both Regression and Classification problems. It builds decision trees on different subsets of the given dataset and takes their average in case of regression and majority votes for classification to enhance the prediction reliability of that dataset and predict the final output. As the number of trees in the forest increases, the accuracy rate increases, preventing the problem of overfitting. Random forest is used to predict the scores on a large scale dataset because it gives higher accuracy. It can also estimate missing values on a larger dataset as well.

**OUTPUT**: Figures 6 and 7 show the confusion matrix for Random forest using TF-IDF vectorizer and count vectorizer, respectively.

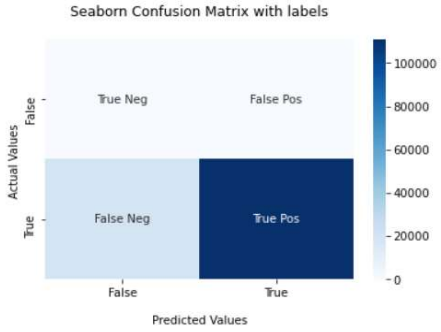



Fig. 6. Confusion Matrix using TF-IDF Vectorizer (using Random forest classifier)

This shows that this algorithm computes a total of 56 and 111031 correct predictions.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:45:04 UTC from IEEE Xplore. Restrictions apply.

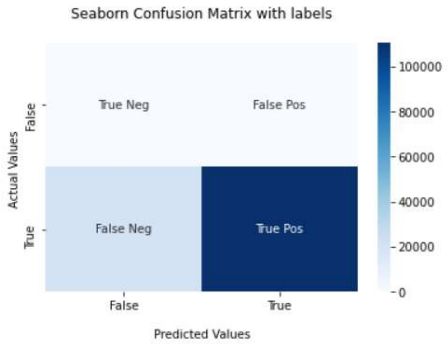



Fig. 7. Confusion matrix using count vectorizer (using Random forest classifier).

This shows that this algorithm computes a total of 39 and 110731 correct predictions.

## V. PERFORMANCE COMPARISON

The above machine learning classification algorithms were used for predicting the score, using the accuracy classification score function in python, by doing the sentiment analysis of online reviews using two NLP-based techniques, i.e., TF-IDF vectorizer and count vectorizer. Table VI shows the performance comparison between the different classification algorithms used in the proposed model. The comparison is based on the feature extraction method's 114969 most relevant features selected. Based on the accuracy, it is observed that Logistic regression gives the highest accuracy for both the NLP techniques while Dummy classification and Random forest classification provides a similar lower percentage. In terms of qualitative analysis, Logistic Regression is 10.72% better in terms of accuracy than the other machine learning algorithm when using the TF-IDF vectorizer and 11.02% better in terms of accuracy than the other machine learning algorithm when using the count vectorizer.

TABLE VI. PERFORMANCE COMPARISON

| **Classification Algorithms** | **Accuracy using TF-IDF Vectorizer [12]** | **Accuracy using Count Vectorizer [12]** |
|---|---|---|
| Logistic Regression [13] | 93.47% | 93.62% |
| Dummy Classifier [14] | 84.34% | 84.39% |
| Random Forest [15] | 84.50% | 84.26% |

## VI. CONCLUSION AND FUTURE WORK

This paper successfully analyses the reviews in the dataset by visualizing the pivot table. Finding the top positive and negative words using the NLP techniques, this paper successfully predicts the customer sentiments. In this paper, different machine learning algorithms and their applicability in analyzing customer reviews are compared predicting their sentiments. Majorly, two different NLP techniques are utilized to analyze the customer reviews, and three different classification algorithms are used to predict the customer's sentiments. This paper provides descriptions of the algorithms and their performance comparison based on the algorithm's accuracy. It is observed that Logistic regression gives the highest accuracy for both the NLP techniques compared to other algorithms. In future, more work can be done in this field, such as providing suggestions to the owners or manufacturers for improvement based on the customers' sentiments.

## ACKNOWLEDGMENT

The authors would like to express their gratitude to all the friends and others who in one way or another shared their support, either morally, financially and physically.
Authors express their appreciation for the useful suggestions provided by all the reviewers and editors to help improve this paper.

## REFERENCES

[1] B. Keith Norambuena, E. F. Lettura, and C. M. Villegas, "Sentiment analysis and opinion mining applied to scientific paper reviews," *Intell. Data Anal.*, vol. 23, no. 1, pp. 191–214, 2019, doi: 10.3233/IDA-173807.

[2] P. Pankaj, P. Pandey, M. Muskan, and N. Soni, "Sentiment Analysis on Customer Feedback Data: Amazon Product Reviews," *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019*, Faridabad, India, pp. 320–322, 2019, doi: 10.1109/COMITCon.2019.8862258.

[3] S. R. D'souza and K. Sonawane, "Sentiment analysis based on multiple reviews by using machine learning approaches," *Proc. 3rd Int. Conf. Comput. Methodol. Commun. ICCMC 2019*, Erode, India, no. Iccmc, pp. 188–193, 2019, doi: 10.1109/ICCMC.2019.8819813.

[4] A. Almjawel, S. Bayoumi, D. Alshehri, S. Alzahrani, and M. Alotaibi, "Sentiment Analysis and Visualization of Amazon Books' Reviews," *2nd Int. Conf. Comput. Appl. Inf. Secur. ICCAIS 2019*, Riyadh, Saudi Arabia, pp. 1–6, 2019, doi: 10.1109/CAIS.2019.8769589.

[5] K. L. Santhosh Kumar, J. Desai, and J. Majumdar, "Opinion mining and sentiment analysis on online customer review," *2016 IEEE Int. Conf. Comput. Intell. Comput. Res. ICCIC 2016*, Chennai, India, 2017, doi: 10.1109/ICCIC.2016.7919584.

[6] Y. Xiao, C. Qi, and H. Leng, "Sentiment analysis of Amazon product reviews based on NLP," *Proc. - 2021 4th Int. Conf. Adv. Electron. Mater. Comput. Softw. Eng. AEMCSE 2021*, Changsha, China, pp. 1218–1221, 2021, doi: 10.1109/AEMCSE51986.2021.00249.

[7] Z. Singla, S. Randhawa, and S. Jain, "Statistical and sentiment analysis of consumer product reviews," *8th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2017*, Delhi, India, pp. 1–6, 2017, doi: 10.1109/ICCCNT.2017.8203960.

[8] F. M. S. Haque, Tanjim Ul, Nudrat Nawal Saber, "Sentiment Analysis on Large Scale Amazon Product Reviews," *2018 IEEE Int. Conf. Innov. Res. Dev.*,Bangkok, Thailand, no. May, pp. 85–88, 2018.

[9] S. Choudhary and C. Chhabra, "Sentiment analysis of Amazon food review data," *Proc. - 2021 4th Int. Conf. Comput. Intell. Commun. Technol. CCICT 2021*, Sonepat, India, pp. 116–120, 2021, doi: 10.1109/CCICT53244.2021.00033.

[10] "Amazon food review dataset." https://snap.stanford.edu/data/web-FineFoods.html (accessed Jan. 20, 2022).

[11] A. Rajput, "Natural language processing, sentiment analysis, and clinical analytics," *Innov. Heal. Informatics A Smart Healthc. Prim.*, no. February, pp. 79–97, 2019, doi: 10.1016/B978-0-12-819043-2.00003-4.

[12] R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, "Sentiment analysis on product reviews using machine learning techniques," *Adv. Intell.*

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:45:04 UTC from IEEE Xplore. Restrictions apply.

*Syst. Comput.*, Sakheer, Bahrain,vol. 768, pp. 639–647, 2019, doi: 10.1007/978-981-13-0617-4_61.

[13] M. Qasem, R. Thulasiram, and P. Thulasiram, "Twitter sentiment classification using machine learning techniques for stock markets," *2015 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2015*, Kochi, India, pp. 834–840, 2015, doi: 10.1109/ICACCI.2015.7275714.

[14] A. Martino, A. Giuliani, and A. Rizzi, "(Hyper)graph embedding and classification via simplicial complexes," *Algorithms*, vol. 12, no. 11, 2019, doi: 10.3390/a12110223.

[15] S. Zahoor and R. Rohilla, "Twitter Sentiment Analysis Using Machine Learning Algorithms: A Case Study," *Proc. - 2020 Int. Conf. Adv. Comput. Commun. Mater. ICACCM 2020*, Dehradun, India, pp. 194–199, 2020, doi: 10.1109/ICACCM50413.2020.9213011.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:45:04 UTC from IEEE Xplore. Restrictions apply.

2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) | 978-1-6654-1120-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAECT54875.2022.9808038

# Grid Tied Photovoltaic Based Electric Vehicle Charging Infrastructure

Mamta Gupta
Department of Electrical Engineering
Delhi Technological University
Delhi, India
mamtagupta_2k20psy10@dtu.ac.in

Aakash Kumar Seth
Department of Electrical Engineering
Delhi Technological University
Delhi, India
akrseth@gmail.com

Mukhtiar Singh
Department of Electrical Engineering
Delhi Technological University
Delhi, India
smukhtiar_79@yahoo.co.in

*Abstract*— **This paper presents detailed simulation of grid tied photovoltaic (PV) based an electric vehicle (EV) charging infrastructure. The EV charger can work in both modes of power flow i.e., grid to vehicle (G2V) and vehicle to grid (V2G), which can be employed in diverse load power requirement scenarios. The overall system uses various converters at each stage, boost converter for harnessing solar PV power, a bi-directional buck-boost converter for EV charging, and a voltage source converter (VSC) with LCL filter for integrating overall system with the grid. The controlling of DC-DC boost converter for extracting peak power output is done using maximum power point tracking (MPPT) technique. The controller associated with bidirectional buck-boost converter is used for controlling power flow of battery pack. Moreover, the controller of grid side bidirectional AC-DC converter is utilized for controlling DC-link voltage. The system presented in this paper contributes towards maximizing energy output from PV system with perturbation and observation (P&O), uninterrupted EV charging, enhancing grid performance and using EV as a form of energy storage system when required. Extensive simulation for various power flow modes have been performed to verify the aforementioned claims. The overall design and simulation work is executed in MATLAB/Simulink software and results are confirmed for various input conditions.**

***Keywords—Photovoltaic, Vehicle to Grid, Grid to Vehicle, EV charging, Grid Power Support***

## I. INTRODUCTION

In the last couple of decades, the growing concern in the energy sector is mainly due to continuous depletion of fossil fuels and emission of carbon di oxide which leads to green-house effect. For sustainable management of conventional resources and for minimizing hazardous environmental condition, in the field of alternate energy sources further research and development is required. Solar powered energy is one of the most widely used non-conventional energy, as it is abundant, inexhaustible, free and clean [1].

One of the major sectors using fossil fuel and natural gas is the transportation sector, hence battery electric vehicle was developed as an alternative to Internal Combustion (IC) engines, for reducing emission of CO2 and to limit the use of fossil fuel. Integration of EV with grid can also provide other ancillary services like regulation of voltage and frequency, peak shaving, can act as a virtual power plant and regulated spinning reserve with the smart grid [2]-[4].

Moreover, integration of EV on a huge scale and its unregulated connection into the power distribution system might poses various problem such as rising peak demand, harmonics, voltage imbalance and current distortion etc. [5]. Another issue is that no emission EVs are majorly charged using electricity grid, where the major concentration of fuel mix comes from conventional sources, hence PV powered EV charging is a sustainable option for long term [6]. Interconnection of solar PV system with EV charging in a coordinated manner mitigates impact of solar PV generation on power grid. However, PV integrated system with EV leads to system complexity and to avoid issues related to solar power fluctuation, there is a need of integration of PV with grid and other multifunctional operating sources for reliable operation of EV charging [6]-[7].

There are three charging levels for EV [8]. Level 1 charging consist of an on-board charger with 120 – 240 V AC and 15 A of maximum current. The Level 2 charging consist of an on-board charger of voltage level 240 V with maximum charging current of 40 A. Level 3 consist of an off-board charger with 200-800 V and 250 A of maximum current.

In the literature [9]-[11], various topologies of grid integrated PV systems have been discussed. An adjustable DC-link with two stage grid tied solar PV system is presented [12], and an elimination of one stage leads to single stage topology having their own benefits and drawbacks. Topologies are also dependent on certain other factors such as isolation, power level requirement and controlling strategy. In [1], the elimination of boost converter is presented, here DC bus directly connects solar PV array with VSC, allowing VSC to harness the peak power out of solar PV array.

Several control strategies have already been presented in literature for different power converters mainly includes Proportional-Integral (PI) control, Proportional-Resonant (PR) control, Repetitive control (RC), Model Predictive Control (MPC) etc. [13]-[16]. In this work, boost converter harnesses peak power out of PV array, which is further interconnected with DC link of VSC for AC-DC conversion. The EV battery charger is interfaced at DC link with the help of DC-DC bi-directional buck-boost converter. The control algorithm prioritizes energy sources utilization based on solar power generation and EV charging requirement. Firstly, it utilizes solar power for EV charging and in case of less availability or over power generation remaining power is taken from or fed to grid respectively. EV can also be used as a form of storage system and during peak load time it can supply energy back to grid. During unavailability of solar power, EV and grid can act as an isolated system having both V2G and G2V mode.

The overall system configuration as shown in Fig. 1. is designed in MATLAB/Simulink software, where total seven power flow modes from mode 1 to mode 7, which has been discussed in Table II., is simulated to validate controller performance upon various input conditions and load power requirements.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:41:58 UTC from IEEE Xplore. Restrictions apply.

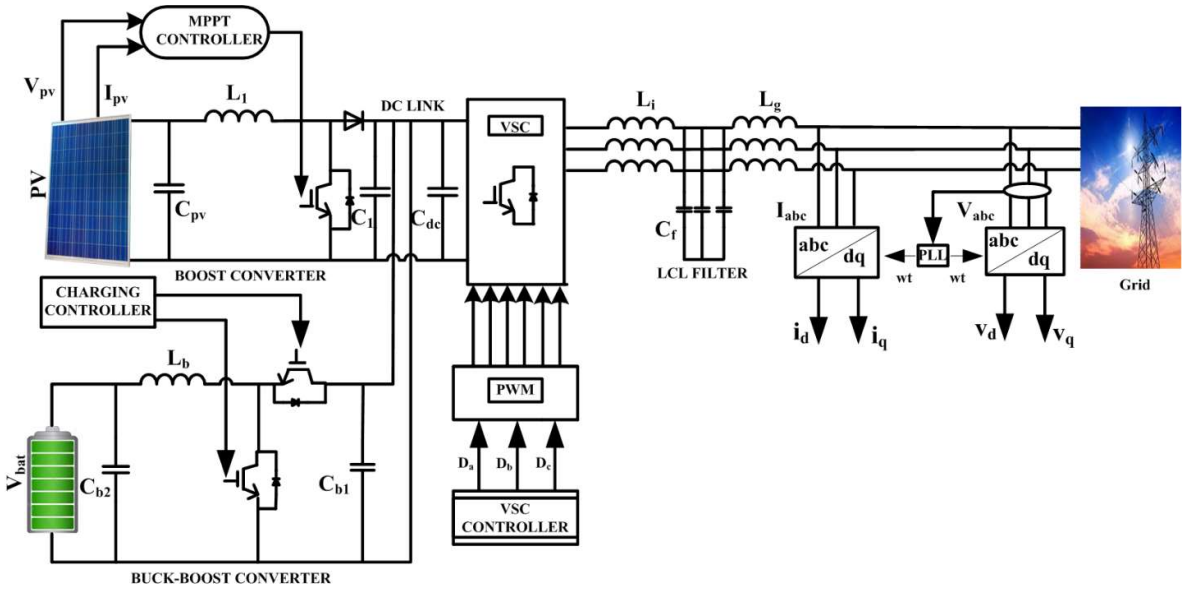



Fig. 1. Overall system configuration

## II. SYSTEM DESCRIPTION

The complete system architecture is shown in Fig. 1. The system consists of PV source, grid and EV charging station.

### A. Solar PV cell model

Fig. 2. shows equivalent electric circuit model of PV cell which is obtained after analyzing its terminal characteristic, which is further utilized to design photovoltaic systems. The terminal equation that governs the output voltage-current performance characteristic of PV cell is as follows:

$$I_D = I_O\left(e^{\frac{V+IR_{se}}{nVt}} - 1\right) \quad (1)$$

$$I = I_{sc} - I_O\left(e^{\frac{V+IR_{se}}{nVt}} - 1\right) - \frac{V+IR_{se}}{R_{sh}} \quad (2)$$

Where $I_{sc}$ represents the photo electric current proportional to the total amount of perpendicular solar energy incident on solar cell surface, $I_o$ represents the reverse saturation current which is dependent on doping and temperature of p-n junction diode; $R_{sh}$ and $R_{se}$ implies shunt and series resistance present due to nonlinearity in PV characteristic; $n$ denotes the diode ideality factor and $V_T$ represents the thermal voltage.

The above equation is an acausal equation but in practice diode is not ideal hence it has the junction capacitance which takes care of the causality problems.

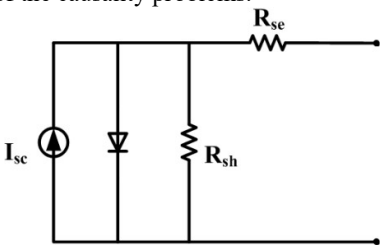



Fig. 2. Electric equivalent PV cell model.

Under different irradiation levels, characteristic of *I-V*, *P-V* is drawn, which is depicted in Fig. 3. The nonlinearity in PV cell characteristic indicates that, the unique *I-V* characteristic results in output current which is constant over a wide voltage range until it extends to a point from where it begins to drop exponentially, giving rise to the concept of maximum power point which has to be tracked for exporting peak power out of solar array.

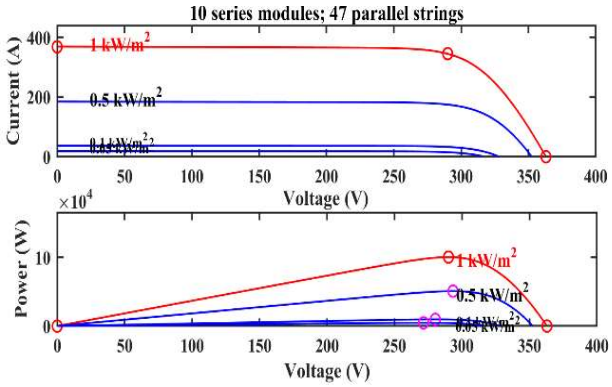



Fig. 3. PV array characteristic.

### B. Boost Converter MPPT control

The main function of DC-DC boost converter is to rise the source side solar PV array voltage to load terminal by adjusting the duty cycle. This duty cycle, for maximum power is maintained by MPPT techniques. Among several MPPT techniques, Perturbation and observation (P&O) method is most widely implemented because of the limited number of parameters required, simplicity in its structure and ease in calculation. The parameter required to track MPP is current and voltage of PV panel as input values, which implies the association between output power and terminal voltage. This method iteratively perturbs, observes and generates output power to track maximum power point. If at certain PV module operating $\frac{\Delta P}{\Delta V} > 0$ then the perturbation of PV module output voltage should be increased toward MPP, if PV module operating point $\frac{\Delta P}{\Delta V} < 0$ then the perturbation of PV module output voltage should be decreased toward MPP. If the status of certain point is $\frac{\Delta P}{\Delta V} = 0$ then that point is the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:41:58 UTC from IEEE Xplore. Restrictions apply.

point of maximum power. The MPPT flow chart is shown in Fig. 4.

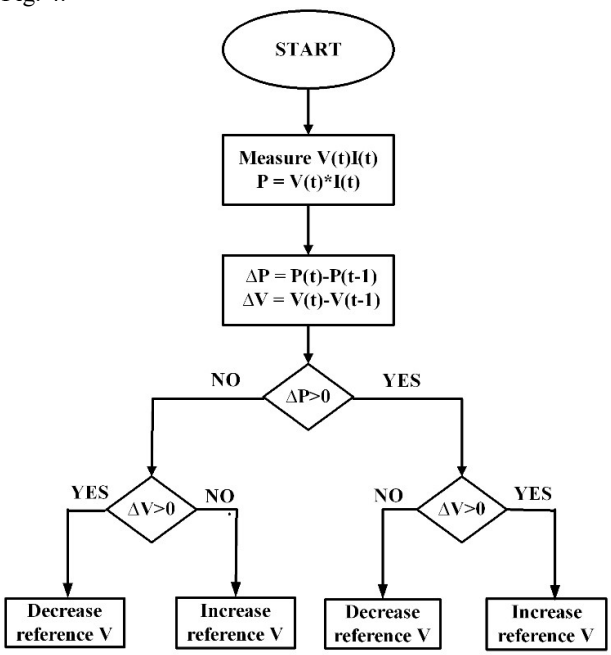



Fig. 4. P & 0 flow chart.

For boost converter control through MPPT, PV current and voltage is first sensed and passed through MPPT to generate $V_{ref}$ which is the reference voltage for tracking MPP, error signal is generated by comparing $V_{ref}$ with PV voltage, which is then passed to the PI controller , which will further generate signal having the required duty ratio, The output PWM signal is a result of comparing this generated signal with the carrier signal for the purpose of getting the required switching pulses for the boost converter as shown in Fig. 5.

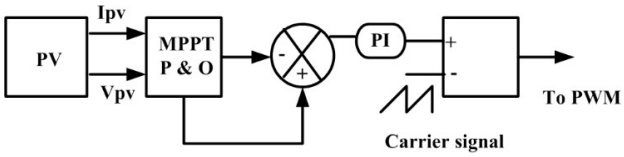



Fig. 5. MPPT control block diagram.

### C. Voltage source converter configuration

The converter implemented for PV to grid connection is bi-directional three phase AC to DC converter with sine pulse width modulation technique. The bidirectional power flow is important for this work as it allows power flow in both forward and reverse direction depending upon load requirement. For such an application appropriate control strategy of power switches of converter is required.

### *Design of LCL filter*

The PV inverter section, on the DC input side consist of PV panel and DC link capacitor and on the output AC side consist of an LCL filter interfaced with the grid. PV panel generated voltage must always maintain a DC-link voltage greater than the peak AC voltage at all time. Switching harmonics created by PWM must be reduced to improve the quality of power provided to local loads and that injected into the grid, for improving this quality, filters are required. There are many filter topologies for example *L* , *LC* and *LCL* filter etc., among various filter topologies *LCL* filter is the most suitable for providing better attenuation and coupling capability.

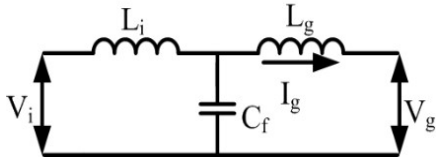



Fig. 6. Single phase equivalent model of LCL filter.

LCL filter has good ripple attenuation capability but can cause resonance related disturbance hence selection of filter components needs to be precise. For the design of filter, selecting filter components depends on several factors. Device constraint, cost and thermal consideration etc. influence switching frequency selection. The switching frequency of 10 kHz is chosen in this case.

The fundamental and switching frequency both must be very well separated from resonant frequency for its selection. The transfer function of LCL filter neglecting damping, all higher order harmonics condition, and by setting condition of $V_g = 0$ as shown in Fig. 6. is as follows:

$$\frac{Ig(s)}{Vi(s)} = \frac{1}{S^3 Li\ Lg\ C_f + S(Li + Lg)} \tag{3}$$

The capacitance value is determined based on reactive power requirement of capacitor. Due to this reactive power, overall output power factor is reduced causing large current to flow from IGBTs and $L_i$. The formula of capacitor $C_f$ as shown in eq. (4) can be derived, with the limitation of reactive power requirement, keeping its value below 5% of inverter's rated output power.

$$C_f = \frac{0.05 * S}{v^2 * 2 * pi * f} \tag{4}$$

For the minimization of ripple current flowing through switching IGBTs inverter side inductor is required. The total change in current during the overall cycle is expressed as:

$$\Delta i = \frac{V_{dc}}{F_{sw} * 2L_i}(1 - M_r \sin \omega_0 t) M_r sin w_0 t \tag{5}$$

Where $F_{sw}$ is the switching frequency for the IGBT's, $V_{dc}$ represents input DC-link voltage , $\omega_0$ refers to the fundamental angular frequency and modulation index is represented by $M_r$.

Considering the design constraints, 10 % ripple variation of rated current is allowed, for designing of inductor maximum value of ripple content is calculated, after further simplification and by putting appropriate value of $\omega_0$, $L_i$ follows the equation:

$$L_i = \frac{V_{dc}}{8 * F_{sw} * \Delta i_{max}} \tag{6}$$

The use of grid side inductor is to mitigate and keep current harmonics and total harmonic distortion (THD) under specified IEEE standards. The following equation gives the minimum value of $L_g$:

$$L_g = \frac{1}{L_i C \omega_n^2 - 1}\left(L_g - \frac{\left|V_g(j\omega_n)\right|}{\omega_n \mu_n I_g}\right) \tag{7}$$

For this computation we require to obtain frequency spectrum of voltage $\left|V_g(j\omega_n)\right|$. The angular frequency ($\omega_n$)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:41:58 UTC from IEEE Xplore. Restrictions apply.

of dominant harmonic must be determined from that equation. The estimated harmonic proportion of the filter output current is $\mu_n$.

### D. Battery charger configuration

DC bus links EV batteries to off board charger. DC-DC converter is the basic building block and act as an interface to above battery charger configuration. The converter configuration as shown in Fig. 1. consist of two IGBT switches which always operate in complimentary fashion.

It permits power to flow in both directions, resulting in two modes of operation: buck and boost. When PV is incapable of operating it provides V2G operation hence EV acts as an auxiliary power supply leading to different modes of operation. The equation governing the above two operations are as follows:

Boost mode: $V_H = \frac{V_L}{1-D}$ (8)

Buck mode: $V_H = \frac{V_L}{D}$ (9)

$V_H$ and $V_L$ denotes voltages on high voltage and low voltage ends.

## III. CONTROL DEVELOPMENT

### A. Voltage source converter control

The PI based controller design of VSC is depicted in the Fig. 7. The controller ensures DC-link voltage regulation. It is required for the reference value to be constant in its steady state for the necessary integral action of a PI based current controller to provide zero steady state error control, this can be achieved by applying Clarke's and park's transformation which is used to transform AC based current quantities into DC quantities and hence simple PI controller gives good result. The phase looked loop (PLL) is used to calculate $\omega t$ which maintains the synchronization between inverter current and grid voltage.

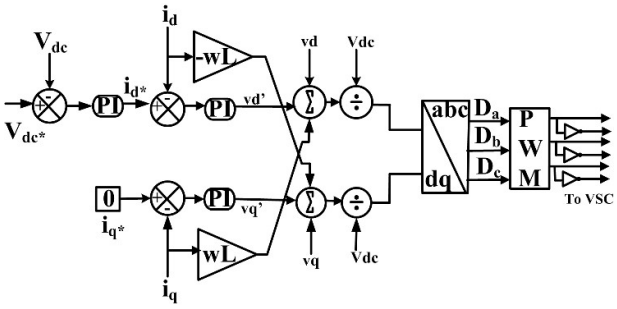



Fig. 7. VSC decoupled current controller.

Since, PV based grid system is in combination with EV charging system, multimode controller combinations are required. Two control loops is used in this PWM based technique i.e. an outer slower voltage loop which is used to regulate DC link voltage and an inner fast current loop for controlling *dq* axis current ($i_d$ and $i_q$). Since, the *d* and *q* axis values are interconnected, controlling the active and reactive flow of power into the grid requires the use of two PI controllers. The DC link voltage is regulated by the PI controller, by minimizing the difference between the reference and actual voltage values, which further act as a reference for the *d* axis current ($i_d^*$). The line current vector must be aligned with the line voltage vector to achieve the condition of Unity Power Factor (UPF), hence the *q* axis current reference is set to zero ($i_q^* = 0$). The expression for *dq* axis current is as follows:

$$\frac{di_d}{dt} = v_d - k_p v_d' - k_i i_d + wLi_q/L \quad (10)$$

$$\frac{di_q}{dt} = v_q - k_p v_q' - k_i i_q - wLi_d/L \quad (11)$$

Where *L* is the line inductance and *w* is the grid's angular frequency. In addition, ignoring inverter losses, the equation for active and reactive power is as follows:

$$P = \frac{3}{2} Re\{v^{dq}(i^{dq})^*\} = \frac{3}{2}(v_d i_d + v_q i_q) \quad (12)$$

$$Q = \frac{3}{2} Im\{v^{dq}(i^{dq})^*\} = \frac{3}{2}(v_q i_d - v_d i_q) \quad (13)$$

Since, $i_q^* = 0$ so voltage vector is aligned with *d* axis, we get:

$$P = \frac{3}{2}(v_d \times i_d) \quad (14)$$

$$Q = 0 \quad (15)$$

### B. Battery charger control

Fig. 8. depicts a PI controller-based constant current control that is used to control the charging and discharging of an electric vehicle battery. Depending on the direction of power flow, buck and boost mode is operated. Buck mode is used to charge the battery if the SOC is less than the threshold value, and boost mode is used to operate V2G if the SOC is larger than the threshold and based on required load power demand on grid.

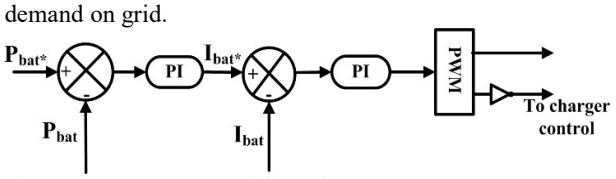



Fig. 8. Constant current control for EV battery.

The operation of a DC-DC bi-directional converter is determined by the amount of PV power generated, the state of charge (SOC) of the electric vehicle battery, and the load demand of the power grid. For proper operation of different modes, $I_{bat}$ and $P_{bat}$ are the two quantities which needs to be controlled by the controller, therefore two PI controllers are required. For outer loop battery actual power is compared with the reference power to obtain the signal which will be further compared with battery current to generate switching pulses i.e. D. Based on the generated power reference and measured power, decision is made by the controller leading to operation of battery charging (buck) and battery discharging/V2G (boost) mode.

$$I_{bat}^* = (k_{p1} + \frac{k_{i1}}{s}) * (P_{bat}^* - P_{bat}) \quad (16)$$

$$D = (k_{p2} + \frac{k_{i2}}{s}) * (I_{bat}^* - I_{bat}) \quad (17)$$

## IV. RESULTS AND DISCUSSION

An integrated solar PV based EV charging system consisting of various controllers, power electronics converters as shown in Fig. 1. is executed in *MATLAB/Simulink* environment. Table I. enumerates the system parameters of the above model.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:41:58 UTC from IEEE Xplore. Restrictions apply.

TABLE I. System parameters

| PARAMETER | SYMBOL | VALUES |
|---|---|---|
| PV Power | $P_{pv}$ | 100 kW |
| Open circuit voltage of PV array | $V_{oc}$ | 363 V |
| DC link voltage | $V_{dc}$ | 600 V |
| Boost inductor | $L_1$ | 1.45 mH |
| Boost capacitor | $C_1$ | 3227 µF |
| Battery voltage | $V_{bat}$ | 350 V |
| Battery SOC | $soc$ | 50 % |
| Buck boost inductor | $L_b$ | 4 mh |
| Buck boost capacitor | $C_{b1}$ | 100 µF |
| Grid voltage | $V_{grid}$ | 400 V |
| Grid frequency | $f$ | 50 Hz |
| Switching frequency | $f_{sw}$ | 10 kHz |
| Filter inductor | $L_i$, $L_g$ | 500 mH |
| Filter capacitor | $C_f$ | 100 µF |

To verify the performance claim of controllers, different modes of power flow as shown in Table II. are obtained depending on PV generated power, EV charging/discharging state, and grid power requirement. The PV system gives maximum power output of 100 kW and power varies according to irradiance condition 1000 $\frac{W}{m^2}$ – 500$\frac{W}{m^2}$ –1000 $\frac{W}{m^2}$ – 0 $\frac{W}{m^2}$ pertaining to different modes as shown in Fig. 9 (a). The DC-link voltage is maintained at its reference value of 600 V as inferred form Fig. 10.

TABLE II. SIMULATION VALUES

| # MODE | IRRADIATION $\frac{W}{m^2}$ | $P_{PV}$ kW | $P_{bat}$ kW | $P_{grid}$ kW |
|---|---|---|---|---|
| 1 | 1000 | 100 | 50 | -50 |
| 2 | 500 | 48 | 80 | 32 |
| 3 | 500 | 48 | 48 | 0 |
| 4 | 1000 | 100 | -50 | -150 |
| 5 | 1000 | 100 | 0 | -100 |
| 6 | 0 | 0 | 50 | 50 |
| 7 | 0 | 0 | -50 | -50 |

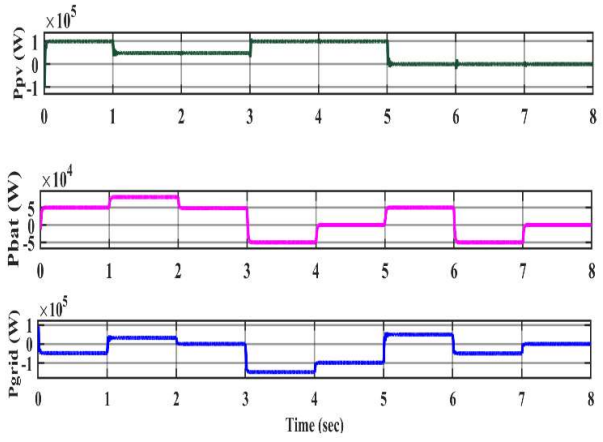



Fig. 9 (a) Solar PV generated power (b) Battery power waveform during charging/discharge mode (c) Grid active power ($P_{grid}$).

The power for solar PV, battery and grid active power for different modes of operation are enumerated in Table II. and shown in Fig. 9. During mode 1 generated PV power is used for charging of EV battery and power which is remaining is absorbed by the grid. Mode 2 shows EV charging through PV and grid both i.e. G2V mode is also shown here which will be required in case of increased EV charging demand and during the case in which PV is not sufficient enough to meet EV load demand. Mode 3 shows PV based EV charger which is isolated from grid. During the case of increase load demand on grid, mode 4 can be used where EV is acting as a storage system i.e. V2G operation and where both PV and EV are supplying power to meet the grid load demand. Mode 5 shows PV grid system isolated from EV. During mode 6 and mode 7, PV is isolated from EV and grid system, here G2V and V2G operations are shown respectively.

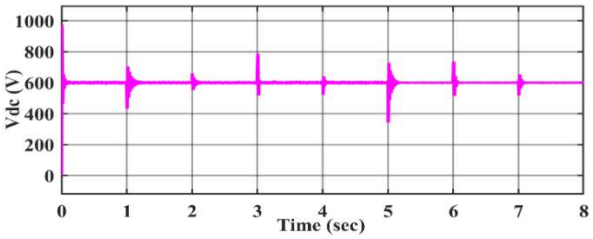



Fig. 10. DC link voltage.

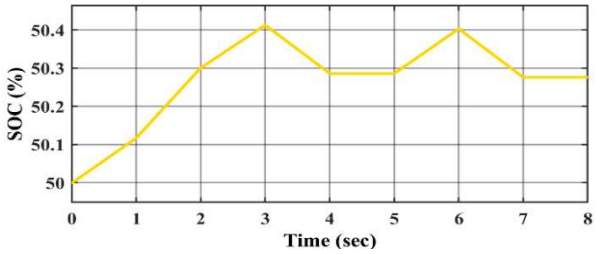



Fig. 11. SOC of battery.

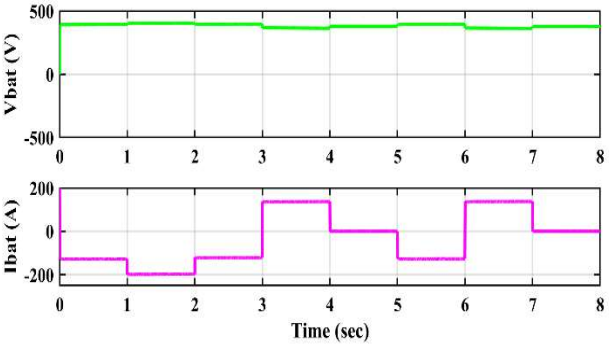



Fig. 12. Simulation result of battery voltage $V_{bat}$ and battery current $I_{bat}$.

SOC, battery voltage ($V_{bat}$) and battery current ($I_{bat}$) are shown in Fig. 11 and 12 respectively. During mode 1 2 3 and 6 battery current is negative and battery power is positive which signifies battery charging state and in mode 4 and 7 battery current is positive and battery power is negative which signifies battery discharging state. Mode 5 is showing no action on EV. Battery voltage value is positive during all the stages. Battery SOC is increasing in mode 1 2 3 6 and decreasing in mode 4 and 7 signifying EV charging and discharging respectively. SOC is increasing faster in mode 2 which means more power is supplied to the battery in this mode compared to other modes.

Fig. 13 (a) and (b) shows instantaneous grid voltage and current. Mode 2 and 6 follows Fig. 13 (a) shows same phase between voltage and current and Fig. 9 (c). also signifies positive power for this modes i.e. power is supplied by the grid in this mode. Mode 1 4 5 7 follows Fig. 13 (b) and shows voltage and current in opposite phase and Fig. 9 (c). also

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:41:58 UTC from IEEE Xplore. Restrictions apply.

showing negative power for all this modes i.e. power is absorbed by the grid in this case.

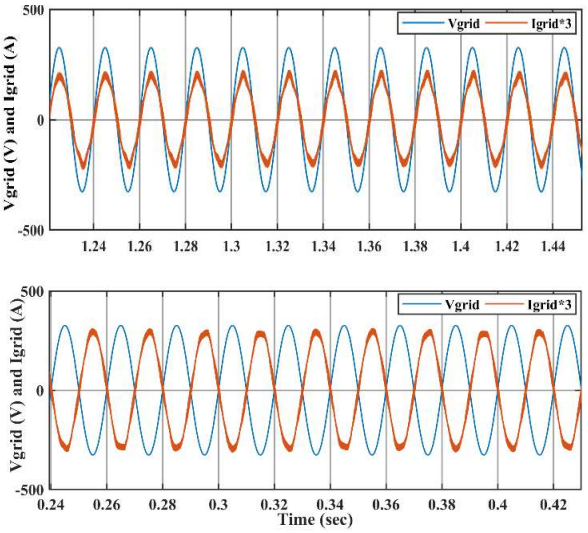



Fig. 13. Simulation results of instantaneous grid voltage and current during (a) mode 2 and 6 is in same phase and (b) and during mode 1, 4, 5 and 7 in out of phase.

## V. Conclusions

In this work detailed modelling and simulation of grid tied PV based EV charging system has been carried out. The maximum power out of PV array is extracted with the help of DC-DC boost converter. The system is modelled to charge or discharge EV based on PV power availability, EV charge state and grid requirement using bi-directional converters. To verify the performance of controllers simulation for various modes have been carried out. This model is cost-effective and simple in design, and it supports PV power efficiency, continuous EV charging, and enhancement in grid power flow under a variety of load demand scenarios. Moreover, this model shows that the grid power quality will not be impacted in any of the power flow mode.

## VI. References


[1] A. Verma, B. Singh, A. Chandra and K. Al-Haddad, "An Implementation of Solar PV Array Based Multifunctional EV Charger," *2018 IEEE Transportation Electrification Conference and Expo (ITEC)*, 2018, pp. 531-536, doi: 10.1109/ITEC.2018.8450191.

[2] L. Cheng, Y. Chang and R. Huang, "Mitigating Voltage Problem in Distribution System With Distributed Solar Generation Using Electric Vehicles," IEEE Trans. Sustainable Energy, vol. 6, no. 4, pp. 1475-1484, Oct. 2015, doi: 10.1109/TSTE.2015.2444390.

[3] M. Brenna, F. Foiadelli and M. Longo, "The Exploitation of Vehicle-to Grid Function for Power Quality Improvement in a Smart Grid," IEEE Trans. Intelligent Transportation Systems, vol. 15, no. 5, pp. 2169-2177, Oct. 2014, doi: 10.1109/TITS.2014.2312206.

[4] N. Saxena, I. Hussain, B. Singh and A. L. Vyas, "Implementation of a Grid-Integrated PV-Battery System for Residential and Electrical Vehicle Applications," in *IEEE Transactions on Industrial Electronics*, vol. 65, no. 8, pp. 6592-6601, Aug. 2018, doi: 10.1109/TIE.2017.2739712.

[5] L. Wang, Z. Qin, T. Slangen, P. Bauer and T. van Wijk, "Grid Impact of Electric Vehicle Fast Charging Stations: Trends, Standards, Issues and Mitigation Measures - An Overview," in *IEEE Open Journal of Power Electronics*, vol. 2, pp. 56-74, 2021, doi: 10.1109/OJPEL.2021.3054601.

[6] A. Verma and B. Singh, "Control and Implementation of Renewable Energy Based Smart Charging Station Beneficial for EVs, Home and Grid," *2019 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2019, pp. 5443-5449, doi: 10.1109/ECCE.2019.8913253.

[7] B. Singh, A. Verma, A. Chandra, and K. Al-Haddad, "Implementation of Solar PV-Battery and Diesel Generator Based Electric Vehicle Charging Station," IEEE Transactions on Industry Applications, vol. 56, no. 4, pp. 4007-4016, 2020, doi: 10.1109/TIA.2020.2989680.

[8] S. Chalia, A. K. Seth and M. Singh, "Electric Vehicle Charging Standards in India and Safety Consideration," *2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, 2021, pp. 1-6, doi: 10.1109/UPCON52273.2021.9667649.

[9] A. Narula and V. Verma, "Dual Port Impedance Converter for PV – Battery fed Remote Telecom Towers," 2018 IEEE 4th Southern Power Electronics Conference (SPEC), 2018, pp. 1-7, doi: 10.1109/SPEC.2018.8636089.

[10] H. Saxena, A. Singh and J. N. Rai, "Application of Time Delay Recurrent Neural Network for Shunt Active Power Filter in 3-Phase Grid-tied PV System," 2019 National Power Electronics Conference (NPEC), 2019, pp. 1-6, doi: 10.1109/NPEC47332.2019.9034769.

[11] M. O. Badawy and Y. Sozer, "Power Flow Management of a Grid Tied PV-Battery System for Electric Vehicles Charging," in *IEEE Transactions on Industry Applications*, vol. 53, no. 2, pp. 1347-1357, March-April 2017, doi: 10.1109/TIA.2016.2633526.

[12] C. Jain and B. Singh, "An Adjustable DC Link Voltage-Based Control of Multifunctional Grid Interfaced Solar PV System," in *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 5, no. 2, pp. 651-660, June 2017, doi: 10.1109/JESTPE.2016.2627533.

[13] A. K. Seth and M. Singh, "Resonant controller of single-stage off-board EV charger in G2V and V2G modes," in IET Power Electronics, vol. 13, no. 5, pp. 1086-1092, 8-4-2020.

[14] A. K. Seth and M. Singh, "Second-Order Ripple Minimization in Single-Phase Single-Stage Onboard PEV Charger," in *IEEE Transactions on Transportation Electrification*, vol. 7, no. 3, pp. 1186-1195, Sept. 2021, doi: 10.1109/TTE.2021.3049559.

[15] T. He, J. Zhu, D. D. Lu and L. Zheng, "Modified Model Predictive Control for Bidirectional Four-Quadrant EV Chargers With Extended Set of Voltage Vectors," in IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 7, no. 1, pp. 274-281, March 2019, doi: 10.1109/JESTPE.2018.2870481.

[16] I. Jayathilaka *et al*., "DQ Transform Based Current Controller for Single-Phase Grid Connected Inverter," *2018 2nd International Conference On Electrical Engineering (EECon)*, 2018, pp. 32-37, doi: 10.1109/EECon.2018.8541004.


Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:41:58 UTC from IEEE Xplore. Restrictions apply.

arXiv:2206.14050v1 [math.OC] 28 Jun 2022

# Identification of storm eye from Satellite image data using fuzzy logic with machine learning

Monu Yadav[1] and Laxmi Narayan Das[1*]

[1*]Department of Applied Mathematics, Delhi Technological University, New Delhi, India.

*Corresponding author(s). E-mail(s): lndas@dce.ac.in;
Contributing authors: yadavm012@gmail.com;

**Abstract**

The method discussed in this paper uses fuzzy logic and image processing to determine the eye of tropical cyclones from images of IR channels from geostationary satellites. This technique has been used to analyze images of four tropical cyclones: Amphan, Nivar, Tauktae, and Yaas. Fuzzy logic provides the ability to model the dynamic behavior of the storm and determines the location of the best eye in an area of interest. It provides an effective method for describing the uncertain behavior of complicated systems. After that, image processing is applied to enable accurate eye positioning based on the search results. A comparison has been made between this technique's experimental results and those of the Indian Meteorological Department and Cooperative Institute of Meteorological Satellite Studies. As a result, the storm's eye location can be found to improve performance significantly. Using the present technique, it is possible to determine the eye entirely automatically, thus replacing the manual method that has been employed in the past.

**Keywords:** Tropical Cyclone, Eye, IR image, Fuzzy logic, Image Processing

## 1 Introduction

Along with gales, rainstorm, and storm surges, tropical cyclones (TCs) are considered extreme weather events that have the potential to cause massive

damage for coastal areas. Meteorologists and warning centers have been studying the atmosphere, the atmospheric boundary layer, and the air-sea interface since the 19th century and have made significant advances in observational technology, intensification physics, the ocean responses, and the interaction of these areas. In Particular, predictions of TCs and their genesis, intensities, and risks remain quite difficult because of climate and geographical variation as well as the climatological characteristics of oceans. As a result, a single forecasting technique will not be able to perform adequately and accurately. Additionally, there are multiple key characteristics of a storm, including wind speed, wind direction [2–10], pressure system [2, 3, 6, 11, 12], temperature [3, 4, 10, 11, 13–15, 26], moisture precipitation [3, 11], humidity[11] and cloud shape [2, 13, 14, 16–20, 44]. Prediction accuracy is affected by all of these parameters, and more complex forecasting is made more difficult.

Tracking the characteristics of prior information has led to the invention of several types of storm forecasting. According to [27], tropical cyclone (TC) variability is defined by the wind circulation patterns within tropical clusters and based on the storm trajectory tracking. [28] presented a time-series approximation technique for the prediction of a TC in the Western North Pacific was used for the development of the modified A-LTK (Approximation with use of Local features at Thinnedout Keypoints) technique for a multi-dimensional time-series similarity calculation method.In this technique, the time-series data is reduced to a smaller number of time points and the nearby time points are used to create a feature vector at the thinned-out time point. Using wind data provided by Joint Typhoon Warning Center and National Hurricane Center, [29] analyzed how peak intensity and lifetime maximum intensity could affect the recurrence of TCs. In several studies, eye location has been obtained from the ocean surface wind vectors extracted from Sea-Wind instrument from the satellite [4, 5, 8, 16, 17, 30]. Based on wind data, Artificial Intelligence Technique's Support Vector Machines were used to identify cyclones fully automatically [5], and Fuzzy C-Means clustering was used to identify cyclones objectively [30].

As an alternative source of information,cloud images can be used to determine the patterns of storm [2, 19, 21–24, 26] or the location of the storm eye[13, 14, 16–18, 20, 25]. For the automatic segmentation of TC patterns, an elastic pattern matching method that is based on a neural network-based model on applied to the Dvorak analysis [2, 15] of predefined TC patterns, while a similar idea was used to recognize TC patterns with Neural Oscillatory Elastic Graph Matching Model (NOEGM) [26]. In addition, the Dvorak algorithm for matching patterns in TC was developed by [21–23]. An algorithm was developed to determine the axisymmetry of the shape of TCs and their flexibility. Additionally, the algorithm predicted wind speed, storm trajectories and the expansion of TCs by the gradient of the shape and the Deviation-Angle Variance (DAV) of satellite infrared. In [18], we proposed an automated approach to detect the center of a TC by fitting logarithmic spiral patterns to the image enhancement and the center of spiral has been assigned to the TC

center. Although, this method may not be accurate in areas with ambiguous boundaries. As a remedy to this problem, mountain-climbing search was proposed by [13]. Since the method requires a long period of searching time and sometimes is unreliable due to cloud shape uncertainty, it is still complicated in practice. We need to take into account a suitable termination condition when designing the search algorithm to solve this problem. In order to deal with this issue [14] introduce the Spiral Curve Model and Ant Colony Optimization algorithm which can used to determine the termination condition. However, it remains challenging to predict the exact storm location, particularly when the storm eye is unsure or multiple.

The use of these techniques has been satisfying the needs of the Meteorological community for quiet a few decades, but now that there is a much large volume of satellite observational, some advance mathematical agents as well as image processing, it seems that there is still room for improvement. As a mathematical agent, fuzzy is used for forecasting TC. The concept of fuzzy-based decision making allows for expert support at a linguistic level while evaluating data to learn statistical characteristics or to improve the rule. To improve detection results, it id necessary to have domain knowledge of cloud images of the storm. TC forecasting was proposed as a case-based fuzzy multi-criteria decision support system in [31]. In case-based reasoning, we collected historical observation data for a ten-year period to develop fuzzy multi-criteria decision support. The FIS techique provide an advantange of reducing large volumes of data into a smaller number of the best cases, which then allows for further filtering and evaluation of the results for use by a forecaster in real time. Nevertheless, the accuracy of the algorithm varied since location of the storm center in a geographical region still played an important role in the forecast. Therefore, in order to handle the uncertainties of the storms, we have introduced a fuzzy-based storm eye identification system designed specifically for improving tracking of storm centers. An image processing technique for predicting TC characteristics can be used in a variety of ways. TC position and intensity are estimated reasonably accurately by passive and active microwave sensors such as the TRMM Microwave Imager and wind scatterometer [35, 36]. As a result of the high temporal sampling and wide coverage provided by the sensors, geostationary satellite IR images provide the best information for locating and estimating the intensity of TCs in real-time [37]. TC analysts can now make use of dynamic or quickly evolving features in TC because geostationary satellites usually provide half hourly observations. As with intensity estimation, positioning procedures are based on the results of TC imagery conducted years ago [38, 39]. Some of the most important techniques for tracking TCs includes wind field analysis and pattern matching [40, 41], the TREC (tracking radar echoes by correlation) algorithm [42] or automatic cloud feature tracking [43]. Several researchers, using radar data and IR images, have attempted to determine the eye of TC in recent years. By determine the center of the biggest circle contour within the TC in a infrared image, [34] developed a method to determine the approximate eye of any TC. [33] demonstrated that TC spiral

bands can be represented analytically by a combination of hyperbolic and logarithmic spirals. According to [32, 44], they determined the eye of TC using motion field analysis. Compared with earlier eye determined studies, this one presents a new technique that is more unique. An automated method is used in this paper to determine the cyclone's. Fuzzy logic and Image processing are included in this. As part of fuzzy logic, we construct 16 If-Then rules of Multiple Input with Single Output system, define Membership function for Moisture Density, Wind speed, Pressure Drop at the Center, Estimated central pressure and defuzzification with Center of Area method to obtain crisp output. As part of Image processing,the techniques used include removal of noise from images by morphological operations as well as determining of center of gravity/mass of the TC image. Compared to other method, the proposed method shows improved performance. A description of data analysis is given in Section 2. The proposed method is presented in Section 3. Experimental results are presented in Section 4 and compared with other method. Finally summary and conclusion are given in Section 5.

## 2 Analyzing the Data

Typically, the eye of the storm has the greatest important factor in TC tracking. Six-hourly IR images were analysed in this study. Satellite images generally will show TCs as clusters of well-organized clouds, although clouds clusters will lose structure during the formation process , causing the system to be difficulty to identify as a TC. The storm eye is difficult to locate due to incomplete and insufficient data.

There are IR images recorded from tropical storms and typhoon occurring in the North Indian Ocean and Bay of Bengal. For weather forecasting, IR images are widely used. It is generally accepted that several agencies collect the best weather tracks of a storm, such as India Meteorological Department (IMD), Joint Typoon Warning Center (JTWC), and Cooperative Institute of Meteorological Satellite Studies (CIMSS). However, the data for the best track recordings from each organization differ quite a bit from one another. Performance estimation will be affected by these track errors.

## 3 A Possible Location of the Storm Eye

To identify a storm, it is usually necessary to extract its characteristics. From satellite data it is possible to see the storm's eye, which is one of the most widely observed storm features. In recent years, researcher have found many ways to detect the storm's eye, using artificial intelligence technique such as support vector machine, feature vector, genetic algorithms, fuzzy C-means, etc. The purpose of this study is to identify the storm eyes based on a novel FIS definition to improve the storm eye specifications for a variety of storm characteristics and outlined in figure 1. A pair of main processes is included in the propose system which are identification of storm eye locations using FIS and refinement of storm eye location using Image processing.

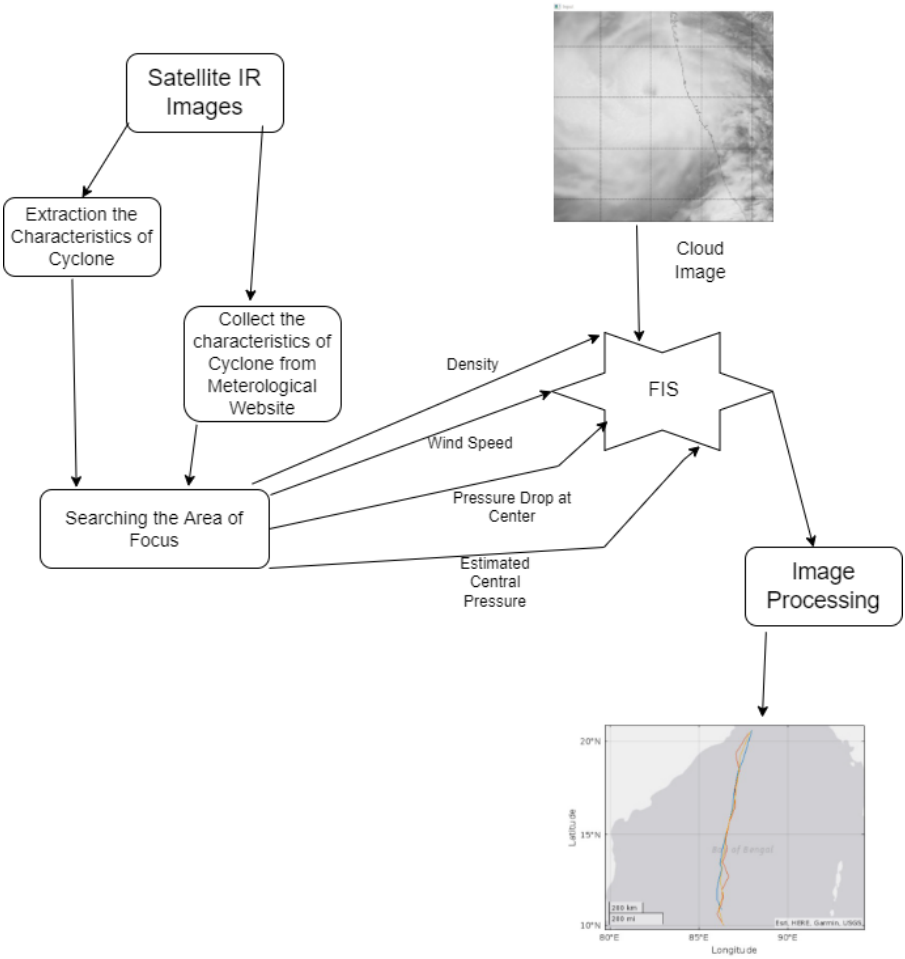



**Fig. 1**: Proposed system

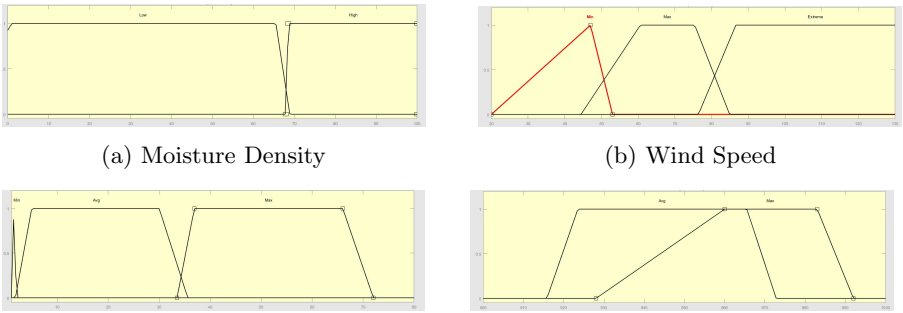
## 3.1 Fuzzy Interference System (FIS)

### 3.1.1 Fuzzification

In this paper, a standard mamdani FIS [1] was proposed to locate storms with expert knowledge. Based on the inputs such as Moisture density, wind speed, pressure drop at the center, and estimated central pressure, Membership Function (MF) are constructed (Figure 2). A detailed description of meteorological input can be found in Table 1.

### 3.1.2 Inference

Within the 300 sub-blocks of the candidate area of focus, fuzzy inference rules are generated for determining the most likely location of the storm eye. In



(a) Moisture Density

(b) Wind Speed

(c) Pressure Drop at the Center

(d) Estimated Central Pressure

**Fig. 2**: Input provide to the proposed system are fuzzy in nature

**Table 1**: A description to the inputs to the proposed FIS method

| Name | Symbol | Range | Unit | Membership Function |
|---|---|---|---|---|
| Moisture Density | D | 1-100 | % | Low, High |
| Wind Speed | W | 20-130 | Knots(kt) | Min, Max, Extreme |
| Pressure Drop at the Center | PD | 1-80 | hPa | Min, Avg, Max |
| Estimated Central Pressure | EP | 900-1000 | hPa | Avg, Max |

Figure 3 , the proposed FIS system is shown as a decision tree. We construct 16 fuzzy IF-THEN based on Multi-Input-Single-Output systems are construction as follows:

1. If (MoistureDensity is low) then (Eye is absent).
2. If (MoistureDensity is High) and (WindSpeed is Min) and (PressureDrop is Min) then (Eye is absent)
3. If (MoistureDensity is High) and (WindSpeed is Min) and (PressureDrop is Avg) and (EstimatedCentralPressure is Avg) then (Eye is absent)
4. If (MoistureDensity is High) and (WindSpeed is Min) and (PressureDrop is Avg) and (EstimatedCentralPressure is Max) then (Eye is present)
5. If (MoistureDensity is High) and (WindSpeed is Min) and (PressureDrop is Max) and (EstimatedCentralPressure is Avg) then (Eye is present)
6. If (MoistureDensity is High) and (WindSpeed is Min) and (PressureDrop is Max) and (EstimatedCentralPressure is Max) then (Eye is absent)
7. If (MoistureDensity is High) and (WindSpeed is Max) and (PressureDrop is Min) then (Eye is ansent)
8. If (MoistureDensity is High) and (WindSpeed is Max) and (PressureDrop is Avg) and (EstimatedCentralPressure is Avg) then (Eye is present)
9. If (MoistureDensity is High) and (WindSpeed is Max) and (PressureDrop is Avg) and (EstimatedCentralPressure is Max) then (Eye is present)
10. If (MoistureDensity is High) and (WindSpeed is Max) and (PressureDrop is Max) and (EstimatedCentralPressure is Avg) then (Eye is present)

11. If (MoistureDensity is High) and (WindSpeed is Max) and (PressureDrop is Max) and (EstimatedCentralPressure is Max) then (Eye is absent)
12. If (MoistureDensity is High) and (WindSpeed is Extreme) and (PressureDrop is Min) then (Eye is absent)
13. If (MoistureDensity is High) and (WindSpeed is Extreme) and (PressureDrop is Avg) and (EstimatedCentralPressure is Avg) then (Eye is present)
14. If (MoistureDensity is High) and (WindSpeed is Extreme) and (PressureDrop is Avg) and (EstimatedCentralPressure is Max) then (Eye is absent)
15. If (MoistureDensity is High) and (WindSpeed is Extreme) and (PressureDrop is Max) and (EstimatedCentralPressure is Avg) then (Eye is present)
16. If (MoistureDensity is High) and (WindSpeed is Extreme) and (PressureDrop is Max) and (EstimatedCentralPressure is Max) then (Eye is absent)

### 3.1.3 Defuzzification

The Center of Area method is the most general form of defuzzification. However, there are several other methods such as the Smallest of Maximas, the Largest of Maximas, and then Mean of Maximas. In Figure 4, we illustrate how FIS can help locate storm eyes. Moisture density is 75%, wind speed is 120, pressure drop at the center is 37.5, and the estimated central pressure is 923 with the crisp inputs of density, wind speed, pressure drop at the center, and the estimated central pressure. We performed COA defuzzification and have obtained 0.703 as our crisp output. It is then concluded that the storm center is indeed within the identified block and the location of eye may subsequently be determined. Image processing should be carried out for eye location refinement in order to achieve an accurate location of the eye, as explained next.

## 3.2 Image processing use for refinement of location of the Eye

In the Image processing process, we refine the location of Eye in order to gain a more accurate forecast, since we known that Eye is one of the main characteristic to predict cyclones.

### 3.2.1 Strategy

1. Raw Data from different sources
   Gathered TCs images from the site of different Meteorological Center like [45, 46]
2. Initialisation of Images

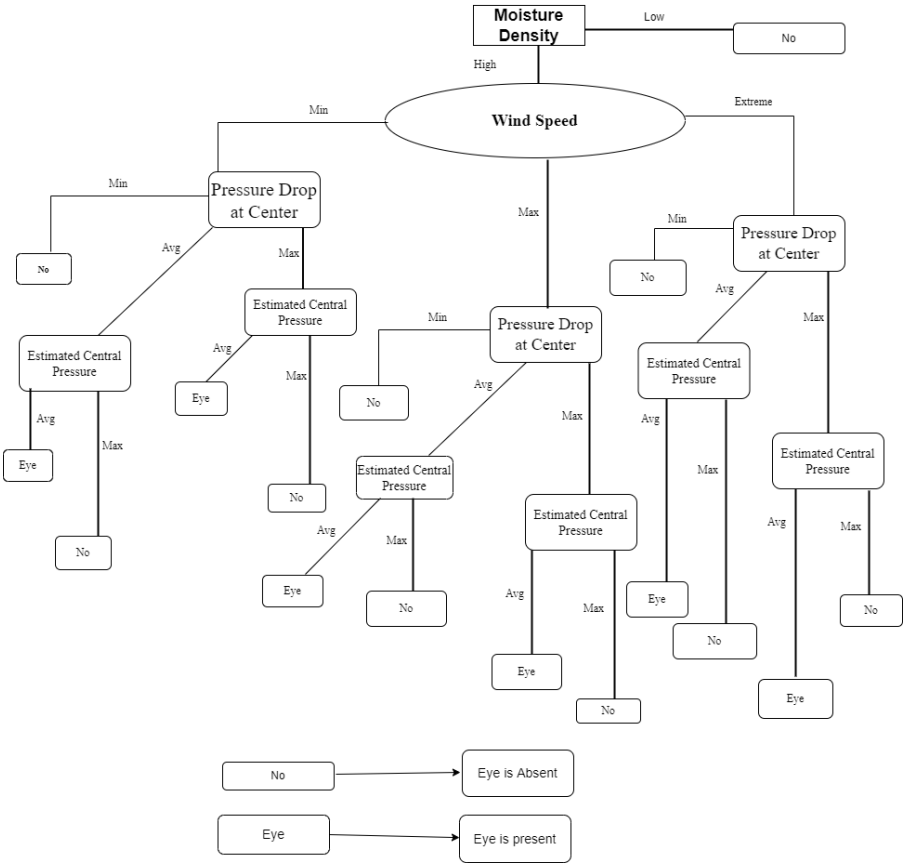



**Fig. 3**: Decision tree for the proposed method

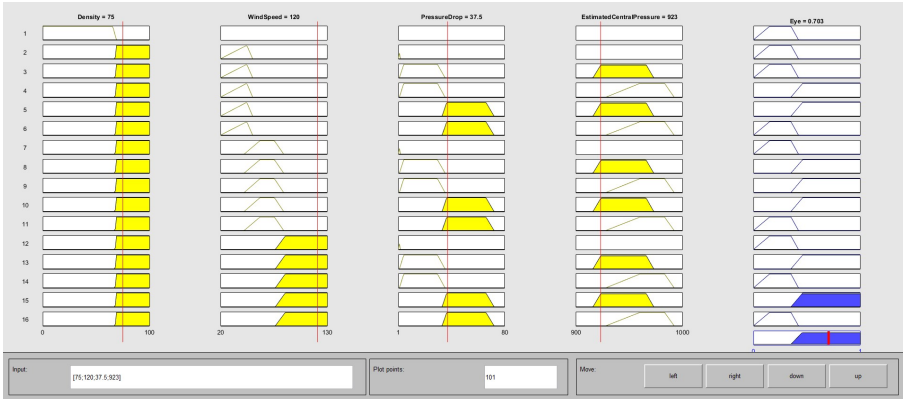



**Fig. 4**: How FIS locate storm eye

- Develop the area of focus and either gather the images of that area or harvest images appropriately.
- Keep up with the consistency between the images like all images is of similar pixels. At first we severe to $150 \times 150$ pixels.
- Changed over images from grayscale to Binary.
- Eliminating the noise from the images by the morphological operation.

3. As the IR image shows, the center of gravity/mass of the storm is supposed to be at its center.

   To calculate the Center of gravity $(c_x, c_y)$ for the image array of $m \times n$ order as follows :

$$c_x = \frac{\sum_m^{i=1} \sum_n^{j=1} i \times C_{ij}}{\sum_m^{i=1} \sum_n^{j=1} C_{ij}}$$

$$c_y = \frac{\sum_m^{i=1} \sum_n^{j=1} j \times C_{ij}}{\sum_m^{i=1} \sum_n^{j=1} C_{ij}}$$

   Among other factors, m refers to the number of horizonal pixels, n to the number of vertial pixels. $C_{ij}$ represents the $(i^{th}, j^{th})$, element of an image.

# 4 Result

Four of the TC cases have been tested with the algorithm discussed above, namely Amphan, Nivar, Tauktae, Yaas. The satellite-derived IR images of these cyclones have been processed and the FIS method employed for determined the eye of TC and Image processing for refinement.

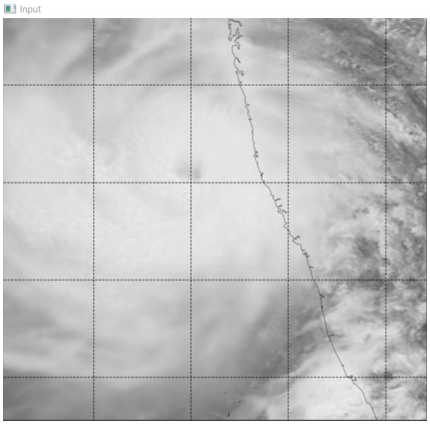
The figure 5 depicts the satellite images after the noise has been removed by using morphological operations, such as erosion and dilation.

On the image, the blue mark represents the center of gravity/mass of the cyclone during that time when satellite take the image, which is illustrated in figure 6.

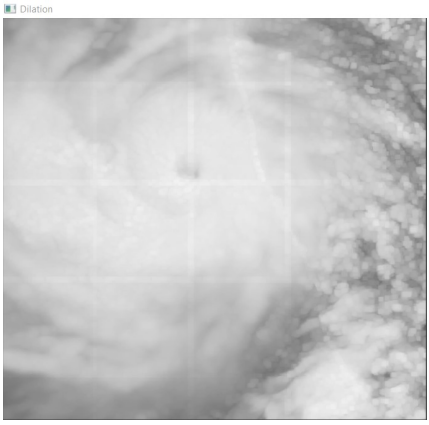
Figure 7 illustrate the track of four TCs derived using the discussed method to determine the eye (center) position, respectively. Comparing the cyclone track obtained from the discussed method with data from CIMSS and IMD shows the cyclone is following a similar path to that observed in satellite data.

An algorithm's accuracy is influenced by several factors, including the focus area, noise removal algorithm, and image processing algorithm. It is important that the center of TC should lie within the focus area in order for the algorithm to produce accurate results.

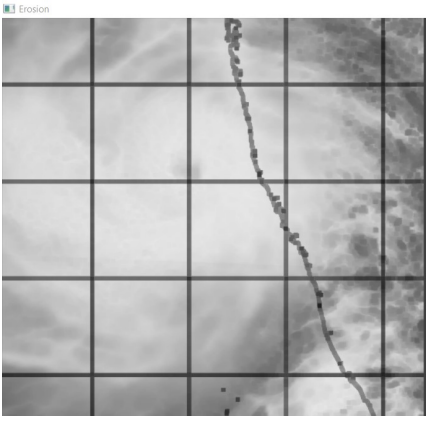
This algorithm starts by performing the fuzzy inference system using Mamdani's fuzzy inference method in MATLAB R2018a on a machine running Windows 11. It requires approximately 60-70 seconds after placing inputs to get the results. As a result of this, Python is used to process the images on the same machine. Noise is removed in 40-50 seconds, while center of gravity/mass is detected in 30-40 seconds.




(a) Input image for removal of noise




(b) Image after removed noise using Dilation morphological operation




(c) Image after removed noise using Erosion morphological operation

**Fig. 5**: Removal of Noise from the Image using Morphological operations

# 5 Conclusion

IR image acquired from the satellites are used in this paper to determine the eye of the TC. The method includes the data processing to extract the eye of the TC using a fuzzy inference system and image processing. The use of this method offers the possibility of establishing the eye in an automated way and can replace the practice of manually determining the center.

A 6-hourly IR image of each of the four TCs listed above is used to test the algorithm.The present method is challenging during the initial phase of TC when the cloud pattern is absent. In this case, the eye of TC is assumed

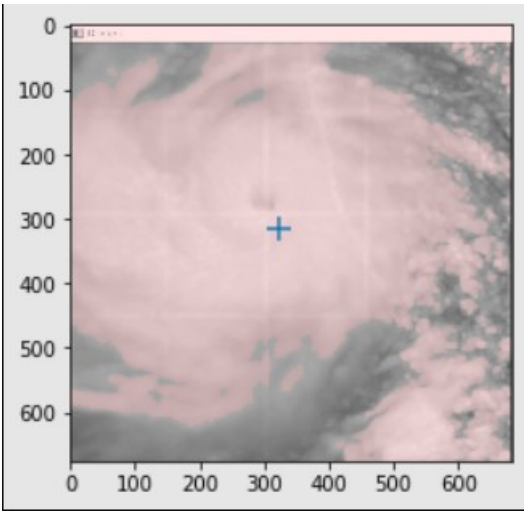



**Fig. 6**: Center of gravity/mass of the cyclone

to be the center of mass of the largest cloud area. More experiments need to be conducted to improve the accuracy and speed of the proposed method.

It will be necessary to test the algorithm on more TCs with different structures such as those with different cloud patterns, wind directions, wind speeds, as well as different oceanographic characteristics before it can used operationally. This information can further be utlilized for estimating TC intensity automatically after the eye of TC has been determine.

# Declarations

## Conflict of Interest

Conflict of interest is not declared by any of the authors.

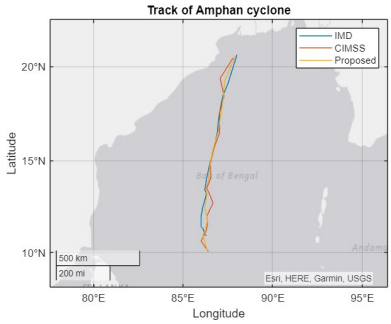
## Author Contribution

Research and manuscript preparation were done equally by each author.

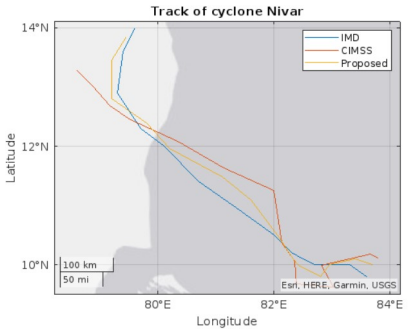
# References

[1] Riza, L.S., Bergmeir, C.N., Herrera Triguero, F., Benítez Sánchez,, J.M.,et al.: frbs: Fuzzy rule-based systems for classification and regression in r. (2015). American Statistical Association

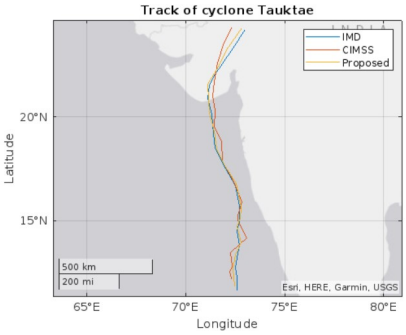
[2] Dvorak, V.F.: Tropical cyclone intensity analysis and forecasting from satellite imagery. Monthly Weather Review 103(5), 420–430 (1975)



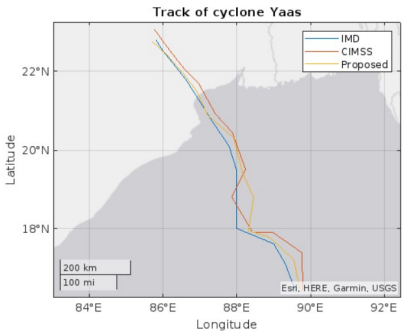

(a) Track of cyclone Amphan by IMD, CIMSS, Proposed Method




(b) Track of cyclone Nivar by IMD, CIMSS, Proposed Method




(c) Track of cyclone Tauktae by IMD, CIMSS, Proposed Method




(d) Track of cyclone Yaas by IMD, CIMSS, Proposed Method

**Fig. 7**: Track of Four cyclone using proposed method and compare with IMD and CIMSS

[3] Wong, K.Y., Yip, C.L., Li, P.W.: Automatic identification of weather systems from numerical weather prediction data using genetic algorithm. Expert systems with Applications 35(1-2), 542–555 (2008)

[4] Zou, J., Lin, M., Xie, X., Lang, S., Cui, S.: Automated typhoon identification from quikscat wind data. In: 2010 IEEE International Geoscience and Remote Sensing Symposium, pp. 4158–4161 (2010). IEEE

[5] Ho, S.-S., Talukder, A.: Automated cyclone identification from remote quikscat satellite data. In: 2008 IEEE Aerospace Conference, pp. 1–9 (2008). IEEE

[6] Kumar, P., Kumar, K.H., Pal, P.K.: Impact of oceansat-2 scatterometer winds and tmi observations on phet cyclone simulation. IEEE transactions on geoscience and remote sensing 51(6), 3774–3779 (2012)

[7] Zhong, J., Fei, J., Huang, S., Du, H., Zhang, L.: An improved quikscat wind retrieval algorithm and eye locating for typhoon. Acta Oceanologica Sinica 31(1), 41–50 (2012)

[8] aiswal, N., Kishtawal, C.M.: Prediction of tropical cyclogenesis using scatterometer data. IEEE transactions on geoscience and remote sensing 49(12), 4904–4909 (2011)

[9] Rodríguez, Pablo García and Polo, María-Eugenia and Cuartero, Aurora and Felicísimo, ÁM and Ruiz-Cuetos, JC, J.: Vecstatgraphs2d, a tool for the analysis of two-dimensional vector data: an example using quikscat ocean winds. IEEE Geoscience and Remote Sensing Letters 11(5), 921–925 (2013)

[10] Lee, R.S., Liu, J.N.: Atmosphere-automatic track mining and objective satellite pattern hunting system using enhanced rbf and egdlm. In: AAAI/IAAI, pp. 603–608 (2000)

[11] Lee, R., Liu, J.: ijade weatherman: a weather forecasting system using intelligent multiagent-based fuzzy neuro network. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 34(3), 369–377 (2004)

[12] Terry, J.P., Feng, C.-C.: On quantifying the sinuosity of typhoon tracks in the western north pacific basin. Applied Geography 30(4), 678–686 (2010)

[13] Bai, Q., Wei, K.: Cloud system extraction in tropical cyclones by mountain-climbing. Atmospheric Research 101(3), 611–620 (2011)

[14] Bai, Q., Wei, K., Jing, Z., Li, Y., Tuo, H., Liu, C.: Tropical cyclone spiral band extraction and center locating by binary ant colony optimization. Science China Earth Sciences 55(2), 332–346 (2012

[15] Jaiswal, N., Kishtawal, C.M.: Objective detection of center of tropical cyclone in remotely sensed infrared images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 6(2), 1031–1035 (2013)

[16] Warunsin, K., Chitsobhuk, O.: Automatic typhoon eye identification using quikscat data and spiral cloud image. In: 2014 14th International Conference on Control, Automation and Systems (ICCAS 2014), pp. 212–216 (2014). IEEE

[17] Warunsin, K., Chitsobhuk, O.: Heuristic search on statistics of wind data and cloud images for automatic typhoon eye location. In: 2015 7th International Conference on Knowledge and Smart Technology (KST), pp. 60–64 (2015). IEEE

[18] Warunsin, K., Chitsobhuk, O.: Heuristic search on statistics of wind data and cloud images for automatic typhoon eye location. In: 2015 7th International Conference on Knowledge and Smart Technology (KST), pp. 60–64 (2015). IEEE

[19] Lakshmanan, V., Rabin, R., DeBrunner, V.: Multiscale storm identification and forecast. Atmospheric research 67, 367–380 (2003)

[20] Pao, T.-L., Yeh, J.-H., Liu, M.-Y., Hsu, Y.-C.: Locating the typhoon center from the ir satellite cloud images. In: 2006 IEEE International Conference on Systems, Man and Cybernetics, vol. 1, pp. 484–488 (2006). IEEE

[21] Piñeros, Miguel F and Ritchie, Elizabeth A and Tyo, J Scott.: Objective measures of tropical cyclone structure and intensity change from remotely sensed infrared image data. IEEE Transactions on Geoscience and Remote sensing 46(11), 3574–3580 (2008)

[22] Piñeros, M.F., Ritchie, E.A., Tyo, J.S.: Estimating tropical cyclone intensity from infrared image data. Weather and forecasting 26(5), 690–698 (2011)

[23] Rodríguez-Herrera, O.G., Wood, K.M., Dolling, K.P., Black, W.T., Ritchie, E.A., Tyo, J.S.: Automatic tracking of pregenesis tropical disturbances within the deviation angle variance system. IEEE Geoscience and Remote Sensing Letters 12(2), 254–258 (2014)

[24] Kossin, J.P., Velden, C.S.: A pronounced bias in tropical cyclone minimum sea level pressure estimation based on the dvorak technique. Monthly weather review 132(1), 165–173 (2004)

[25] Yan, W.K., Lap, Y.C., Wah, L.P., Wan, T.W.: Automatic template matching method for tropical cyclone eye fix. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., vol. 3, pp. 650–653 (2004). IEEE

[26] Lee, R.S., Liu, J.N.: Tropical cyclone identification and tracking system using integrated neural oscillatory elastic graph matching and hybrid rbf network track mining techniques. IEEE Transactions on Neural Networks 11(3), 680–689 (2000)

[27] Harr, P.A., Elsberry, R.L.: Large-scale circulation variability over the tropical western north pacific. part i: Spatial patterns and tropical cyclone characteristics. Monthly Weather Review 123(5), 1225–1246 (1995)

[28] Fang, Y., Sugano, K., Oku, K., Kawagoe, K.: Applying a multi- dimensional time-series similarity method to typhoon-track prediction. In: 2015

IEEE 11th International Conference on e-Science, pp. 259–262 (2015). IEEE

[29] Knaff, J.A.: Revisiting the maximum intensity of recurving tropical cyclones. International Journal of Climatology: A Journal of the Royal Meteorological Society 29(6), 827–837 (2009)

[30] Warunsin, K., Chitsobhuk, O.: Cyclone identification using fuzzy c mean clustering. In: 2013 13th International Symposium on Communications and Information Technologies (ISCIT), pp. 369–373 (2013). IEEE

[31] San Pedro, J., Burstein, F., Sharp, A.: A case-based fuzzy multicriteria decision support model for tropical cyclone forecasting. European Journal of Operational Research 160(2), 308–324 (2005)

[32] Wong, K.Y., Yip, C.L., Li, P.W.: A novel algorithm for automatic tropical cyclone eye fix using doppler radar data. Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling 14(1), 49–59 (2007)

[33] Yurchak, B.S.: Description of cloud-rain bands in a tropical cyclone by a hyperbolic-logarithmic spiral. Russian Meteorology and Hydrology 32(1), 8–18 (2007)

[34] Liu, J.N.-K., Feng, B., Wang, M., Luo, W.: Tropical cyclone forecast using angle features and time warping. In: The 2006 IEEE International Joint Conference on Neural Network Proceedings, pp. 4330–4337 (2006). IEEE

[35] Badarinath, K., Kharol, S.K., Dileep, P., Prasad, V.K.: Satellite observations on cyclone-induced upper ocean cooling and modulation of surface winds—a study on tropical ocean region. IEEE Geoscience and Remote Sensing Letters 6(3), 481–485 (2009)

[36] Hawkins, J.D., Turk, F.J., Lee, T.F., Richardson, K.: Observations of tropical cyclones with the ssmis. IEEE transactions on geoscience and remote sensing 46(4), 901–912 (2008)

[37] Piñeros, M.F., Ritchie, E.A., Tyo, J.S.: Objective measures of tropical cyclone structure and intensity change from remotely sensed infrared image data. IEEE Transactions on Geoscience and Remote sensing 46(11), 3574–3580 (2008)

[38] Fritz, S., Hubert, L.F., Timchalk, A.: Some inferences from satellite pictures of tropical disturbances. Monthly Weather Review 94(4), 231–236 (1966)

[39] Gentry, R.C., Fujita, T.T., Sheets, R.C.: Aircraft, spacecraft, satellite and radar observations of hurricane gladys, 1968. Journal of Applied Meteorology and Climatology 9(6), 837–850 (1970)

[40] Rao, B., Kishtwal, C., Pal, P., Narayanan, M.: Ers-1 surface wind observations over a cyclone system in the bay of bengal during november 1992. Remote Sensing 16(2), 351–357 (1995)

[41] Chaurasia, S., Kishtawal, C., Pal, P.: An objective method of cyclone centre determination from geostationary satellite observations. International Journal of Remote Sensing 31(9), 2429–2440 (2010)

[42] Tuttle, J., Gall, R.: A single-radar technique for estimating the winds in tropical cyclones. Bulletin of the American Meteorological Society 80(4), 653–668 (1999)

[43] Hasler, A., Palaniappan, K., Kambhammetu, C., Black, P., Uhlhorn, E., Chesters, D.: High-resolution wind fields within the inner core and eye of a mature tropical cyclone from goes 1-min images. Bulletin of the American Meteorological Society 79(11), 2483–2496 (1998)

[44] Yan, W.K., Lap, Y.C.: An intelligent tropical cyclone eye fix system using motion field analysis. In: 17th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'05), p. 5 (2005). IEEE

[45] CIMSS, url=http://tropic.ssec.wisc.edu/tropic.php

[46] Satellite images, url=http://satellite.imd.gov.in/insat.htm

Eur. Phys. J. Plus (2022) 137:823
https://doi.org/10.1140/epjp/s13360-022-03005-x

THE EUROPEAN PHYSICAL JOURNAL PLUS

Regular Article

# Impact of plasma process parameters on the growth of vertically aligned carbon nanotube array and its optimization as field emitters

**Umang Sharma, Suresh C. Sharma**[a]

Plasma and Nano-Simulation Lab, Department of Applied Physics, Delhi Technological University (DTU), Bawana Road, New Delhi, Delhi 110042, India

Received: 12 May 2022 / Accepted: 24 June 2022
© The Author(s), under exclusive licence to Società Italiana di Fisica and Springer-Verlag GmbH Germany, part of Springer Nature 2022

**Abstract** The present theoretical study models a vertically aligned carbon nanotube (VACNT) array in plasma and structures its growth based on the variation of plasma operating conditions and process parameters. We investigate the consequence of plasma concentration, power, pressure and substrate temperature on the aspects of the VACNT array. Furthermore, the amorphous carbon deposition between the CNTs of the array is also examined. As field emission is one of the most outstanding applications of the VACNT array, the field enhancement factor for the array is also deliberated and optimized. The results are modelled, considering the balance of plasma species, plasma and CNT energy exchange, hydrocarbon and hydrogen generation over the catalyst nanoparticle, CNT array growth, and carbon deposition over the substrate between CNTs in VACNT array growth using plasma-enhanced chemical vapour deposition. The model is numerically solved by deliberating the experimental literature's initial conditions and plasma parameters. The aspects of an array, i.e. its length and average CNT diameter, can be optimized as desired by altering the operating conditions and glow discharge parameters. Additionally, the influence of array aspects on the field emission properties is observed. The results of the study are validated by the available experimental data. This theoretical study can be effectively used to grow the VACNT array and its optimization to obtain better field emitters.

## 1 Introduction

Carbon nanotubes (CNTs) have been extensively deliberated worldwide and recognized as one of the most attractive nanomaterials. They show some extraordinary and unique electrical properties, high conductivity, and incredible strength, which holds tremendous potential for its use in electrical, thermal, biomedical and chemical applications. The synthesis of CNTs can be attained via arc discharge [1, 2], laser ablation [3, 4], chemical vapour deposition (CNT) [5, 6] and plasma-enhanced vapour deposition (PECVD) [7–9]. The plasma activation prevents thermal damage to the substrate as PECVD operates at low temperatures and allows better vertical alignment due to the plasma electric field [10]. The alignment of CNTs is critical for various applications, and vertically aligned carbon nanotubes (VACNTs) are stated to be used for electronic applications like field emission displays, field-effect transistor channels, and quantum wires [11–13]. The synthesis of VACNTs forest is obtained via various processes such as thermal CVD (TCVD), one-step catalytic chemical vapour deposition (CCVD), and PECVD over a metal catalyst. As mentioned before, PECVD can produce VACNT arrays at lower temperatures. The VACNTs have good alignment, uniform length, and coverage of the catalyst–substrate surface. The properties of VACNT arrays can be influenced by their pattern and intertube distance and are demonstrated in fibre lasers, photonics, solar and nano-applications (as a vertical CNT-field effect transistor). As the VACNT arrays deteriorate the shielding of the electric field, thus improving the electric field distribution, the upkeep of closely packed VACNT array areal density is essential for its use in field emission and nanoelectrode arrays [14]. The pre-patterned catalyst sites for the growth of nanotubes are used to decrease the areal density. This vertical alignment of carbon nanotubes results from an electrostatic force acting on growing nanotubes' tips, resulting from the sheath electric field.

The growth of CNT forest is a consequence of continuous carbon species diffusion through the catalyst surface either via tip-mode growth (where the nanoparticle is detected on the nanotube tip) or base-mode growth (where the nanoparticle is detected at the nanotube base, i.e. on the substrate). A catalyst metal thin film (commonly Fe, Ni, Co) deposited over the substrate fragments into catalyst nanoparticles upon heating, serving as a seed layer for nanotube nucleation and growth [15]. The feedstock gases dissociate due to plasma power and over the catalyst surface into carbon species, which stimulates the growth of the nanotubes. Several dissociations and diffusion processes are liable for carbon species movement over the catalyst. Various groups have acknowledged the base-mode growth of VACNTs, suggesting the feedstock gases transport through the CNT array to reach the catalyst located on the substrate and contribute to the growth. Although the CVD growth of VACNTs has been enlightened experimentally and theoretically, the studies done for the plasma-assisted growth of CNT arrays are ambiguous.

[a] e-mail: suresh321sharma@gmail.com (corresponding author)

Published online: 17 July 2022

Kato et al. [16] demonstrated a pioneering work by synthesizing single-walled CNTs (SWCNTs) in the form of SWCNT bundles and individual SWCNT using PECVD. Hoffman et al. [17] revealed a controlled PECVD synthesis of VA-carbon nanofibers on pre-marked substrates at very low temperatures, i.e. 120 °C, suggesting a diffusion growth mechanism. Srivastava et al. [18] synthesized CNTs using the MPECVD (microwave plasma-enhanced chemical vapour deposition) technique to study the influence of gas composition on their growth. They also synthesized nitrogen-doped CNT films using a similar technique to study their growth morphology and field emission characteristics [19]. Hayashi et al. [20] grew well-aligned CNTs under the influence of a high plasma sheath electric field and pressure for the rf-dc PECVD method. Bell et al. [21, 22] synthesized a VACNT film via a DC PECVD set-up to study the plasma chemistry of species, calculated the optimum gas ratio for nanotube film growth and dominant precursor. In our previous work [23], we have developed a parametric model for the vertical alignment of a CNT, where we established that an electrostatic force acting on the CNT during its growth is responsible for its vertical alignment in plasma. We have also studied the effect of process parameters such as plasma temperature, plasma density, and negative substrate voltage on the growth and alignment of CNTs.

Okuyama et al. [24] made an effort to control the density and diameter of VACNTs in a DC-plasma set-up using a Ni catalyst. They suggested that the density can be maximized by optimizing the plasma pre-treatment time. Ikuno et al. [25] examined the relation between film morphology and PECVD plasma conditions. They proposed that the growth was stimulated due to increasing inter-electrode distance, which increases the generation of radicals and increases in self-bias, which induces vertical alignment in nanotubes. Lee et al. [26] provided new insights into the CVD growth of CNT forest and hypothesized that the film growth is mass transport limited. Louchev et al. [27] provided a detailed analysis of CVD growth of nanotubes involving the diffusion processes over metal catalyst surfaces. Zhong et al. [28] synthesized VA-SWCNT mats via point-arc microwave plasma chemical vapour deposition over Fe catalyst and confirmed that gas-phase diffusions of feedstock gases bound the film growth rate.

Few applications such as field emitter devices require a patterned growth of CNTs to avoid any screening effects of high-density films. The spacing between the CNTs must be high to disregard the field screening effect. Thapa et al. [29] synthesized VACNT arrays on a stainless-steel substrate in a dc-PECVD set-up and studied field emission characteristics of these arrays. They also simulated the effect of bundling of CNTs and observed that field emission from individual CNTs or less dense CNT bundles was better in comparison with denser CNT bundles. Campo et al. [30] synthesized SWNTs via alcohol catalytic chemical vapour deposition (AC-CVD) on TiN conductive substrates and further tested them for field emission, concluding that SWNTs grown directly on these substrates showed excellent field emission characteristics. Morassutto et al. [31] synthesized dense arrays of aligned CNTs via thermal CCVD over Fe catalyst on a thin Ta layer. They observed that the catalyst particle size and density govern the growth of VACNTs and addition of Au or Pt between substrate and Ta layer increases the CNT conductivity and improves field emission.

Various studies have already established that plasma-assisted growth of nanotubes provides high quality and uniformly deposited array. Also, using highly ionized gaseous fluxes enhances the array growth uniformity and quality, which are critical for their electron-emitting and dimensional characteristics; hence, studying the effects of plasma parameters on VACNTs growth becomes significant. This study investigates the impact of plasma conditions on VACNT array growth in a PECVD set-up. We propose that the catalyst is attached to the substrate during the growth so that the gaseous species are required to travel through the growing CNT array, i.e. the growth occurs via base mode. CNT growth only occurs at the catalyst sites, enabling a patterned or selective growth approach. Taking into account the characteristics of bulk plasma species and the chamber conditions, we have provided a kinetic mechanism model, which examines the growth of nanotube arrays considering various nucleation and growth processes like surface and volume diffusion of carbon species, precipitation of these species to create nanotube interlayers, and formation of amorphous carbon layer (other processes mentioned in Sect. 2). The geometrical field enhancement factor of the array is also studied. The outcomes of the analytical model calculations are discussed comprehensively in Sect. 3, and lastly, in Sect. 4, the study is concluded.

## 2 Detailed model description

The following section is an extensive study of the growth model of a VACNT film in a PECVD set-up to provide a deep understanding of the kinetics of all the species involved and various nucleation and growth processes involved. The feedstock gases considered in the model are acetylene ($C_2H_2$) as a hydrocarbon and hydrogen ($H_2$) as a carrier gas, respectively.

The growth of VACNT arrays in a complex reactive plasma follows the scenario illustrated schematically in Fig. 1—A metal catalyst film deposited over a substrate surface under the effect of hydrogen plasma gets fragmented into nanoparticles, which are nucleation sites for the carbon nanotubes. The plasma species are dissociated into carbon species owing to many dissociation processes and surface decomposition over catalyst surface. These carbon atoms diffuse over the catalyst's surface and volume, initiating CNT growth. This growth begins when the carbon precursors deposit on the substrate area between the catalyst nanoparticles and diffuse into and over the catalyst until the nanoparticles become saturated with carbon. Once the saturation of metal catalysts is achieved, nanotubes start to nucleate on their surface by increasing their proportions. Few carbon atoms get deposited on the substrate surface due to the diffusion of gaseous fluxes between these growing nanotubes leaving a deposition layer (typically amorphous carbon layer). These nanotubes continue to grow until the metal catalysts get poisoned.

A plasma sheath region is formed near the substrate (space charge separation region), where a negative bias is supplied. This region starts at $z = 0$, and substrate is located at $z = \lambda_s$, where the sheath potential equals the substrate potential. An electric

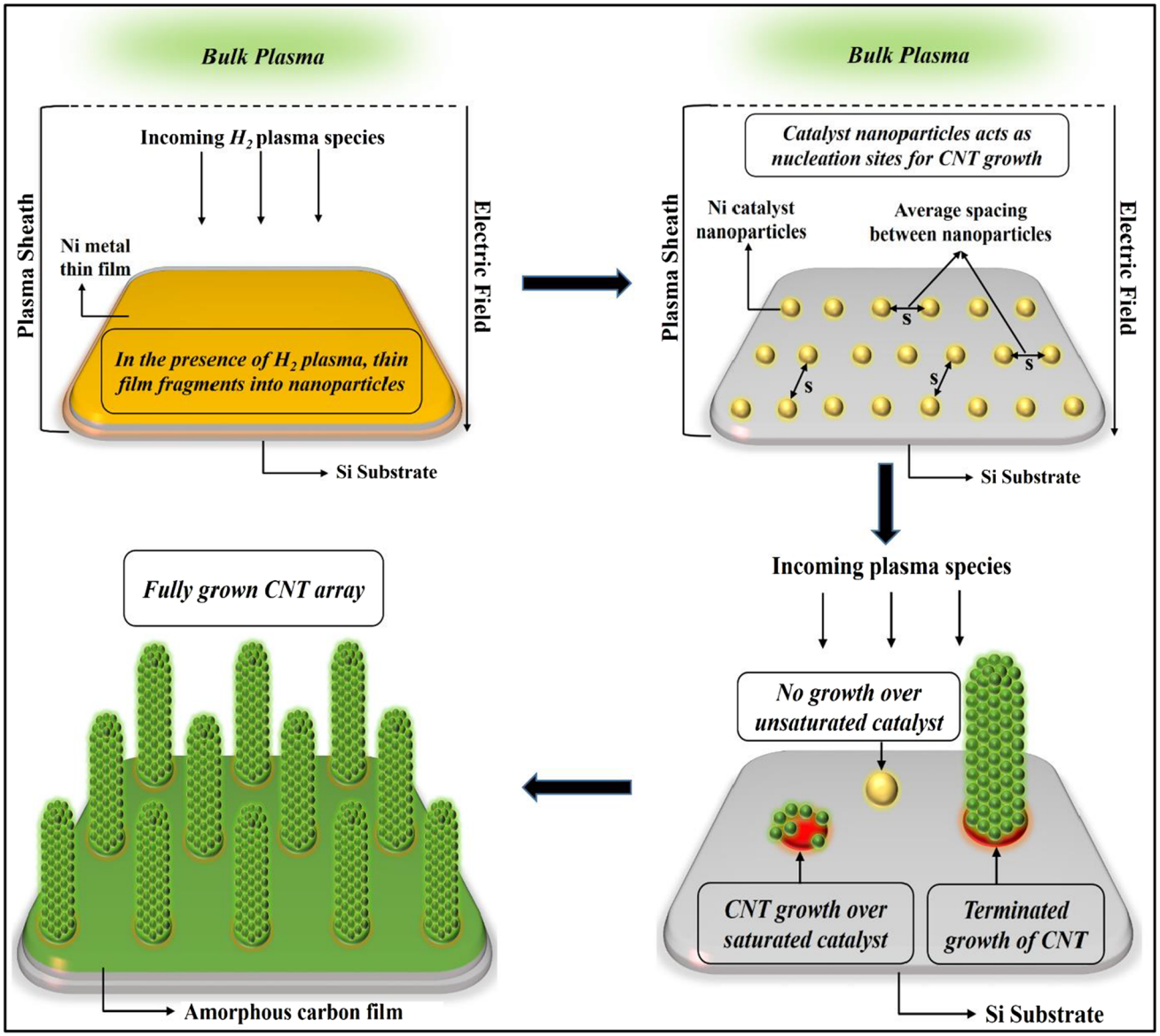



**Fig. 1** Schematic illustration of CNT array growth in plasma

**Table 1** Gaseous species deliberated in the study

| Type | *A* | *B* |
|---|---|---|
| Neutral species | $C_2H_2$, $C_2H$, $C_2H_3$, $C_2H_4$, $C_4H$, $C_4H_3$, $C_4H_2$, $C_6H$, $C_6H_4$, $C_6H_2$ (10) | H, $H_2$ (2) |
| Positive species | $C_2H_2^+$, $C_2H^+$, $C_2H_3^+$, $C_2H_4^+$, $C_4H^+$ $C_4H_2^+$, $C_4H_3^+$, $C_6H_2^+$, $C_6H^+$, $C_6H_3^+$ (10) | $H^+$, $H_2^+$ (2) |

sheath field is established due to the charge separation that prevails along the $z$-axis, which directs all the ionic species towards the negatively biased substrate, while the electrons are averted. Following the 1D fluid steady-state model, the particle energy and momentum balance equations are calculated from Boltzmann's transport equation, and the electric field is calculated from Poisson's equation combined with balance equations to make the model self-consistent. The ionization and collisional effects of electrons and ions with the neutral species are considered; however, ion–electron collisions are neglected. The temperatures of electrons, ionic, and neutral species are presumed to be constant throughout the plasma sheath, and additionally, ions and neutrals have an equivalent temperature, i.e. $T_i = T_n$. The study considers acetylene ($C_2H_2$) as a hydrocarbon gas and hydrogen ($H_2$), denoted as *A* and *B* in the script. Only dominant gaseous species are considered in the model, as mentioned in Table 1.

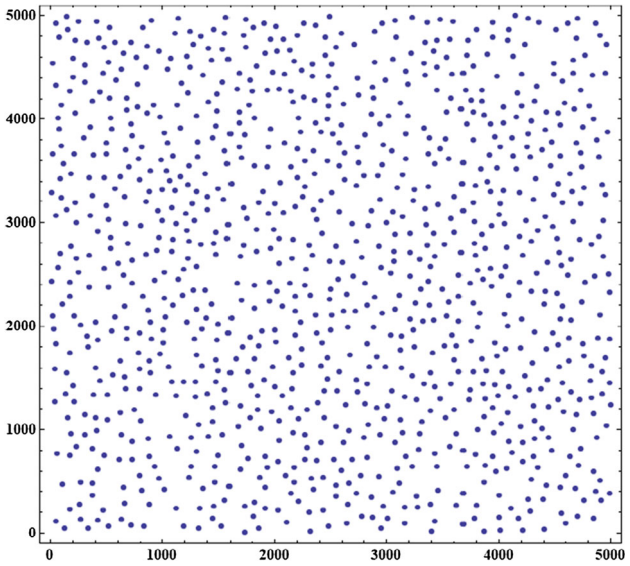



**Fig. 2** Computational domain with 1000 catalytic nanoparticles over $5000 \times 5000$ μm.$^2$ substrate area

For all the plasma species (ions, neutrals, and electrons), the balance equations are represented as [32, 33]:

$$\frac{\partial}{\partial z}\left(n_{ij}u_{ij}\right) = f_{ij}n_e, \tag{1}$$

$$\frac{\partial}{\partial z}(n_e u_e) = \sum_j f_{ij}n_e, \tag{2}$$

$$m_{ij}n_{ij}u_{ij}\left(\frac{\partial u_{ij}}{\partial z} + f_{cj}\right) = -n_{ij}\left(e\frac{\partial \varphi}{\partial z} + \frac{T_{ij}}{n_{ij}}\frac{\partial n_{ij}}{\partial z}\right), \tag{3}$$

$$m_e u_e f_{ce} = \left(e\frac{\partial \varphi}{\partial z} - \frac{T_e}{n_e}\frac{\partial n_e}{\partial z}\right). \tag{4}$$

Using Poisson's equation, we obtain the potential $\varphi$ and electric field $E$ as—

$$\frac{\mathrm{d}^2\varphi}{\mathrm{d}z^2} = -\frac{e}{\varepsilon_o}\left(\sum_{ij}^{A,B}\sigma_{ij}n_{ij} - n_e\right), \quad E = -\frac{\mathrm{d}\varphi}{\mathrm{d}z}. \tag{5}$$

where $\sigma_{ij} = \left\{\begin{array}{l}\sum \sigma_{ij} = 0 \\ 0 < \sigma_{ij} < 1\end{array}\right\}$ defines the $j$th ion–electron ratio. Using the solutions of Eqs. (1–5), the plasma sheath is investigated around the substrate region and essential parameters of the CNT growth are determined with the following boundary conditions –

$$u_e|_z = 0, \;\; u_{ij}\big|_z = v_{ij0}\big|_z\left(= \frac{T_\mathrm{e}}{m_{ij}}\right), \;\; \varphi = 0, \;\; \frac{\partial \varphi}{\partial z} = -\frac{T_\mathrm{e}}{e\lambda_i},$$

where $\lambda_i = \frac{T_\mathrm{n}}{\sigma p_0}$ and $v_{ij}$ and $p_0$ are the ion-acoustic speed and gas pressure, respectively.

In this complex numerical model, we have presumed that the catalyst nanoparticles remain anchored to the substrate throughout the nanotube growth, i.e. the base-mode growth of CNT is deliberated. As perceived from Fig. 2, on a substrate with an area of $5000 \times 5000$ μm$^2$, a pre-patterned Ni metal nanoparticle film is considered with 1000 metal catalysts with an average diameter of 40 nm and a minimum average spacing of 100 nm (centre-to-centre distance between catalysts). The catalyst nanoparticles act as nucleation sites for carbon nanotubes; hence, 1000 nanotubes for this pre-patterned catalyst film would be accomplished. The main equations accounting for the nucleation and CNT array growth in a reactive plasma are preceded by the bulk plasma balance equations taking into account all the process changes loss during the growth. Other predetermined parameters considered for the computation are mentioned in Sect. 2.

2.1 Formation of hydrogen radicals and carbon species over the catalyst nanoparticle

$$\sum_{p=1}^{N}\partial_\tau(n_C^p) = \Bigg[\underbrace{\left\{\sum_{p=1}^{N}\left[\sum_{A=1}^{10}(1-\theta_t)J_A^p\right]+\sum_{p=1}^{N}\left[\sum_{A=1}^{10}\overset{\text{surf}^p}{n}_A\,\upsilon\exp\left(-\frac{\delta E_{\text{CH}}^{\text{diss}}}{k_{\text{B}}T_{\text{S}}}\right)\right]+\sum_{p=1}^{N}\left[\sum_{iA=1}^{10}\left(\sum_{A=1}^{10}\overset{\text{surf}^p}{n}_A\frac{y_d}{\upsilon_0}\right)J_{iA}^p\right]\right.}_{\text{First Part (Gain)}}$$
$$\underbrace{\left.+\sum_{p=1}^{N}\left(\sum_{iA=1}^{10}J_{iA}^p\right)+\sum_{p=1}^{N}J_C^p+\sum_{p=1}^{N}\left[\sum_{iB=1}^{2}\left(\sum_{iA=1}^{10}\frac{J_{iA}^p\sigma_{\text{ads}}}{\upsilon}\right)J_{iB}^p\right]\right\}}_{\text{First Part (Gain)}}$$
$$\left.-\underbrace{\left\{\sum_{p=1}^{N}\left[\sum_{A=1}^{10}\overset{\text{surf}^p}{n}_A\,\upsilon\exp\left(-\frac{\delta E_{\text{dCH}}}{k_{\text{B}}T_{\text{S}}}\right)\right]+\sum_{p=1}^{N}\left[\sum_{B=1}^{2}\left(\sum_{A=1}^{10}\overset{\text{surf}^p}{n}_A\sigma_{\text{ads}}\right)J_B^p\right]\right\}}_{\text{Second Part (Loss)}}\right]$$
$$\left.+\underbrace{\left\{\sum_{p=1}^{N}\left[n_C^p\upsilon\exp\left(-\frac{\delta E_{\text{ev}}}{k_{\text{B}}T_{\text{S}}}\right)\right]+\sum_{p=1}^{N}\left(n_C^p\sigma_{\text{ads}}\upsilon n_H^p\right)\right\}}_{\text{Second Part (Loss)}}\right]\times d_\tau(\Omega_C) \tag{6}$$

Equation (6) describes the carbon atoms produced over the $p$th catalyst nanoparticle's surface. The increase in carbon atom density is described in the first part of the equation. This gain is ascribed to the hydrocarbon neutrals absorption over the nanoparticle's surface, hydrocarbon thermal dissociation, hydrocarbon ion-induced dissociation, positive ion decomposition, direct carbon influx onto the nanoparticle, and hydrocarbon–hydrogen ion interaction, respectively. The decrease in carbon atom density is described in the second part of the equation. This loss is ascribed to the hydrocarbon desorption from the nanoparticle's surface, the interaction of hydrocarbon–hydrogen neutrals, carbon species evaporation from the nanoparticle, and interaction between catalyst-surface-created carbon and hydrogen species. The incessant hydrocarbon dissociation over the $p$th nanoparticle's surface generates an enormous amount of amorphous carbon that forms a vague layer over it, which depletes the active area of the catalyst and gradually saturates the CNT growth. Hence, the last part of Eq. (6) administers the catalyst contamination rate due to the creation of this carbon layer as a function of carbon number density.

$$\sum_{p=1}^{N}\partial_\tau(n_H^p) = \underbrace{\left\{\sum_{p=1}^{N}\left(\sum_{B=1}^{2}J_B^p(1-\theta_t)\right)+\sum_{p=1}^{N}\left[\sum_{B=1}^{2}\overset{\text{surf}^p}{n}_B\,\upsilon\exp\left(-\frac{\delta E_{\text{CH}}^{\text{diss}}}{k_{\text{B}}T_{\text{S}}}\right)\right]+\sum_{p=1}^{N}\left[\sum_{iA=1}^{10}\left(\sum_{A=1}^{10}\overset{\text{surf}^p}{n}_A\frac{y_d}{\upsilon_0}\right)J_{iA}^p\right]\right.}_{\text{First Part (Gain)}}$$
$$\underbrace{\left.+\sum_{p=1}^{N}\left(\sum_{iB=1}^{2}J_{iB}^p\right)\right\}}_{\text{First Part (Gain)}}-\underbrace{\left\{\sum_{p=1}^{N}\left[\sum_{B=1}^{2}\overset{\text{surf}^p}{n}_B\,\upsilon\exp\left(-\frac{\delta E_{\text{dH}}}{k_{\text{B}}T_{\text{S}}}\right)\right]+\sum_{p=1}^{N}\left(\sum_{B=1}^{2}\overset{\text{surf}^p}{n}_B\sigma_{\text{ads}}J_B^p\right)\right.}_{\text{Second Part (Loss)}}$$
$$\underbrace{\left.+\sum_{p=1}^{N}\left[\sum_{iB=1}^{2}\left(\sum_{B=1}^{2}\overset{\text{surf}^p}{n}_B\sigma_{\text{ads}}J_{iB}^p\right)\right]+\sum_{p=1}^{N}n_H^p\sigma_{\text{ads}}\upsilon n_C^p\right\}}_{\text{Second Part (Loss)}} \tag{7}$$

Equation (7) describes the hydrogen atoms produced over the $p$th catalyst nanoparticle's surface. The increase in hydrogen atom density is described in the first part of the equation. This gain is ascribed to the hydrogen neutrals adsorption over the nanoparticle's surface, hydrocarbons' thermal dissociation, hydrocarbons ion-induced dissociation, and the influx of hydrogen ion

**Table 2** Particulars of functions/parameters termed in Eqs. (6–7)

| Parameters/functions | Description |
|---|---|
| $\upsilon_0 (\approx 10^{15}\ \mathrm{cm}^{-2})$ | Adsorption sites per unit area [34] |
| $\sigma_{\mathrm{ads}} \left(= 6.8 \times 10^{-16}\ \mathrm{cm}^2\right)$ | Cross section of plasma species interaction [34] |
| $\overset{\mathrm{surf}}{n}_j \left(= \theta_j \upsilon_0\right)$ | Surface density of the $j = (A, B)$ neutral atom [34] |
| $\upsilon\ (\approx 10^{13}\ \mathrm{Hz})$ | Thermal vibration frequency [35] |
| $\delta E_{\mathrm{CH}}^{\mathrm{diss}} (= 2.1\ \mathrm{eV}),\ \delta E_{\mathrm{ev}} (= 1.8\ \mathrm{eV}),$ $\delta E_{\mathrm{dCH}} (= 1.8\ \mathrm{eV}),\ \delta E_{\mathrm{dH}} (= 1.8\ \mathrm{eV})$ | Hydrocarbon dissociation, C-atom evaporation, hydrocarbon desorption and hydrogen desorption energies [33] |
| $d_\tau(\Omega_C) = \left(1 - n_C \pi d_p^2\right)$ | Rate of contamination of catalyst with $\Omega_C$ as catalyst particle blocking function [23] |

on $p$th catalyst. The decrease in hydrogen atom density is described in the second part of the equation. This loss is ascribed to the hydrogen desorption from the $p$th nanoparticle's surface, the surface-adsorbed hydrogen and neutral hydrogen interaction, the surface-adsorbed hydrogen and ionic hydrogen interaction, and interaction between catalyst-surface-created carbon and hydrogen species. The functions/parameters stated in Eqs. (6–7) are termed in Table 2.

## 2.2 Plasma species equilibrium in the glow discharge (neutrals, ions, electrons)

$$\begin{aligned}
&\partial_\tau (n_A) \\
&= \underbrace{\left\{ \alpha_A n_e n_{iA} + \sum_{p=1}^{N} \frac{n_{\mathrm{CNT}}}{\lambda_s} (1 - \gamma_{iA}) \left( \overset{\mathrm{tip}^p}{Q}_{iA} + \overset{\mathrm{cyl}^p}{Q}_{iA} \right) + \sum_{p=1}^{N} \overset{\mathrm{desp}^p}{J}_A + \left( V_A^{\mathrm{IF}} - V_A^{\mathrm{OF}} \right) + \left( \sum_{qAB} k_q n_B n_{iA} - \sum_{lBA} k_l n_A n_{iB} \right) \right\}}_{\text{First Part (Gain)}} \\
&\quad - \underbrace{\left\{ \beta_A n_A + \sum_{p=1}^{N} \frac{n_{\mathrm{CNT}}}{\lambda_s} \gamma_A \left( \overset{\mathrm{tip}^p}{Q}_{A} + \overset{\mathrm{cyl}^p}{Q}_{A} \right) + \sum_{p=1}^{N} \overset{\mathrm{adsp}^p}{J}_A + \mathrm{K}_{\mathrm{wall}}^A n_A \right\}}_{\text{Second Part (Loss)}}
\end{aligned} \tag{8}$$

$$\begin{aligned}
&\partial_\tau (n_B) \\
&\underbrace{\left\{ \alpha_B n_e n_{iB} + \sum_{p=1}^{N} \frac{n_{\mathrm{CNT}}}{\lambda_s} (1 - \gamma_{iB}) \left( \overset{\mathrm{tip}^p}{Q}_{iB} + \overset{\mathrm{cyl}^p}{Q}_{iB} \right) + \sum_{p=1}^{N} \overset{\mathrm{desp}^p}{J}_B + \left( V_B^{\mathrm{IF}} - V_B^{\mathrm{OF}} \right) + \left( \sum_{\ell AB} k_\ell n_A n_{iB} - \sum_{qBA} k_q n_B n_{iA} \right) \right\}}_{\text{First Part (Gain)}} \\
&\quad - \underbrace{\left\{ \beta_B n_B + \sum_{p=1}^{N} \frac{n_{\mathrm{CNT}}}{\lambda_s} \gamma_B \left( \overset{\mathrm{tip}^p}{Q}_{B} + \overset{\mathrm{cyl}^p}{Q}_{B} \right) + \sum_{p=1}^{N} \overset{\mathrm{adsp}^p}{J}_B + \mathrm{K}_{\mathrm{wall}}^B n_B \right\}}_{\text{Second Part (Loss)}} + J_{\mathrm{thd}}
\end{aligned} \tag{9}$$

Equations (8) and (9) describe the evolution of neutral density $j = (A, B)$ type in bulk plasma as a consequence of various processes. The gain in neutral atom density represented in the first part of the equations is attributed to ion–electron recombination, ion neutralization on sticking to the nanotube surface, desorption of ions from catalyst surface, variation in chamber inflow and outflow of neutrals, and influence of ion-neutral reactions considered in the study (listed in Table 3), respectively.

The loss in neutral atom density represented in the second part of the equations is due to the dissociative neutral ionization, neutral accretion at the nanotube surface, neutrals species adsorption, and reactive chamber wall discharge loss of neutrals, respectively. The final term in Eq. (9) analyses the rise in carrier gas neutrals density due to thermal dehydrogenation. The specifics of the equations mentioned above are given in Table 4.

**Table 3** Ion–neutral reactions with their reaction rates considered in the study [36]

| Reactions | $k$ ($cm^3 sec^{-1}$) |
|---|---|
| $H^+ + C_4H \rightarrow C_4H^+ + H$ | $2.0 \times 10^{-9}$ |
| $H^+ + C_6H \rightarrow C_6H^+ + H$ | $2.0 \times 10^{-9}$ |
| $H^+ + C_2H_2 \rightarrow C_2H^+ + H_2$ | $4.3 \times 10^{-9}$ |
| $H^+ + C_4H_2 \rightarrow C_4H^+ + H_2$ | $2.0 \times 10^{-9}$ |
| $H_2^+ + H \rightarrow H^+ + H_2$ | $6.4 \times 10^{-10}$ |
| $H_2^+ + C_2H_2 \rightarrow C_2H_2^+ + H_2$ | $5.3 \times 10^{-9}$ |
| $H_2^+ + C_4H \rightarrow C_4H_2^+ + H$ | $1.7 \times 10^{-10}$ |
| $H_2 + C_2H^+ \rightarrow C_2H_2^+ + H$ | $1.7 \times 10^{-9}$ |
| $C_2H + C_4H^+ \rightarrow C_6H^+ + H$ | $6.0 \times 10^{-10}$ |
| $C_2H + C_4H_2^+ \rightarrow C_6H_2^+ + H$ | $1.3 \times 10^{-9}$ |
| $C_2H_2 + C_2H^+ \rightarrow C_4H_2^+ + H$ | $1.2 \times 10^{-9}$ |
| $C_2H_2 + C_2H_3^+ \rightarrow C_4H_3^+ + H_2$ | $2.4 \times 10^{-10}$ |
| $C_2H_2 + C_4H^+ \rightarrow C_6H_2^+ + H$ | $1.5 \times 10^{-9}$ |
| $C_2H_2 + C_4H_2^+ \rightarrow C_6H_4^+$ | $1.4 \times 10^{-10}$ |
| $C_2H_2 + C_2H_2^+ \rightarrow C_4H_3^+ + H$ | $9.5 \times 10^{-10}$ |
| $C_2H_2 + C_2H_2^+ \rightarrow C_4H_2^+ + H_2$ | $1.2 \times 10^{-9}$ |
| $C_2H_2^+ + H_2 \rightarrow C_2H_3^+ + H$ | $1.0 \times 10^{-11}$ |
| $C_2H_2^+ + C_6H_2 \rightarrow C_6H_2^+ + C_2H_2$ | $5.0 \times 10^{-10}$ |
| $C_2H_2^+ + C_2H_3 \rightarrow C_4H_3^+ + H_2$ | $3.3 \times 10^{-10}$ |
| $C_4H_2^+ + C_2H_3 \rightarrow C_6H_4^+ + H$ | $1.2 \times 10^{-9}$ |
| $C_4H_3^+ + C_2H_3 \rightarrow C_6H_4^+ + H_2$ | $5.0 \times 10^{-10}$ |
| $C_2H_3^+ + H \rightarrow C_2H_2^+ + H_2$ | $6.8 \times 10^{-11}$ |
| $C_2H_3^+ + C_4H \rightarrow C_6H_2^+ + H_2$ | $4.0 \times 10^{-10}$ |
| $C_2H_3^+ + C_6H \rightarrow C_6H_2^+ + C_2H_2$ | $5.0 \times 10^{-10}$ |

$$\partial_\tau(n_{iA}) = \underbrace{\left\{\beta_A n_A + \left(\sum_{\ell AB} k_\ell n_A n_{iB} - \sum_{qBA} k_q n_B n_{iA}\right) + \left(\frac{P}{E_{\mathrm{CH}}^{\mathrm{diss}}\mathrm{vol}}\right)\right\}}_{\text{First Part (Gain)}} - \underbrace{\left\{\alpha_A n_e n_{iA} + \sum_{p=1}^{N} \frac{n_{\mathrm{CNT}}}{\lambda_s}\left(\overset{\mathrm{tip}^p}{Q}_{iA} + \overset{\mathrm{cyl}^p}{Q}_{iA}\right) + \mathrm{K}_{\mathrm{wall}}^{iA} n_{iA}\right\}}_{\text{Second Part (Loss)}} \tag{10}$$

$$\partial_\tau(n_{iB}) = \underbrace{\left\{\beta_B n_B + \left(\sum_{qBA} k_q n_B n_{iA} - \sum_{\ell AB} k_\ell n_A n_{iB}\right) + \left(\frac{P}{E_{\mathrm{H}}^{\mathrm{diss}}\mathrm{vol}}\right)\right\}}_{\text{First Part (Gain)}} - \underbrace{\left\{\alpha_B n_e n_{iB} + \sum_{p=1}^{N} \frac{n_{\mathrm{CNT}}}{\lambda_s}\left(\overset{\mathrm{tip}^p}{Q}_{iB} + \overset{\mathrm{cyl}^p}{Q}_{iB}\right) + \mathrm{K}_{\mathrm{wall}}^{iB} n_{iB}\right\}}_{\text{Second Part (Loss)}} \tag{11}$$

Equations (10) and (11) describe the evolution of positive ion density $j = (A, B)$ type in bulk plasma as an outcome of various processes. The gain in ion density represented in the first part of the equations is due to dissociative ionization of neutrals, the influence of ion-neutral reactions considered in the study (listed in Table 3), and the plasma power ion dissociation process, respectively. The loss in ion number density represented in the second part of the equations is due to ion–electron recombination, ion accretion on the nanotube's surface, and reactive chamber wall discharge loss of ions, respectively. The specifics of the equations mentioned above are given in Table 5.

**Table 4** Particulars of functions/parameters termed in Eqs. (8–9)

| Parameters/functions | Description |
|---|---|
| $V_j^{\mathrm{IF}} = \left(\frac{4.4\times10^{17}}{\mathrm{vol}} J_j\right)$ | Inflow of neutrals from the reactive plasma chamber [37] |
| $V_j^{\mathrm{OF}} = \left(\frac{\upsilon_{\mathrm{pump}}}{\mathrm{vol}} n_j\right)$ | Outflow of neutrals from the reactive plasma chamber where $\upsilon_{\mathrm{pump}}$ is the pumping rate [37] |
| $J_{\mathrm{thd}}$ | Thermal dehydrogenation flux [34] |
| $\overset{\mathrm{adsp}}{J}_j, \overset{\mathrm{desp}}{J}_j$ | Adsorption and desorption fluxes |
| $K_{\mathrm{wall}}^j = \sigma_j \frac{D_j}{\Lambda^2}$ | Neutral atom discharge loss [38] |
| $\Lambda = RL/\left\{(2.405L)^2+(\pi R)^2\right\}^{1/2}$ | Cylindrical chamber effective diffusion length [38] |
| $\sigma_j,\ D_j$ | Wall reaction and diffusion coefficient of respective species [38] |

**Table 5** Particulars of functions/parameters termed in Eqs. (10–11)

| Parameters/functions | Description |
|---|---|
| $K_{\mathrm{wall}}^{ij} = \left(\frac{\gamma_{ij} u_{ij} A_{chm}}{4\,\mathrm{vol}}\right)$ | Ion discharge loss to the wall [37] |
| $k_q,\ k_l$ | Neutral–ion reaction rate coefficients [39] |
| $\delta E_{\mathrm{CH}}^{\mathrm{diss}}(\approx 2.1\ \mathrm{eV})$, $\delta E_{\mathrm{H}}^{\mathrm{diss}}(\approx 2.1\ \mathrm{eV})$ | Hydrocarbon and hydrogen dissociation energies [33] |
| $P$ | Plasma power |

**Table 6** Particulars of functions/parameters termed in Eq. (12)

| Parameters/functions | Description |
|---|---|
| $\beta_j$ | Ionization coefficient of $j = (A, B)$ type neutral atom |
| $\alpha_j(T_e) = \alpha_{j0}\left(300/T_e\right)^k$ | Recombination coefficient for ions and electrons where is $k = -1.2$ constant [40] |
| $n_e,\ n_j,\ n_{ij}$ | Electron, neutral, and ion density, respectively |
| $n_{CNT}$ | Number density of CNT |
| $K_{\mathrm{wall}}^e = \left(\frac{\gamma_e u_e A_{chm}}{4\,\mathrm{vol}}\right)$ | Loss rate of electrons over the volume on the discharge wall [37] |
| $A_{chm},\ vol$ | Cylindrical plasma chamber's area and volume (length $L = 23$ cm and internal diameter $D = 32$ cm) [37] |

$$\partial_\tau(n_e) = \underbrace{\left\{\sum_{A=1}^{10}\beta_A n_A + \sum_{B=1}^{2}\beta_B n_B\right\}}_{\text{First Part (Gain)}} - \underbrace{\left\{\left[\sum_{A=1}^{10}\alpha_A n_{iA} n_e + \sum_{B=1}^{2}\alpha_B n_{iB} n_e\right] + \sum_{p=1}^{N}\frac{n_{\mathrm{CNT}}}{\lambda_s}\gamma_e\left(\overset{\mathrm{tip}^p}{Q}_e + \overset{\mathrm{cyl}^p}{Q}_e\right) + \mathrm{K}_{\mathrm{wall}}^e n_e\right\}}_{\text{Second Part (Loss)}} \tag{12}$$

Equation (12) describes the evolution of electron density in the bulk plasma attributed to various processes. The increase in electron density represented in the first term of the equation is attributed to the dissociative ionization of neutral. The second part of the equation represents the loss in electron density owing to ion–electron recombination, electron accretion on the nanotube surface, and reactive chamber wall discharge loss of electrons, respectively. The specifics of the equation mentioned above are given in Table 6.

In Eqs. (8–12), the surface collection currents of ions, electrons and neutrals define the amount of charge collected by the nanotube's tip and cylindrical surfaces, which are given by- $\overset{\mathrm{tip}^p}{Q}_j = n_j(z)\pi d_p^2\left(\frac{8k_{\mathrm{B}}T_j}{\pi m_j}\right)^{1/2}$, $\overset{\mathrm{cyl}^p}{Q}_j = n_j(z)\pi d_p l\left(\frac{2\pi k_{\mathrm{B}}T_j}{m_j}\right)^{1/2}$, $\overset{\mathrm{tip}^p}{Q}_e = n_e(z)\,\pi d_p^2\left(\frac{8k_{\mathrm{B}}T_e}{\pi m_e}\right)^{1/2}\exp\left[Z\alpha_e + \frac{eU_S}{k_{\mathrm{B}}T_S}\right]$, $\overset{\mathrm{cyl}^p}{Q}_e = n_e(z)d_p l\left(\frac{2\pi k_{\mathrm{B}}T_S}{m_e}\right)^{1/2}\exp\left[\frac{eV_S}{k_{\mathrm{B}}T_e} + \frac{eU_S}{k_{\mathrm{B}}T_S}\right]$, $\overset{\mathrm{tip}^p}{Q}_{ij} = n_{ij}(z)\pi d_p^2\left(\frac{8k_{\mathrm{B}}T_{ij}}{\pi m_{ij}}\right)^{1/2}\left[1 - Z\alpha_{ij}\right]\exp\left[\frac{-E_b}{k_{\mathrm{B}}T_S}\right]\exp\left[\frac{-eU_S}{k_{\mathrm{B}}T_S}\right]$,

**Table 7** Particulars of functions/parameters in equations for collection currents

| Parameters/functions | Description |
|---|---|
| $V_S$ | CNT tubular surface potential |
| $E_b$ | Energy necessary for bulk diffusion |
| $n_{ij}(z) = n_{ij0}\left(1 - \frac{2e\varphi(z)}{m_{ij}u_{ij0}^2}\right)^{-1/2}$ | Plasma sheath ion density |
| $n_e(z) = n_{e0}\exp\left(\frac{\lvert e\rvert\varphi(z)}{k_B T_e}\right)$ | Plasma sheath electron density |
| $n_{e0}, \quad n_{ij0}$ | Primary number density of electrons and ions |
| $\varphi(z) = \varphi_0\exp\left(-\frac{z}{\lambda_D}\right)$ | Electrostatic sheath potential |
| $\lambda_D = \sqrt{k_B T_e/n_e e^2}$ | Debye length |

$\overset{\text{cyl}^p}{\underset{ij}{Q}} = n_{ij}(z)d_p l\left(\frac{2\pi k_B T_{ij}}{m_{ij}}\right)^{1/2}\exp\left[-\frac{E_b}{k_B T_S}\right]\exp\left[-\frac{eU_S}{k_B T_S}\right]\times\left\{\frac{2}{\sqrt{\pi}}\left(\frac{eV_S}{k_B T_{ij}}\right)^{1/2} + \exp\left[\frac{eV_S}{k_B T_{ij}}\right]\text{erfc}\left[\left(\frac{eV_S}{k_B T_{ij}}\right)^{1/2}\right]\right\}$. The other terms are given in Table 7.

2.3 Charge acquired by the carbon nanotubes of the array –

$$\partial_\tau(Z^p) = \sum_j^{A,B}\left(\overset{\text{tip}^p}{\underset{ij}{Q}} + \overset{\text{cyl}^p}{\underset{ij}{Q}}\right) - \gamma_e\left(\overset{\text{tip}^p}{\underset{e}{Q}} + \overset{\text{cyl}^p}{\underset{e}{Q}}\right) \tag{13}$$

Equation (13) evaluates the charge acquired by the $p$th CNT's tip and cylindrical surfaces due to the accretion of $j = (A, B)$ type ions and electrons impending from the plasma. $Z^p$ describes the aggregate charge on the $p$th CNT on the array.

$$\mathbb{Q} = \frac{\sum_{p=1}^N \partial_\tau(Z^p)}{N} \tag{14}$$

Equation (14) estimates the average charge acquired by CNTs of an array growing in a reactive plasma surrounding. The equation considers the summation of surface charges over all $N$ numbers of CNTs.

2.4 Growth of CNT array in plasma

$$\partial_\tau\left[\pi\left\{\left(\frac{D_{\text{ct}}}{2}\right)^2 - \left(\frac{d_p}{2}\right)^2\right\}l_p\right] = \left[\left\{D_{\text{surf}} + D_{\text{volm}} + 2\kappa\upsilon\left(\frac{D_E}{2}\right)^2_p\right\}\times\frac{P\left\{8\pi\left(\frac{D_{\text{ct}}}{2}\right)_p\right\}n_C}{\upsilon\,\rho_{\text{CNT}}\sum_{A=1}^{8}\overset{p}{\underset{iA}{Q}}}\right.$$
$$\left. + \left\{\frac{D_{\text{sd}}}{4\pi\left(\frac{D_{\text{ct}}}{2}\right)^2_p} + \left\{4\pi\left(\frac{D_{\text{ct}}}{2}\right)^2_p\right\}\sum_{B=1}^{2}\overset{p}{\underset{iB}{Q}}\right\}n_H\right]\times\frac{m_{\text{ct}}}{\rho_{\text{ct}}}. \tag{15}$$

Equation (15) calculates the growth of $p$th CNT growing over $p$th catalyst nanoparticle where $l_p$ and $d_p$ are the length and diameter of $p$th CNT. This growth rate is calculated by considering the total volumetric change (reliant on CNT diameter and height) during the continuous CNT growth on the $p$th catalyst nanoparticle. The former part of the equation assesses the carbon surface and volume diffusion process over the nanoparticle surface and volume along with the precipitation process of carbon atoms from the catalyst surface, giving rise to graphitic cylindrical tubes. When volume diffusion dominates over surface diffusion, MWCNTs are produced otherwise SWCNTs; hence, the supremacy of one diffusion process over the other administers the type of CNTs being produced. The former part of the equation considers the metal nanoparticle self-diffusion process and elimination of growth hindering amorphous layer formed over the catalyst due to its contamination. The removal of this layer promotes the CNT growth further by getting rid of carbon atoms attached to the extremities of growing CNT or on the nanoparticle's surface. The vertically aligned structure of CNT is attained due to a continuous electrostatic force, resulting from the plasma sheath electric field [23]. Presuming the CNTs to have a perfectly aligned and concentric cylindrical structure, the difference in CNT diameters can estimate the number of CNT layers created, i.e. outer CNT diameter ($\sim D_{ct}$) − inner CNT diameter ($d$) ≈ number of CNT layers.

$$\gamma_{\text{array}}(d, l) = \frac{\sum_{p=1}^N \partial_\tau\left[\pi\left\{\left(\frac{D_{\text{ct}}}{2}\right)^2 - \left(\frac{d_p}{2}\right)^2\right\}l_p\right]}{N} \tag{16}$$

**Table 8** Particulars of functions/parameters termed in Eq. (15)

| Parameters/functions | Description |
|---|---|
| $D_{\text{surf}} = a_0^2 \upsilon \exp\left(-\frac{E_{\text{surf}}}{k_B T_S}\right)$ | Surface diffusion coefficient of C atoms [34] |
| $D_{\text{volm}} = \frac{\upsilon D_{ct}}{2\pi} \exp\left(-\frac{E_{\text{volm}}}{k_B T_S}\right)$ | Volume diffusion coefficient of C atoms [34] |
| $D_{\text{sd}} = D_{\text{sd0}} \exp\left(-\frac{\delta E_{\text{sd}}}{k_B T_S}\right)$ | Self-diffusion coefficient for metal atoms [34] |
| $\kappa = a_0 \upsilon \exp\left(-\frac{\delta E_{\text{inc}}}{k_B T_S}\right)$ | C-atom incorporation rate into the CNT wall [34, 41] |
| $E_{\text{surf}} (= 0.3 \text{ eV})$ | Surface diffusion C-atom energy barrier on the catalyst [34] |
| $E_{\text{volm}} (= 1.6 \text{ eV})$ | Volume diffusion C-atom energy barrier through the catalyst [34] |
| $\delta E_{\text{inc}} (= \delta E_d = 0.8 \text{ eV})$ | C-atom energy barrier to diffuse along nanotube–catalyst boundary [41] |

Hence, Eq. (16) estimates the average growth rate of CNTs in an array growing over catalyst nanoparticles in a reactive plasma where $l$ and $d$ are the average length and diameter of CNTs in an array. The equation considers the summation of the growth rate of all $N$ numbers of CNTs. The details of various growth processes deliberated in Eq. (15) are given in Table 8.

### 2.5 Amorphous carbon film deposition

Along with VACNT array growth, an amorphous carbon film can also get deposited between the spaces of the nanotubes of the array. The ionic and neutral species get incorporated at the level where the catalysts are anchored on the substrate ($z = \lambda_s$). The deposition flux rate of carbon film can be determined as [42–44]:

$$J_{cf} = J_{iA} + J_{iA}\theta_t y_s(\lambda_s) + n_C \upsilon \exp\left(\frac{-\delta E_f}{k_B T_S}\right)(\lambda_s) - \frac{1}{4}(1-\theta_t)\upsilon_0 \sigma_{\text{bulk}} J_H - J_{iA}(1-\theta_t) y_{\text{sp}} \tag{17}$$

The first term of Eq. (17) describes the direct absorption flux of hydrocarbon ions, the second term determines the neutral stitching owing to ion-induced incorporation with the stitching rate as $y_s = 2.49 \times 10^{-2} + 3.29 \times 10^{-2}(E_i)$, the third term is the incorporation of carbon atoms into the film, the fourth term expresses the atomic hydrogen etching assuming the concentration of carbon is higher than hydrogen. Lastly, the fifth term defines the sputtering loss of carbon atoms with sputtering yield as $y_{sp} = -3.89 \times 10^{-7}\left(E_i^2\right) + 7.04 \times 10^{-4}(E_i - 100) + 8.14 \times 10^{-2}$, where $E_i = 2.1$ eV is the ion energy. Other parameters in the equation are $\delta E_f = 2.3$ eV, and $\sigma_{\text{bulk}} = 6.8 \times 10^{-15}$ cm$^2$. The growth rate of this carbon film in terms of its width $w_{cf}$ is expressed in Eq. (18):

$$\partial_\tau \{w_{cf}\} = \frac{J_{cf} M_{cf}}{\rho N_A}, \tag{18}$$

where $J_{cf}$ is the deposition flux rate, $M_{cf}$ is the film mole mass, $\rho$ is the film density, $N_A$ is the Avogadro number.

### 2.6 Field emission from the CNT array grown in plasma

Carbon nanotubes are considered pre-eminent electron field mission materials attributed to their high aspect ratio, i.e. length-to-radius ratio. The field enhancement factor is crucial to understanding the field emission aspects. When an external electric field of adequate strength is applied around the tip of the nanotube, it leads to electron emission. This factor is usually measured higher for individual CNTs around 30,000–50,000 than for CNT films around 1000–3000, which reduces furthermore for dense films due to the electric field screening from the close nanotubes. The factor is known to be dominated by the length of nanotubes and spacing between them; hence, it is essential to recognize their effects to achieve optimum field emission properties [45].

$$\beta_0 = 1.2\left(2.15 + \frac{2l}{d}\right)^{0.9} \tag{19}$$

Equation (19) approximates the primary geometrical field enhancement factor for an individual VACNT where $l$ and $d$ signify the dimensions of VACNT [46]. The field emission factor is also defined as the ratio of the CNT tip electric field to the macroscopic electric field. Considering the spacing $s$ between VACNTs, the geometrical field enhancement factor for CNT films [47] can be estimated by Eq. (20).

$$\beta = \beta_0\left[1 - \exp\left(-2.3172\frac{s}{l}\right)\right] \tag{20}$$

## 3 Results and discussion

The present computation would help to accomplish a catalysed CNT array growth in $C_2H_2/H_2$ reactive plasma. The effect of substrate temperature, acetylene concentration, plasma power and plasma pressure on the CNT array aspects is examined. The

**Table 9** Computational parameters considered in the model

| Terms | Initial value |
|---|---|
| Computational substrate area | $5000 \times 5000\ \mu m^2$ |
| Average catalyst diameter $D_{ct}$ | 40 nm |
| Minimum spacing between catalysts | 100 nm |
| $T_e$ (electron temperature) | 2.0 eV |
| $T_n$ (neutral temperature) | 0.20 eV |
| $T_{ij}$ (ion species temperature) | 0.20 eV |
| $T_s = T_{ct}$ (substrate and catalyst temperature) | 750 K |
| Plasma power | 300 W |
| Working pressure | 100 mTorr |
| $U_S$ (substrate voltage) | − 400 V |
| $\theta_t$ (total coverage) | 0.01 |
| Acetylene gas flow | 50 sccm |
| Hydrogen gas flow | 200 *sccm* |
| $\gamma_e$, $\gamma_{ij}$, $\gamma_j$ (sticking coefficients) | 1 |

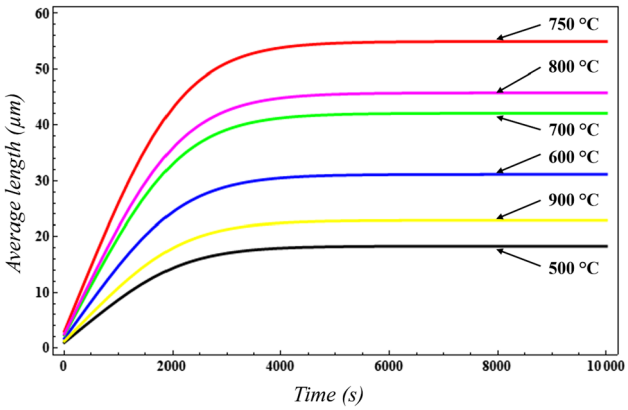



**Fig. 3** Time variation of average length of CNTs of the array during plasma-assisted growth for different substrate temperatures

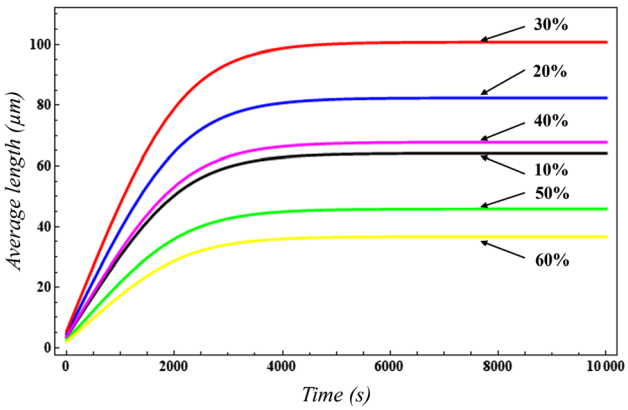



**Fig. 4** Time variation of average length of CNTs of the array during plasma-assisted growth for different $C_2H_2$%

consequence of gas ratio and substrate temperature on the width of amorphous carbon film between CNTs is also studied. The influence of CNT array aspects on the field enhancement factor of CNT films is also deliberated. The equations mentioned in Sect. 2 are first-order time-varying equations that are solved for plasma-aided CNT array growth considering the experimentally acquired primary conditions and glow-discharge parameters anticipated from the literature [21, 48, 49] and others mentioned in Table 9.

The time variation of average CNT length at different substrate temperatures is depicted in Fig. 3. The CNT array growth proceeds via hydrocarbon and hydrogen dissociation, creation of diffusion through or over the catalyst nanoparticle, generation of amorphous carbon layer and termination of growth as catalyst gets encapsulated with carbon. As the substrate temperature increases, the ion dissociation increases; hence, the incoming flux of carbon-generating species increases. Therefore, more carbon atoms diffuse into and over the catalyst, which correspondingly increases the average length of the CNTs in an array. As the temperature is increased beyond an optimum value, i.e. 750 °C [c.f. Fig. 3], the accretion of carbon atoms increases excessively over the catalyst consuming

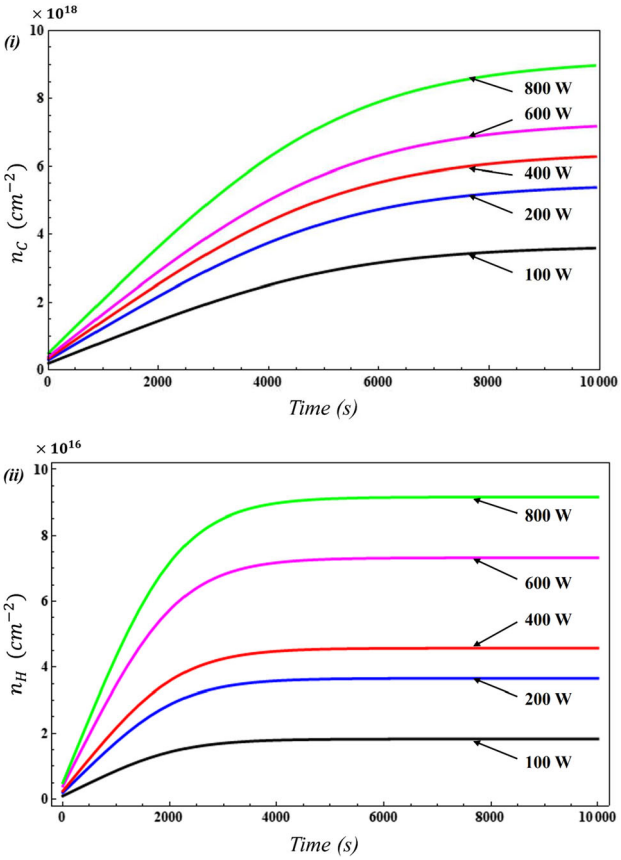



**Fig. 5** Time variations of (i) carbon density and (ii) hydrogen density during plasma-assisted growth of CNT array for different plasma powers

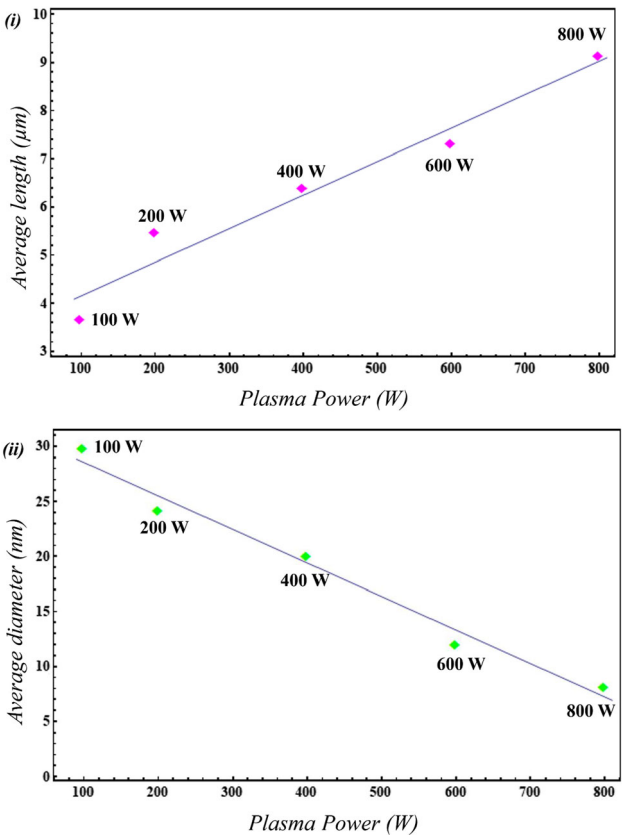



**Fig. 6** Variation of (i) average length and (ii) average diameter of CNTs of the array for different plasma powers. (Line represents best-fit curve)

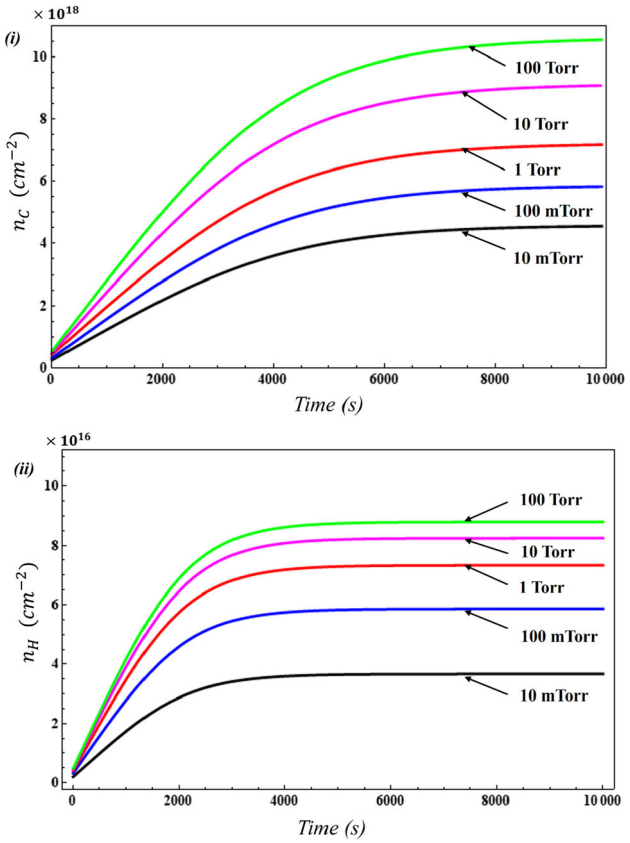



**Fig. 7** Time variations of (i) carbon density and (ii) hydrogen density during plasma-assisted growth of CNT array for different plasma pressures

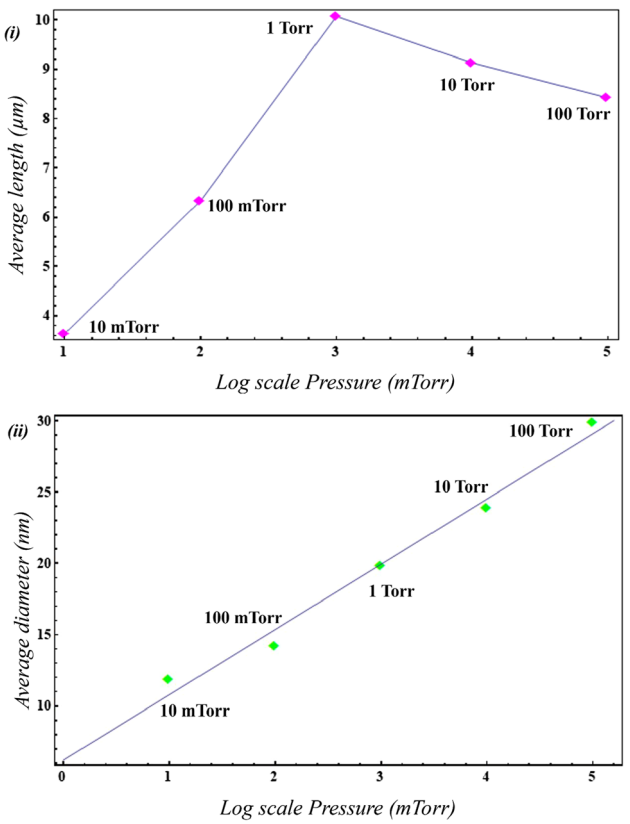



**Fig. 8** Variation of (i) average length and (ii) average diameter of CNTs of the array for different plasma pressures. (Line represents best-fit curve)

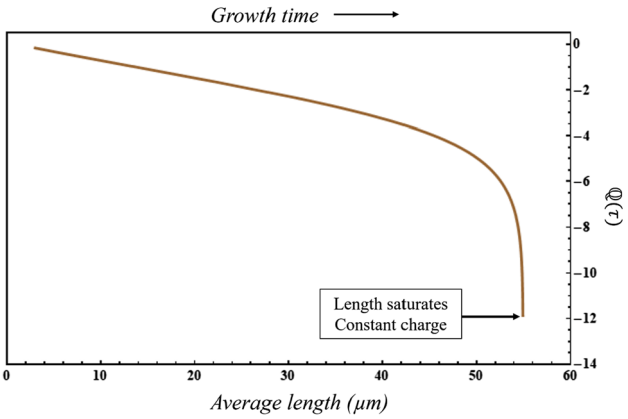



**Fig. 9** A plot of average length and average charge of CNT with time

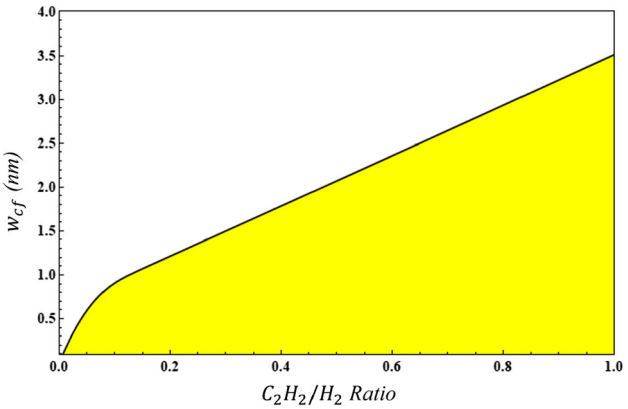



**Fig. 10** Width of carbon film formed between CNTs with changing gas ratio

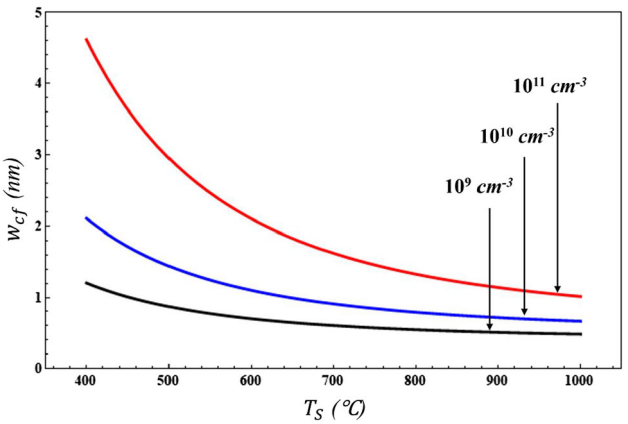



**Fig. 11** Width of carbon film formed between CNTs for changing substrate temperature for different hydrocarbon ion densities

rapidly its active area required for the growth. This accelerates the generation of the amorphous carbon layer and terminates its growth sooner. Hence, the average length of the array at a higher temperature is inferior. This theoretical result is consistent experimentally with the observation of Chhowalla et al. [9], Cui et al. [50], and Bulyarskiy et al. [51, 52].

The time variation of average CNT length at different acetylene concentrations is depicted in Fig. 4. The presence of $H_2$ in $C_2H_2/H_2$ plasma has two roles in the growth of CNTs: the generation of atomic hydrogen and the removal of excess carbon and delay in creating amorphous carbon on the catalyst surface. This means that the etching rate of the amorphous layer decreases for lower hydrogen content. On the lower side of the increase in acetylene concentration in plasma, the carbon-creating species essential for growth and hydrogen-creating species required for excess carbon etching collectively give the initial increase in the array length. However, increasing the $C_2H_2$ concentration over an ideal value, i.e. 30% [c.f. Fig. 4], results in a plasma with excess carbon for growth and low hydrogen for etching. Hence, the growth rate dominates the etching rate giving rise to quick termination of growth and shorter nanotube arrays. The result is verified experimentally by Chhowalla et al. [9], Bell et al. [21], and Chen et al. [53].

When the plasma power increases, the plasma electron density and electron temperature also increase [54], which gives rise to plasma neutral atom ionization and an increase in plasma ion number density. Bell et al. [21] studied that the decomposition of hydrocarbon rises on increasing power. The increase in neutral decomposition enhances the ion-induced dissociation, which generates

**Table 10** Dimensions of the array obtained from the pressure study [cf Fig. 8(i, ii)] which considered for field enhancement study

| Case | $l$ (μm) | $d$ (nm) |
|---|---|---|
| I | 3.64 | 11.88 |
| II | 6.34 | 14.19 |
| III | 10.07 | 19.81 |
| IV | 9.13 | 23.91 |
| V | 8.42 | 29.92 |

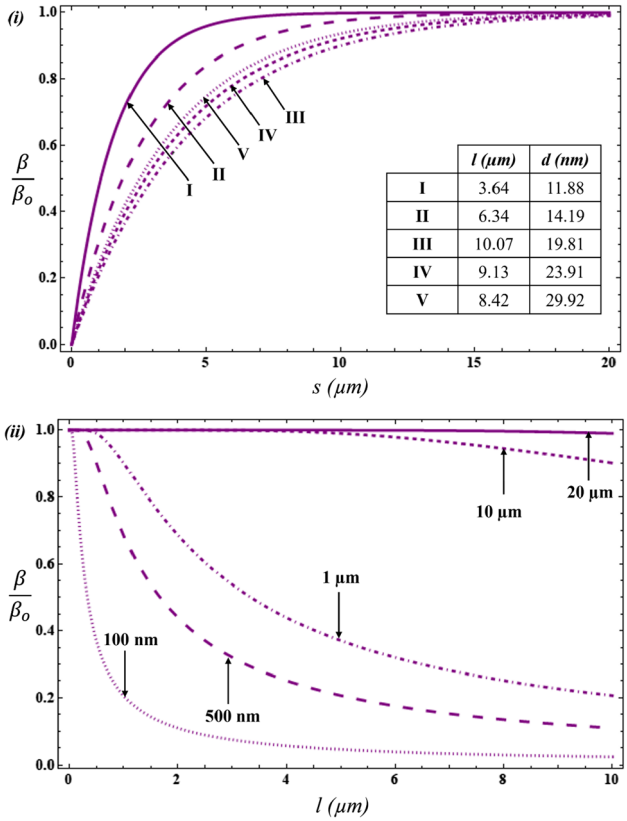



**Fig. 12** $\frac{\beta}{\beta_0}$ variation for an array (i) with spacing $s$ between CNTs and (ii) with length for different spacing between CNTs

a significant amount of carbon and hydrogen species. The time variations of carbon and hydrogen number density at different plasma powers are depicted in Fig. 5(i) and (ii), respectively. Since carbon atoms are responsible for the growth of nanotubes and hydrogen atoms' etching, both the rates collectively heighten the nanotubes' array. The decrease in the nanotube diameter is due to stresses exerted by the graphitic walls and etching due to hydrogen species. For increasing plasma power, the hydrogen etching leads to the removal of graphitic walls of nanotubes, thereby diminishing its diameter. (The initial diameter was comparable to the catalyst diameter.) Thus, for higher power, the reduction in diameter is more significant. The variations of average length and average diameter of CNTs at different plasma power are depicted in Fig. 6(i) and (ii), respectively, where the points represent the value of average length and diameter of CNTs of the array for specific plasma power and the line represents the best fit. The outcomes are comparable to the experimental observations of Ikuno et al. [55], Wei et al. [56], Merkulov et al. [57], and Chang et al. [58].

When the plasma pressure increases, the plasma electron density and electron temperature decrease [59], increasing the number of electrons' and neutrals' collisions. This indicates that plasma species would lose energy, decreasing the temperature of neutral species and resulting in lower neutral ionization, i.e. the number of neutrals in plasma (neutral species density) is higher at higher pressures. Hence, at higher pressures, the neutral flux to the catalyst increases, thus raising the amount of C and H atoms generated over the catalyst surface. The time variations of carbon and hydrogen number density with plasma pressure are depicted in Fig. 7(i) and (ii), respectively. The increases in the density of carbon and hydrogen atoms lead to more significant nanotube growth; hence, the average length of CNTs of the array increases. At extremely high pressures, hydrogen neutrals dissociation is higher than hydrocarbons as they require lower energy to dissociate. This increases the number of carbon atoms excessively as opposed to hydrogen. Hence, deposition rate dominates etching rate, which rapidly consumes the catalyst active area required for growth, leading to shorter nanotubes. The fall in average CNT diameter is the greatest for the lowest plasma pressure. The etching rate is lower for a high plasma pressure than the deposition rate; hence, the diameter reduction is slight. (The initial diameter was comparable to the catalyst

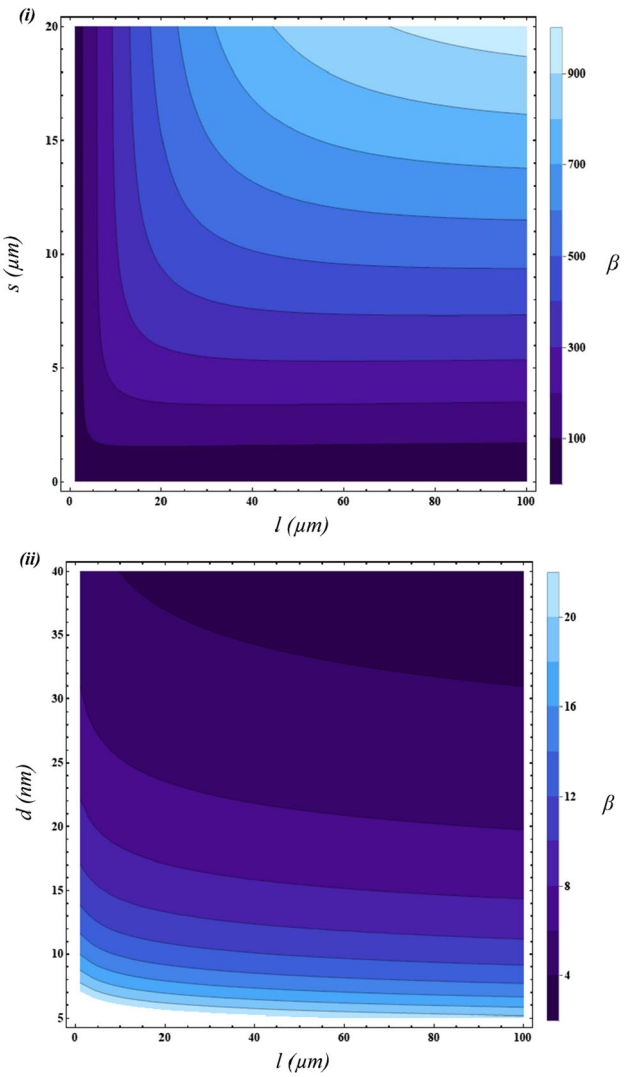



**Fig. 13** Contour plots for geometrical field enhancement factor of an array for varying (i) average CNT spacing and length, (ii) average CNT diameter and length

diameter.) The variations of average length and average diameter of CNTs at different plasma pressures are depicted in Fig. 8(i) and (ii), respectively, where the points represent the value of average length and diameter of CNTs of the array for specific plasma pressure and the line represents the best fit. The results for the initial increase in length for low pressure and decrease with higher pressures are experimentally detected by Cruden et al. [60], Tsakadze et al. [61], and Li et al. [62], whereas the increase in diameter with pressure is in agreement with Li et al. [62] and Tanemura et al. [7].

A plot of the instantaneous change in length and corresponding charge attained by the CNT with time is depicted in Fig. 9. As the substrate where the catalyst is anchored and array growth occur is attached to a negatively biased substrate, the CNTs of the array acquire a negative charge. With time, the average length of the CNTs of the array increases and attains a saturated value on termination [cf. Figs. 3, 4]. At shorter CNT length, negative charge is less, and at larger length, this negative charge is more. Hence, the negative charge acquired by the growing CNTs would also increase due to increase in CNT length and lastly attain a constant value on growth termination ($l = 55$ μm). The variation of the nanotube with growth time is experimentally verified through various studies [9, 56].

The variation of the width or thickness of amorphous carbon film generated over the substrate between CNTs with $C_2H_2/H_2$ gas ratio is depicted in Fig. 10. As the amount of acetylene compared to hydrogen in plasma increases, i.e. the $C_2H_2/H_2$ ratio increases, more hydrocarbon species are available for dissociation and incorporation, which correspondingly increases carbon atoms density. This increases the flux of carbon atoms and creates an amorphous carbon film. In another way, if the amount of hydrogen in plasma increases, i.e. the $C_2H_2/H_2$ ratio decreases, the film etching increases due to an increase in hydrogen flux that produces a thinner amorphous carbon film. This result is comparable to the result numerically modelled by Denysenko et al. [44].

The variation of the width of amorphous carbon film generated over the substrate between CNTs with substrate temperature for different hydrocarbon ion densities is depicted in Fig. 11. As the substrate temperature increases, the width of carbon film decreases due to an increase in desorption flux, which decreases the total surface coverage. The deposition flux rate of carbon film decreases

significantly, depositing a thin carbon film. Also, increasing the hydrocarbon ion density increases the ion flux increases, which collectively increases their direct and ion-induced incorporation to the carbon film [cf Eq. (17)], and the film gets thicker. This outcome is observed experimentally by Mantzaris et al. [42] and simulated numerically by Denysenko et al. [41].

The variation of the ratio of geometrical field enhancement factor $\beta$ of a CNT film and enhancement factor $\beta_0$ of an individual CNT (i.e. the ratio of field enhancement of an array film and an individual nanotube $\beta/\beta_0$) with spacing $s$ for different arrays obtained from the pressure study mentioned in Table 10 is depicted in Fig. 12(i). The field emission property is significant for arrays with the highest $\beta$; therefore, arrays with the highest $\beta/\beta_0$ ratio would have the best field emission characteristics. As seen from the figure, for less dense arrays, i.e. for greater spacing $s$ between CNTs, $\beta/\beta_0$ increases. This can be attributed to the reduction in screening effects of individual CNTs of the array on one another. This suggests that the field emission from sparse nanotube arrays is better than denser ones. For case I, where $l$ and $d$ possess minimum values, $\beta/\beta_0$ is maximum signifying that small and thin CNT arrays are appropriate for best emission characteristics. Also, the result shows that as $s$ increases beyond 2 $l$, the exponential term in Eq. (20) becomes insignificant, and the CNTs of the array are spaced so far apart that they behave as individual emitters showing the best emission characteristics.

Figure 12(ii) shows the variation of $\beta/\beta_0$ with the length of the nanotubes of the array for different $s$. For constant $d = 40$ nm, an increase in $l$ decreases $\beta/\beta_0$, indicating poor field emission characteristics. The results show that as s increases, $\beta/\beta_0$ also increases, which substantiates our previous result. However, again it is observed that as $s$ increases beyond 2 $l$, i.e. the array becomes sparse, CNTs start behaving as individual emitters. Hence, taking advantage of these results, one can design a CNT array by varying the plasma process parameters as reflected in this study to attain the best field emission characteristics.

The contour plots for field enhancement factor on the variation of inter CNT spacing and length (constant $d = 40$ nm) are depicted in Fig. 13 i), and the average CNT diameter and length (constant $s = 100$ nm) are depicted in Fig. 13(ii), respectively. The plots verify that an array with short and thin CNTs with large spacing shows the highest field enhancement. Hence, it is evident that short and sparse CNT arrays will be appropriate for producing the best field emitters. These results are in agreement with the field emission studies of CNT arrays by Bonard et al. [47], Jo et al. [63], Xu et al. [64], and Chhowalla et al. [65].

## 4 Conclusions

The current computational study is modelled to investigate the growth kinetics of a carbon nanotube array in a plasma. A pre-patterned film of catalyst nanoparticles was considered, and a model was developed that computes the influence of substrate temperature, plasma concentration, power and pressure on the aspects of the CNT array. From the outcomes, it is concluded that plasma processing conditions greatly influence the array's growth. The average CNT length increases with substrate temperature and decreases beyond an optimum value. Similarly, the plasma concentration can also be adjusted to accomplish appropriate CNT length. The effect of power and pressures on the carbon and hydrogen atom density that corresponds to deposition of carbon atoms and etching by hydrogen is also observed. The change in average length and diameter of CNTs of the array with power and pressures is revealed. The temporal change in average charge acquired by the nanotube and its length is also represented. The effect of gas ratio and substrate temperature on the width of a carbon film deposited on the substrate between CNTs is also presented. The field enhancement of the nanotube array is studied, and ways to improve the emission properties are discussed. The current study can support optimizing various plasma parameters as discussed in the script for achieving a suitable CNT array that can be used as the best field emitters. This model can also be comprehensively used to design a study of the plasma-assisted growth of several nanostructured materials and optimization of their properties.

**Acknowledgements** Umang Sharma is grateful to the Department of Science and Technology (DST), Government of India, for providing the financial support under DST – INSPIRE fellowship.

**Data Availability Statement** This manuscript has associated data in a data repository. [Authors' comment: The data supporting the findings of this study are available within the article and available from the corresponding author upon reasonable request.]

**Declarations**

**Conflict of Interest** The authors have no conflicts to disclose.

## References

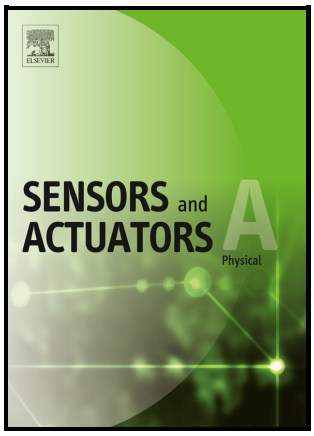
1. Z. Shi, Y. Lian, X. Zhou, Z. Gu, Y. Zhang, S. Iijima, L. Zhou, K. Yue, S. Zhang, Carbon N. Y. **37**, 1449 (1999)
2. A. Maiti, C.J. Brabec, C.M. Roland, J. Bernholc, Phys. Rev. Lett. **73**, 2468 (1994)
3. A.A. Puretzky, D.B. Geohegan, H. Schittenhelm, X. Fan, M.A. Guillorn, Appl. Surf. Sci. **197–198**, 552 (2002)
4. R.L. Vander Wal, G.M. Berger, T.M. Ticich, Appl. Phys. A Mater. Sci. Process **77**, 885 (2003)

5. M. Jung, Diam. Relat. Mater. **10**, 1235 (2001)
6. R. Brukh, S. Mitra, Chem. Phys. Lett. **424**, 126 (2006)
7. M. Tanemura, K. Iwata, K. Takahashi, Y. Fujimoto, F. Okuyama, H. Sugie, V. Filip, J. Appl. Phys. **90**, 1529 (2001)
8. C. Bower, O. Zhou, W. Zhu, D.J. Werder, S. Jin, Appl. Phys. Lett. **77**, 2767 (2000)
9. M. Chhowalla, K.B.K. Teo, C. Ducati, N.L. Rupesinghe, G.A.J. Amaratunga, A.C. Ferrari, D. Roy, J. Robertson, W.I. Milne, J. Appl. Phys. **90**, 5308 (2001)
10. L. Delzeit, C.V. Nguyen, R.M. Stevens, J. Han, M. Meyyappan, Nanotechnology **13**, 280 (2002)
11. Y. Luo, X. Wang, M. He, X. Li, H. Chen, J. Nanomater. **2012**, 542582 (2012)
12. P.H. Lin, C.L. Sie, C.A. Chen, H.C. Chang, Y.T. Shih, H.Y. Chang, W.J. Su, K.Y. Lee, Nanoscale Res. Lett. **10**, 1 (2015)
13. X. Tang, H. Yue, L. Liu, J. Luo, X. Wu, R. Zheng, G. Cheng, A.C.S. Appl, Nano Mater. **3**, 7659 (2020)
14. L. Nilsson, O. Groening, C. Emmenegger, O. Kuettel, E. Schaller, L. Schlapbach, H. Kind, J.M. Bonard, K. Kern, Appl. Phys. Lett. **76**, 2071 (2000)
15. U. Sharma, S.C. Sharma, IEEE Trans. Plasma Sci. **50**, 888 (2022)
16. T. Kato, G.H. Jeong, T. Hirata, R. Hatakeyama, K. Tohji, K. Motomiya, Chem. Phys. Lett. **381**, 422 (2003)
17. S. Hofmann, C. Ducati, J. Robertson, B. Kleinsorge, Appl. Phys. Lett. **83**, 135 (2003)
18. S.K. Srivastava, V.D. Vankar, V. Kumar, Thin Solid Films **515**, 1552 (2006)
19. S.K. Srivastava, V.D. Vankar, D.V. Sridhar Rao, V. Kumar, Thin Solid Films **515**, 1851 (2006)
20. Y. Hayashi, T. Fukumura, K. Odani, T. Matsuba, R. Utsunomiya, Thin Solid Films **518**, 3506 (2010)
21. M.S. Bell, K.B.K. Teo, R.G. Lacerda, W.I. Milne, D.B. Hash, M. Meyyappan, Pure Appl. Chem. **78**, 1117 (2006)
22. M.S. Bell, R.G. Lacerda, K.B.K. Teo, N.L. Rupesinghe, G.A.J. Amaratunga, W.I. Milne, M. Chhowalla, Appl. Phys. Lett. **85**, 1137 (2004)
23. U. Sharma, S.C. Sharma, Phys. Plasmas **25**, 103509 (2018)
24. H. Okuyama, N. Iwata, H. Yamamoto, Mol. Cryst. Liq. Cryst. **472**, 209/[599] (2007)
25. T. Ikuno, S. Takahashi, K. Kamada, S. Ohkura, S.I. Honda, M. Katayama, T. Hirao, K. Oura, Surf. Rev. Lett. **10**, 611 (2003)
26. J. Lee, E. Oh, T. Kim, J.H. Sa, S.H. Lee, J. Park, D. Moon, I.S. Kang, M.J. Kim, S.M. Kim, K.H. Lee, Carbon N. Y. **93**, 217 (2015)
27. O.A. Louchev, Y. Sato, H. Kanda, Appl. Phys. Lett. **80**, 2752 (2002)
28. G. Zhong, T. Iwasaki, J. Robertson, H. Kawarada, J. Phys. Chem. B **111**, 1907 (2007)
29. A. Thapa, K.L. Jungjohann, X. Wang, W. Li, J. Mater. Sci. **55**, 2101 (2020)
30. T. Campo, S. Pinilla, S. Gálvez, J.M. Sanz, F. Márquez, C. Morant, Nanomaterials **9**, 571 (2019)
31. M. Morassutto, R.M. Tiggelaar, M.A. Smithers, J.G.E. Gardeniers, Mater. Today Commun. **7**, 89 (2016)
32. H. Mehdipour, K. Ostrikov, A.E. Rider, Z. Han, Plasma Process. Polym. **8**, 386 (2011)
33. Z. Marvi, S. Xu, G. Foroutan, K. Ostrikov, Phys. Plasmas **22**, 013504 (2015)
34. I. Denysenko, K. Ostrikov, J. Phys. D. Appl. Phys. **42**, 015208 (2009)
35. H. Mehdipour, K. Ostrikov, A.E. Rider, Nanotechnology **21**, 455605 (2010)
36. M. Mao, A. Bogaerts, J. Phys. D. Appl. Phys. **43**, 205201 (2010)
37. I.B. Denysenko, S. Xu, J.D. Long, P.P. Rutkevych, N.A. Azarenkov, K. Ostrikov, J. Appl. Phys. **95**, 2713 (2004)
38. M.A. Lieberman, A.J. Lichtenberg, *Principles of plasma discharges and materials processing*, 2nd edn. (Wiley, New York, 2005)
39. UMIST Database Astrochem. (2012)
40. M.S. Sodha, S. Misra, S.K. Mishra, S. Srivastava, J. Appl. Phys. **107**, 103307 (2010)
41. I. Denysenko, N.A. Azarenkov, J. Phys. D. Appl. Phys. **44**, 174031 (2011)
42. N.V. Mantzaris, E. Gogolides, A.G. Boudouvis, A. Rhallabi, G. Turban, J. Appl. Phys. **79**, 3718 (1996)
43. O.A. Louchev, C. Dussarrat, Y. Sato, J. Appl. Phys. **86**, 1736 (1999)
44. I. Denysenko, K. Ostrikov, M.Y. Yu, N.A. Azarenkov, J. Appl. Phys. **102**, 074308 (2007)
45. K.B.K. Teo, M. Chhowalla, G.A.J. Amaratunga, W.I. Milne, G. Pirio, P. Legagneux, F. Wyczisk, D. Pribat, D.G. Hasko, Appl. Phys. Lett. **80**, 2011 (2002)
46. C.J. Edgcombe, U. Valdrè, J. Microsc. **203**, 188 (2001)
47. J.M. Bonard, N. Weiss, H. Kind, T. Stöckli, L. Forró, K. Kern, A. Châtelain, Adv. Mater. **13**, 184 (2001)
48. J. Benedikt, J. Phys. D. Appl. Phys. **43**, 043001 (2010)
49. C. Bower, W. Zhu, S. Jin, O. Zhou, Appl. Phys. Lett. **77**, 830 (2000)
50. H. Cui, O. Zhou, B.R. Stoner, J. Appl. Phys. **88**, 6072 (2000)
51. S.V. Bulyarskiy, G.G. Gusarov, A.V. Lakalin, M.S. Molodenskiy, A.A. Pavlov, R.M. Ryazanov, Diam. Relat. Mater. **103**, 107665 (2020)
52. S.V. Bulyarskiy, A.V. Lakalin, M.S. Molodenskii, A.A. Pavlov, R.M. Ryazanov, Inorg. Mater. **57**, 20 (2021)
53. G.Y. Chen, C.H.P. Poa, V. Stolojan, S. Henley, S.R.P. Silva, Mater. Res. Soc. Symp. Proc. **858**, 91 (2004)
54. W.Z. Collison, T.Q. Ni, M.S. Barnes, J. Vac. Sci. Technol. A Vac. Surf. Film **16**, 100 (1998)
55. T. Ikuno, M. Katayama, K. Kamada, S. Hiwatashi, S. Ohkura, S.I. Honda, K. Oura, Jpn. Appl. Phys. Part 1 Regul. Pap. Short Notes Rev. Pap. **42**, 6717 (2003)
56. H.W. Wei, K.C. Leou, M.T. Wei, Y.Y. Lin, C.H. Tsai, J. Appl. Phys. **98**, 044313 (2005)
57. V.I. Merkulov, D.K. Hensley, A.V. Melechko, M.A. Guillorn, D.H. Lowndes, M.L. Simpson, J. Phys. Chem. B **106**, 10570 (2002)
58. S.C. Chang, T.C. Lin, C.Y. Pai, Microelectron. J. **38**, 657 (2007)
59. G. Shivkumar, S.S. Tholeti, M.A. Alrefae, T.S. Fisher, A.A. Alexeenko, J. Appl. Phys. **119**, 113301 (2016)
60. B.A. Cruden, A.M. Cassell, D.B. Hash, M. Meyyappan, J. Appl. Phys. **96**, 5284 (2004)
61. Z.L. Tsakadze, K. Ostrikov, C.H. Sow, S.G. Mhaisalkar, Y.C. Boey, J. Nanosci. Nanotechnol. **10**, 6575 (2010)
62. W.Z. Li, J.G. Wen, Z.F. Ren, Appl. Phys. A Mater. Sci. Process. **73**, 259 (2001)
63. S.H. Jo, Y. Tu, Z.P. Huang, D.L. Carnahan, D.Z. Wang, Z.F. Ren, Appl. Phys. Lett. **82**, 3520 (2003)
64. Z. Xu, X.D. Bai, E.G. Wang, Appl. Phys. Lett. **88**, 133107 (2006)
65. M. Chhowalla, C. Ducati, N.L. Rupesinghe, K.B.K. Teo, G.A.J. Amaratunga, Appl. Phys. Lett. **79**, 2079 (2001)

Journal Pre-proof

# Impedance Based Damage Assessment of Concrete Under the Combined Effect of Impact and Temperature Using Different Piezo Configurations

Indrajeet Singh, Nirendra Dev, Shilpa Pal




PII: S0924-4247(22)00399-5

DOI: https://doi.org/10.1016/j.sna.2022.113763

Reference: SNA113763

To appear in: *Sensors and Actuators: A. Physical*

Received date: 8 May 2022
Revised date: 4 July 2022
Accepted date: 12 July 2022

Please cite this article as: Indrajeet Singh, Nirendra Dev and Shilpa Pal, Impedance Based Damage Assessment of Concrete Under the Combined Effect of Impact and Temperature Using Different Piezo Configurations, *Sensors and Actuators: A. Physical,* (2022) doi:https://doi.org/10.1016/j.sna.2022.113763

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.

© 2022 Published by Elsevier.

Journal Pre-proof

# Impedance based damage assessment of concrete under the combined effect of impact and temperature using different piezo configurations

**Indrajeet Singh[1], Nirendra Dev[2], Shilpa Pal[3*]**

[1] Research scholar, Civil Engineering Department, Delhi Technological University, Delhi
[2] Professor, Civil Engineering Department, Delhi Technological University, Delhi
[3] Associate Professor, Civil Engineering Department, Delhi Technological University, Delhi

** Corresponding Author shilpapal@dtu.ac.in*

**Abstract-** The concrete structures undergo severe damage under the impact loading, which may cause further fatal casualties to life and property and these damage become more dangerous as the temperature rises. This is controlled by performing the proper structural health monitoring (SHM) of structures. In recent years, the electromechanical impedance technique (EMI) has emerged as a promising method for monitoring the health of concrete structures. The objective of this study is to determine the damage in concrete under the impact and thermal loading circumstances using the EMI technique. For this purpose, an experimental study was conducted to investigate the effect of impact loading on concrete at varying temperatures using different piezo sensor configurations. The concrete cube was tested for three piezo sensor configurations, bonded, non-bonded and embedded with a lead zirconium titanate (PZT) patch. A free-falling iron ball was dropped from a fixed height above the top surface of the test samples to simulate the impact loading. Temperature heating was performed using the oven for each concrete cube at 50 $^{o}$C, 100 $^{o}$C and 150 $^{o}$C. The impedance measurement device was used to acquire the signatures from the piezo sensors. Statistical parameters such as root mean square deviation and mean absolute percentage deviation were calculated from the signature, demonstrating a high degree of damage detection under the impact loading at higher temperatures. These signatures are also used to determine the PZT equivalent system parameters (damping, stiffness, and mass) of the specimens under various damage conditions, as well as to establish the remaining life of the model. The study revealed that the EMI technique in SHM is capable of accurately assessing damage under impact loading at rising temperatures. Although all three sensor configurations are capable of detecting damage but, the study shows that the embedded sensors provide more accurate results.

**Keywords:-**Damage; EMI Technique; PZT; Impact Load; SHM**.**

## 1 Introduction:

The exposure of structures to severe dynamic events such as terrorist attacks, blast waves, earthquakes and industrial accidents has become more in recent years. These sorts of occurrences, which are associated with high

loading rates, are frequently followed by fire outbreaks and spread, exposing structures to highly difficult conditions**.** Very few publications demonstrate the damage assessment of concrete under the action of impact load at varying temperatures. The majority of the literature only indicates the effect on concrete when impact and temperature are individually applied. The researcher has not examined many areas in identifying damage produced by the combined action of impact and temperature. This study uses the electromechanical impedance (EMI) technique to measure concrete damage caused by the combined effects of impact and temperature. The EMI technique has been used in a variety of applications, such as in concrete hydration monitoring [1]–[4] and damage detection in RC structures in both embedding and surface-bonded conditions [5][6], etc. Due to its advantages of high sensitivity, quick response, easy installation, cheap cost, and intelligent real-time detection, the EMI approach has gained importance as a useful non-destructive technique for SHM.

Under impact loading, concrete structures experience two types of failures: global and local failure. Crushing, plug development, and scabbing are common causes of local concrete failure, whereas bending causes global failure. Drop impact tests were successfully performed on the concrete materials to observe the impact behavior of concrete [7]. The behavior of concrete under the static and impact Loading at the rate of 10 micro strain per second and 5 -10 strains per second simultaneously and compared the responses of both loadings using parameters like maximum strength and corresponding axial strain, the energy-absorption capacity at failure, and the secant modulus [8]. Many researchers have investigated the behavior of the slab under impact loading [9]–[15]. Youmn Al Rawi et al. [16] studied the influence of impact loading on the behavior of post-tensioned concrete slabs. They examined the displacements, impact forces, cracks, and damage types of two different grades of post-tensioned RC slabs and compared it from the conventional RC slab subjected to a drop impact of 605 kg and a height of 20 m. Impact load analysis of beam was also studied by many researchers [17]–[20]. Liu Jin et al. [21] has been studied the behavior of SFRC beam they have analyzed the impact resistance of 12 simply-supported steel fiber reinforced concrete (SFRC) beams with different stirrup ratios (0%, 0.253%, and 0.502% and different volume fractions of steel fibers (0%, 1%, 2%, and 3%) was tested. A three-dimensional finite element simulation was also performed. Steel fibers have been discovered to increase the impact resistant performance of SFRC beams, such as crack pattern, ductility, energy consumption capacity, and deformation recovery ability.

The effect of low and high temperatures on concrete has been already studied by many researchers [22]–[28]. Concrete's ability to resist high temperatures is one of its most significant advantages as a building material. The disturbance in concrete caused by high temperatures presents itself in the form of surface cracking and spalling

[29]. Xia. et al. [30] has been found that thermal distribution and transmission in concrete is generally non-uniform and time-dependent due to its properties of anisotropy and porosity, which leads to a negative correlation between concrete Young's modulus/natural frequencies and temperatures. Concrete structures such as beams and slabs are also very much affected by the high temperature condition [30]–[33].

Many experiments were carried out under the combined effect of impact and temperature without the use of piezo sensors. The damage behaviour of concrete was investigated using the combined effect of Thermal and Dynamic Loadings [34]. They performed the experiment at the high temperature and impact loadings conditions and observed that the decrement of strength of concrete up to 80 % under the combined effect of impact and temperature at high temperature level. The numerical analysis of the three-dimensional RC slab under the higher temperature and low velocity impact was done to examine the performance. A total of six slabs were tested, three of which were tested at high temperatures and the remaining three after cooling down. They used a constant impact velocity of 6.86 m/s and a drop weight of 105 kg. They discovered that as the heating duration of the temperature increases, the mid span deflection of the RCC slab increases dramatically [35]. Fire resistance of steel fibre reinforced concrete(SFRC) beam were studied after the application of low velocity impact loading [36]. They were performing an experiment on a six SFRC beam with the different volume fractions of steel fibers 0%,1%, 2% and 3%. Using the cylindrical hammer as impactor of size 200mm, weight of 393 kg under the 5.4 m/s impact velocity. They discovered the SFRC beam's failure patterns, thermal profiles, and mid-span deflection curves and validated all of their findings using numerical simulation and conclude that as the dose of steel fibres increases, the failure patterns of SFRC beams change from shear to flexure under impact loadings.

In EMI technique, vibrational response in the form of electrical impedance or admittance is actuated and recorded using an impedance measurement device (LCR meter)**.** A thin lead zirconium titanate (PZT) patch has been used in the crystal form, behaves as both an actuator and a sensor simultaneously. The exciting range of PZT is typically 30 to 600 kHz. The PZT patch is either embedded, non-bonded or surface bonded to the RC structure that is being monitored. When a sinusoidal signal actuates a PZT patch, the shape of a patch changes and mechanical vibrations occur. Change in mechanical behavior was recorded in the form of electrical admittance signature using LCR meter. The coupled EMI equation of effective impedance model as per shown in Equation 1 [37].

$$\bar{Y} = G + Bj = 4\omega j \frac{l^2}{h}\left[\overline{\varepsilon_{33}^{T}} - \frac{2d_{31}^2 \overline{Y^E}}{(1-\nu)} + \frac{2d_{31}^2}{(1+\nu)}\left(\frac{Z_{a,eff}}{Z_{s,eff}+Z_{a,eff}}\right)\bar{T}\right] \quad 1$$

Where G and B is the conductance and susceptance simultaneously with its imaginary unit *(j)*; width *(w)*, length *(l)*, height *(h)*, angular frequency *(ω)*, effective impedance of the structure *($Z_{s,eff}$)*, effective impedance of the PZT sensor *($Z_{a,eff}$)*, $\overline{Y^E}$ is the complex Young's modulus of the PZT sensor as per given in Equation 2.

$$\overline{Y^E} = Y^E(1 + \eta j) \quad 2$$

Electrical permittivity of the PZT material $\overline{\varepsilon_{33}^{\mathrm{T}}}$ is

$$\overline{\varepsilon_{33}^{T}} = \varepsilon_{33}^{T}(1 - \delta j) \quad 3$$

Mechanical loss factor *(η)*, dielectric loss factor *(δ)*, piezoelectric strain coefficient *($d_{31}$)*, complex tangent ratio $\bar{T}$

$$\bar{T} = \frac{tan(kl)}{kl} \quad 4$$

where, $k$ is the 2D wave number, related to the density ($\rho$) and the elastic properties.

$$k = \sqrt{\frac{\rho(1-\mu^2)}{\overline{Y^E}}\rho} \quad 5$$

and $\mu$ is the Poisson's ratio. Any change in the monitored structure's mechanical characteristics induces a change in the mechanical impedance, which affects the signature of the admittance *Y*, indicating damage. Soh and Bhalla [37] were the first to use EMI data for non-destructive strength monitoring of conventional structural systems. Later, employing three types of piezo sensors composition, such as surface-bonded sensors [38]–[39], embedded sensors [40]–[45], and non-bonded sensors [47]–[52], numerous studies used the EMI technique for strength and damage monitoring of the concrete structures. Many researchers have already investigated the effect of temperature on the EMI approach and discovered that the EMI technique is capable of identifying damage in a wide range of temperatures. An increase in temperature causes a slight magnitude change in the real part of the electrical impedance of a free PZT patch, resonance frequency shifts, and peak magnitude changes collected from a single carbon-steel beam [53]. By establishing a theoretical model of a spring-mass-damper (SMD) system, separation of the impedance measurement into active portion and analyzed the temperature effect, and the active signatures were found to have considerably greater tolerance to temperature changes [54].

Although a large body of literature shows the individual effects of impact and temperature under the EMI technique, the damage identification under the effect of impact loading at varying temperatures using the EMI technique has yet to be investigated. Because conventional non-destructive techniques do not allow for real-time

monitoring of structures, the EMI technique can play an important role in real-time monitoring of structures under the combined effect of impact and temperature. In this current study experimental work was performed on the concrete under the combined effect of impact and temperature using different type of piezo sensors configurations. This study explains the feasibility of different type of piezo sensors like non bonded piezo sensors (NBPS), surface bonded piezo sensors (SBPS) and jacketed piezo sensors (JKTPS) under the impact loading at varying temperature.

## 2 Experimentation Details

This section describes the entire experimental setup as well as the experimental procedure used to examine the combined effect of temperature and impact on the concrete cube of size 150 x 150 x 150 mm having 28 days strength as 30 N/mm$^2$.For making the concrete cube, mix in the proportion of 1:1.84:3.48(C:FA: CA) was used. All the cube specimens were demoulded after 24 hours and cured in a water tank for 28 days. After that, the specimens were placed in a laboratory environment until they were tested.The whole experimental setup with the guided tube for impact tests is shown in Figure 1.A free-falling cast iron ball of mass 5.45 kg was dropped from heights of 3 m on the concrete cubes resting on the cemented floor. A 4 mm thick PVC (Polyvinyl chloride) pipe with a 150 mm internal diameter was clamped to the roof parapet wall to serve as a guiding tube. Purpose of guided tube is only that, at each impact, the steel ball was able to hit precisely at the top of the cube. Rectangular oven of: W 605 x D 605 x H 910 mm were used for heating the concrete cube provided by Khera Instruments Pvt. Ltd. having maximum heating temperature limit is up to 150$^o$C .

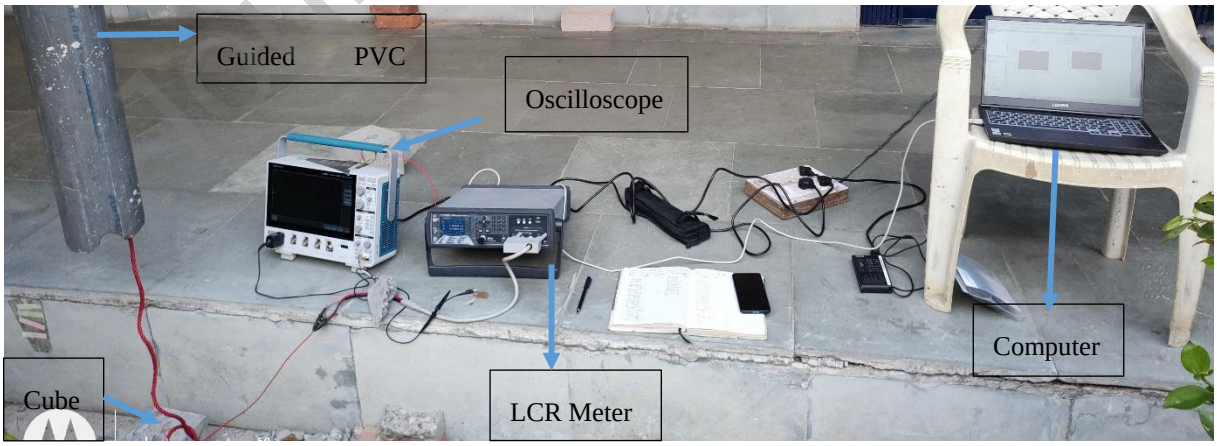



*Figure 1 Total Experimental Setup for impact loading*

## 2.1 Fabrication of resign jacketed piezo sensors (JKTPS) and non bonded piezo sensors (NBPS)

The JKTPS as shown in Figure 2 is of diameter 25mm and height 20mm are made up of rectangular PZT patches of size 10 x 10 x 0.2 mm place at the center and the epoxy resin provided at top and bottom portion for covering the PZT to protect them from the impacts of water and ambient conditions during the construction stages. JKTPS are made in a step-by-step process, in which a cap is first half-filled with epoxy, then the PZT are placed, and the cap is fully filled with epoxy and left out for 24 hours for the epoxy to set. Table 1 shows the material parameters of the PZT patch in detail.

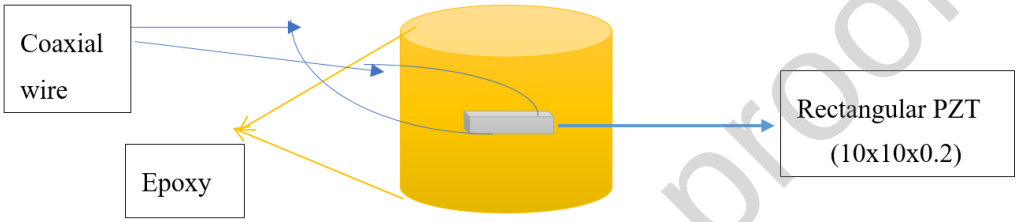



*Figure 2 Jacketed piezo sensors (JKTPS) with PZT patch*

*Table 1Material Properties of PZT*

| **Material parameters Piezoelectric** | **Value** |
|---|---|
| Piezoelectric charge coefficient $d_{33}(10^{-12}$ C/N) | 500 |
| Piezoelectric voltage constants ($g_{33}$,$10^{23}$ Vm/N) | 21.2 |
| Young's modulus $YE^{33}$ ($10^{10}$ N/m) | 5.0 |
| Mechanical quality factor $Q_m$ | 76 |
| Curie temp ($^oC$) | 240 |
| Density ($10^3$ kg/$m^3$) | 7400 |
| Dielectric constant | 2500 |
| Poisson's ratio | 0.35 |
| Dissipation factor at 1KHz | 0.02 |

The fabrication of the NBPS was accomplished by using an aluminium strip ( 100 x 10 x 1 mm) that had been surface cleaned and a PZT patch ( 10 mm x 10 mm x 0.2 mm) was fixed in the middle of the aluminium strip. For the bonding of aluminium sheet and PZT, Epoxy was used as a binding material, and special care was taken to ensure that they were properly bonded. If the bond layer is not adequately accounted for in the EMI approach, it can have a considerable impact on the signal transmission from the PZT material. The PZT sensor generates a permanent finite thickness interfacial layer between the structure and the patch when it is glued to it with an adhesive mix (such as epoxy). Force transmission from the PZT patch occurs via shear mechanism through the interfacial bond layer [55], [56].

Journal Pre-proof

As described in the preceding section, three different piezo sensors configurations were made and concrete cubes are attached with these three distinct piezo sensor configurations, embedded (JKTPS), Surface bonded (SBPS) and non-bonded (NBPS) as shown in Figure 3. JKTPS are placed at the center of the concrete cubes at the time of casting considering equal distance from the top and bottom. SBPS are directly attached to the concrete with the help of epoxy and NBPS are indirectly attached with the concrete in which first PZT patch were placed at the center of the aluminum sheet and the aluminum sheet were bonded with the concrete cube using the epoxy. In the current study total number of twelve JKT sensors has been made for performing the experiment.

## 2.2 Experimental procedure

To begin, baseline data were collected for all three different sensor configurations in the form of conductance in function of frequency and susceptance in function of frequency plots in the 50 kHz to 600 kHz frequency ranges. After that, the cube was subjected to three different temperatures, ranging from 50 °C, 100 °C and 150 °C. Cubes were heated at a rate of 2 $^{o}$/s for 120 minutes and then cooled at room temperature for 60 minutes.

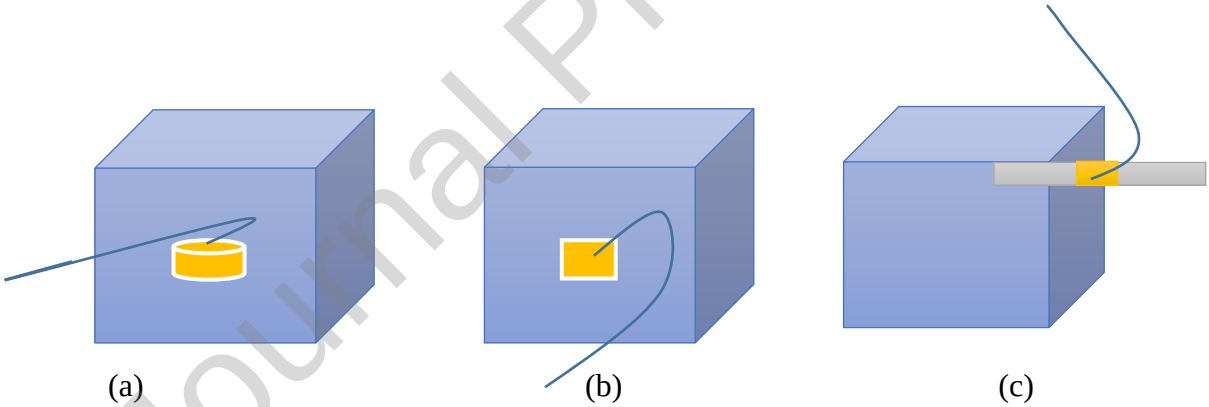



*Figure 3 placing of different sensors through host structure a) JKTPS b) SBPS c) NBPS*

Signatures are taken for both conditions, i.e. immediately after heating and after 60 minutes cooling down at room temperature. After cooling down, each cubes are suspended to the impact loading and record the signatures after the each impact up to the failure. The features of the specimens, the height of the fall, Velocity and the energy absorption for different number of impacts are summarized in Table 2. Figure 4 shows the block diagram of the concrete cube with the impactor.

*Table 2 Features of the specimen*

Journal Pre-proof

| No. of Impacts | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Height of Fall *(m)* | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| Velocity*(m/s)* | 7.6 | 7.6 | 7.6 | 7.6 | 7.6 | 7.6 | 7.6 | 7.6 | 7.6 | 7.6 | 7.6 |
| Energy absorption*(J)* | 160.3 | 320.7 | 481.1 | 641.5 | 801.9 | 962.3 | 1122.7 | 1283.1 | 1443.5 | 1603.9 | 1764.3 |

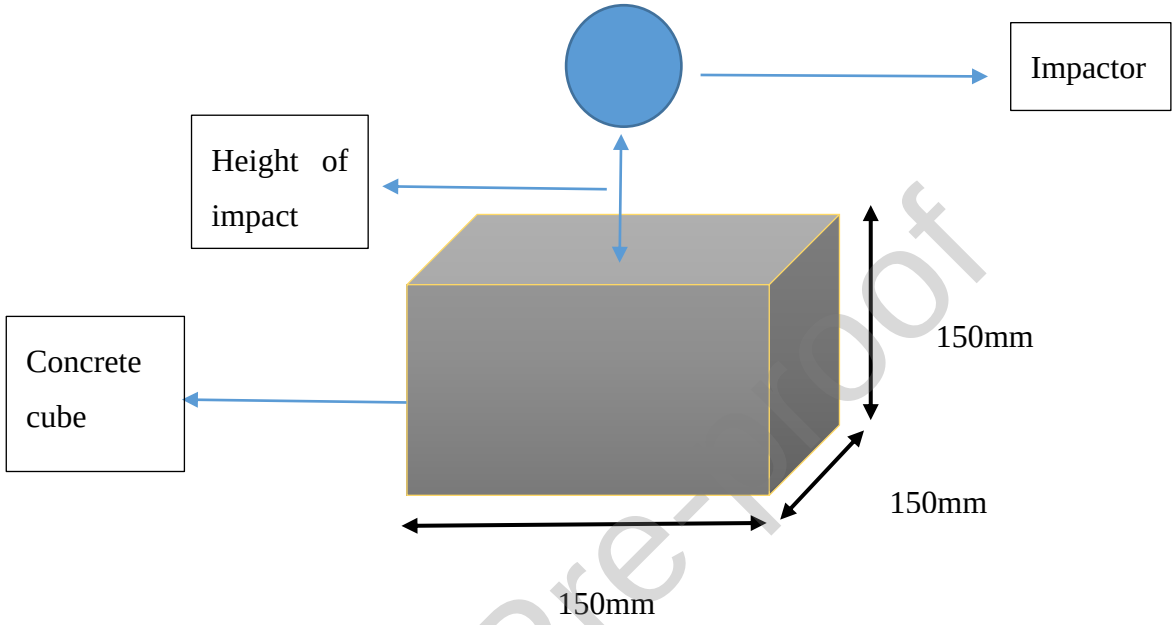



*Figure 4 Concrete cube with the steel ball impactor*

### 2.3 Data Acquisition

The data acquisition has been acquired using an LCR meter. The Keysight E4980A Precision LCR meter, 20 Hz to 2 MHz [57] was used to acquire the conductance (G) and susceptance (B) data from the PZT. The frequency step used for data acquisition was 100 Hz, with a voltage of 1 V applied across the PZT patch's terminals .Each of the sensors is connected with coaxial wire through soldering, and other side of the coaxial wire are directly attached to the LCR meter. The V.Pro software provided by Keysight, i.e., already compatible to the LCR meter, was used for collecting the data of *G* in function of *f* and *B* in function of *f*. Table 3 shows a summary of the total cube tested with the various sensor configurations at varying temperatures, as well as the total number of impacts up to failure

*Table 3 Summery of the different sensors configuration with varying temperature and number of impact up to failure*

| S.No. | Sensors Configuration | Height of Impact | Temperature | No. of impact up to failure |
|---|---|---|---|---|
| 1 | JKTPS (Embedded) | 3m | Ambient | 7 |
| | | | 50 $^{o}$C | 6 |
| | | | 100 $^{o}$C | 7 |
| | | | 150 $^{o}$C | 10 |

| | | | | |
|---|---|---|---|---|
| 2 | SBPS (Surface Bonded) | 3m | Ambient | 8 |
| | | | 50 $^{o}$C | 8 |
| | | | 100 $^{o}$C | 7 |
| | | | 150 $^{o}$C | 12 |
| 3 | NBPS (Non Bonded) | 3m | Ambient | 11 |
| | | | 50 $^{o}$C | 11 |
| | | | 100 $^{o}$C | 9 |
| | | | 150 $^{o}$C | 7 |

## 3 Experimental results and discussion

The experimental findings and results are discussed and interpreted in this section in the following phases. First, case of the physical assessment under impact's damage for each concrete cube was analyzed. In addition, the signature data obtained from the LCR meter was thoroughly discussed.

Journal Pre-proof

### 3.1 Inspection of visible surface cracks after each impact

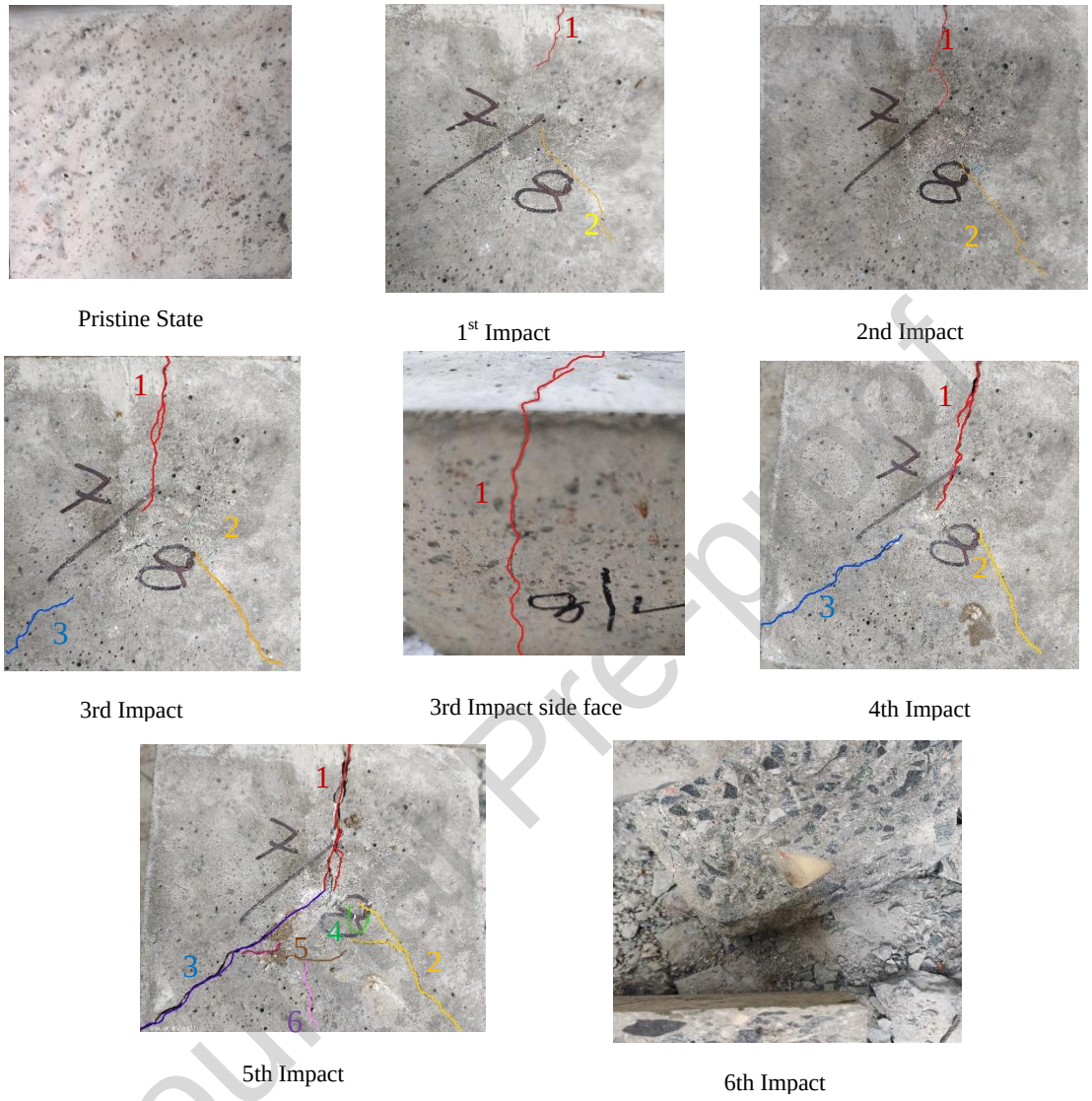



*Figure 5 Crack pattern for cube up to failure under 3m height at 50$^{o}$ C temperature*

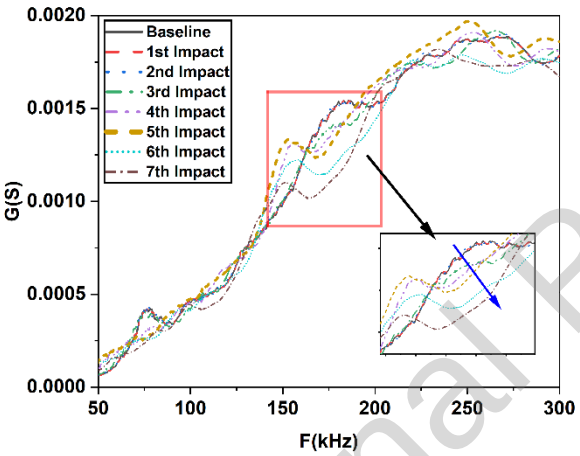
As the number of impacts increases, the cube sustains more cracks until it fails. Under the 3m height of impact, one cube specimen considered JKTPS was fully failed in six numbers of impacts at 50$^{o}$C temperature. The surface condition of the cubes at the impact face and side face can be analyzed after each impact, as shown in Figure 5. The first and second cracks that formed after the first impact are highlighted in red and yellow in Figure 5. During the second impact, the width of the cracks grows. The third crack is highlighted in blue; it was formed after the third impact, and the first crack was spread to the side face as well; these three previous cracks developed during the third impact became wider after the fourth impact, and in the fifth impact, three more cracks are formed around the center of the cube before failing completely in the sixth impact.

## 3.2 Damage assessment using EMI Technique

This section discusses the assessment of damage in concrete under impact loading for various temperature conditions using the EMI technique, as well as the effect of temperature on concrete in detail.

### 3.2.1 Damage assessment concrete at ambient temperature for different sensors

The admittance signatures of cubes were obtained in their pristine state from all the different sensors as a baseline signature and compared to the damage state signatures. The variation in the conductance signature verses frequency with increasing damage under the impact loading at ambient temperature is plotted in Figure 6 to Figure 8 for various sensor configurations of JKTPS, SBPS and NBPS respectively.



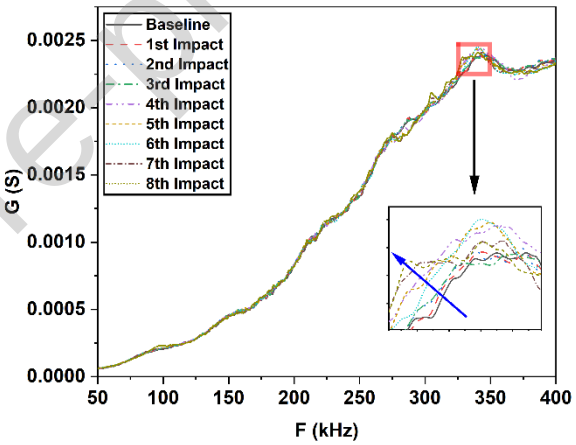
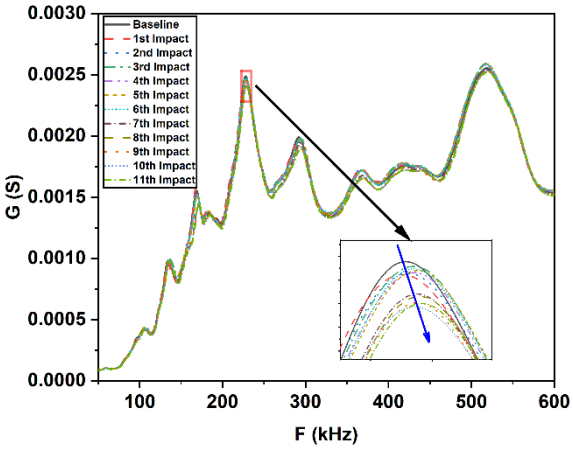




*Figure 6 conductance signature for SBPS at ambient temperature*




Journal Pre-proof

*Figure 8 conductance signature for NBPS at ambient temperature*

Signatures are plotted under the frequency range of 50 kHz to 600 kHz. When compared to the baseline signature and the damage state signature, the conductance signature plays an important role in assessing structural damage. Changes in mass, damping, and stiffness is due to the development of cracks, damage, corrosion, and other factors were calculated using the conductance signature. Susceptance signatures are used to assess the state of the PZT patch. If the PZT patch was attached to the host structure and had to be broken down or detached from the host surface throughout the experiment, the variation of the susceptance signature with increasing frequency was shown to be a downward or uneven variation. Figure 9 shows a baseline conductance signature of different sensors configuration JKTPS, SBPS, NBPS for 3m height of impact at ambient temperature.

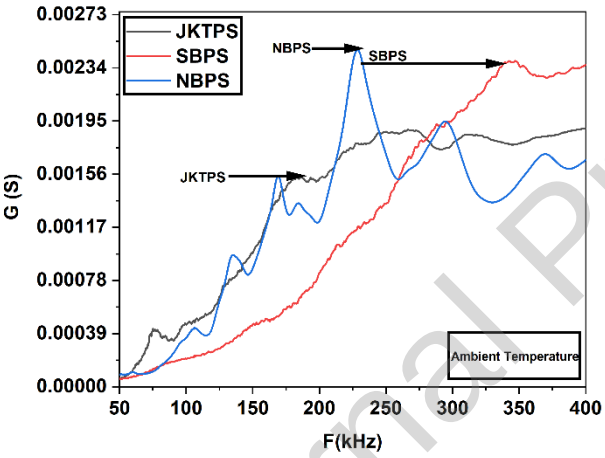



*Figure 9 conductance signature of baseline data for different sensors configuration*

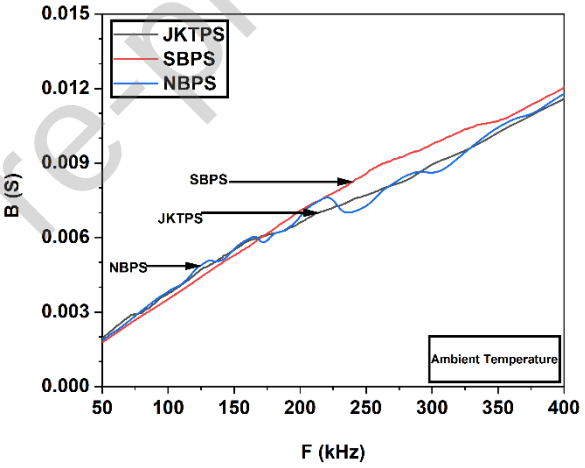



*Figure 10 Baseline suseptance signature for different sensors configuration*

Because of their different sensing capabilities, each sensor has a unique conductance signature. SBPS sensors have achieved their first peak between 300 kHz and 375 kHz. JKTPS obtained its first peak between 150 and 225 kHz. The first peak of the NBPS was obtained in the frequency range of 100-150 kHz, but the optimum sharp peak with a high amplitude was obtained in the frequency range of 175-275 kHz. All of this is due to the sensors' varying sensing behaviour. Figure 10 depicts a baseline susceptance signature of the various sensor configurations at ambient temperature for a 3 m height impact loading. All of the sensors recorded flat signatures in comparison to conductance signatures because they are less sensitive to mechanical and environmental changes. They are only sensitive to PZT sensors, as well as PZT debonding from the host structure and loose connections**.** After acquiring the signature of various sensor configurations at ambient temperature, concrete cubes are subjected to different temperature conditions of 50, 100, and 150 $^{o}$C for 2 hours,

then cooled for one hour, and then subjected to impact loading at a heights of 3m. Figures 11–13 depict the temperature variation of the conductance signature for JKTPS, SBPS, and NBPS, respectively. All of the signatures were recorded after heating the cube in the oven for three distinct temperature conditions ranging from 50 to 150 $^{o}$C and cooled it for one hour in an open atmosphere.

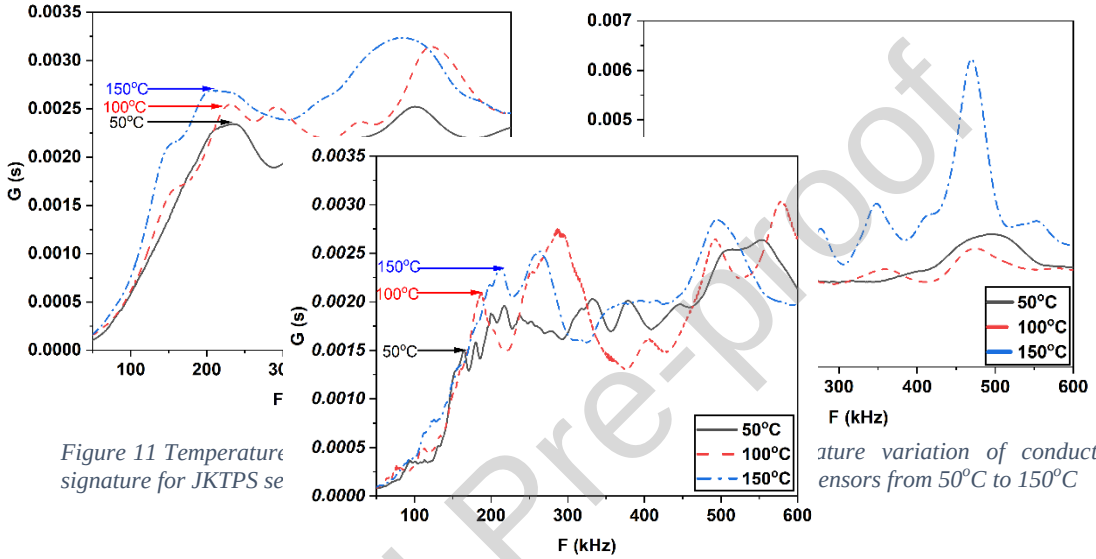



*Figure 11 Temperature signature for JKTPS se* *ature variation of conductance ensors from 50$^{o}$C to 150$^{o}$C*

*Figure 13 Temperature variation of conductance signature for NBPS sensors from 50$^{o}$C to 150$^{o}$C*

All the three different sensors configuration have shown higher amplitudes of conductance as the temperatures rises. This is due to the PZT's greater sensitivity to temperature conditions. Susceptance signature graphs of all the different sensor configurations with the varying temperatures 50 $^{o}$C to 150 $^{o}$C are shown in Figure 14 to Figure 16 and all these susceptance signatures were showing the similar variation against the frequency for each sensors configuration.

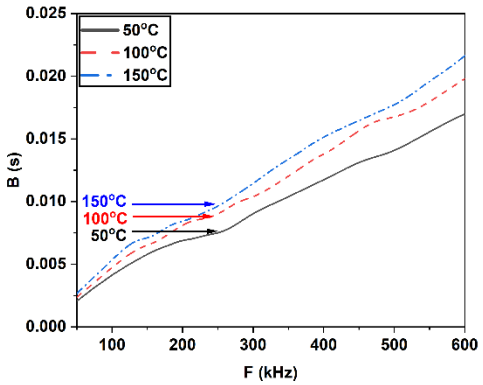



*Figure 14 Temperature variation of susceptance signature for JKTPS sensors from 50$^{o}$C to 150$^{o}$C*

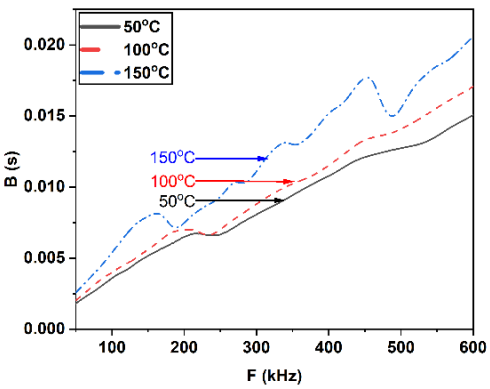



*Figure 15 Temperature variation of susceptance signature for SBPS sensors from 50$^{o}$C to 150$^{o}$C*

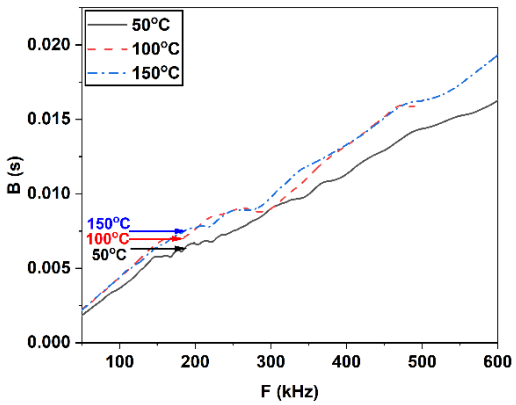



*Figure 16 Temperature variation of susceptance signature for NBPS sensors from 50°C to 150°C*

Journal Pre-proof

### 3.2.2 Damage assessment of concrete under the impact loading at varying temperature using JKTPS sensors.

Figure 17 to Figure19 depicts the graph between *G* in function of *f* for JKTPS sensors at 3 m height of impact loading up to the failure under the three different temperature variation of 50 $^{o}$C, 100 $^{o}$C and 150 $^{o}$C. As compared to the baseline signature, for the first impact, the signature shifts upward due to temperature, and for subsequent impact loads, the signature shifts downward up to the failure as shown in Figure 17, and this pattern of shifting signature is the same for the 100 $^{o}$C and 150 $^{o}$C temperatures as shown in Figures 18 and Figure 19 respectively.

Journal Pre-proof

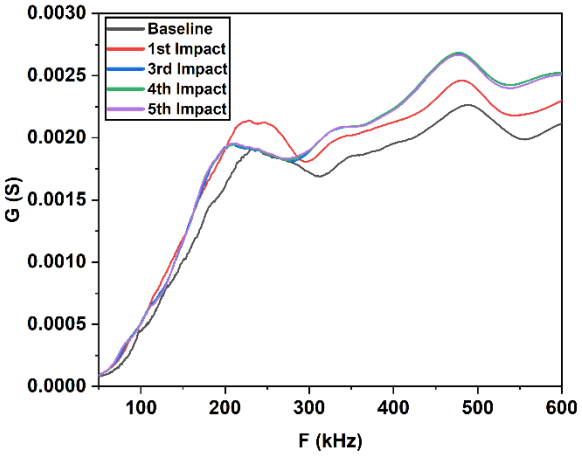



*Figure 17 Conductance signature for JKTPS at 50°C*

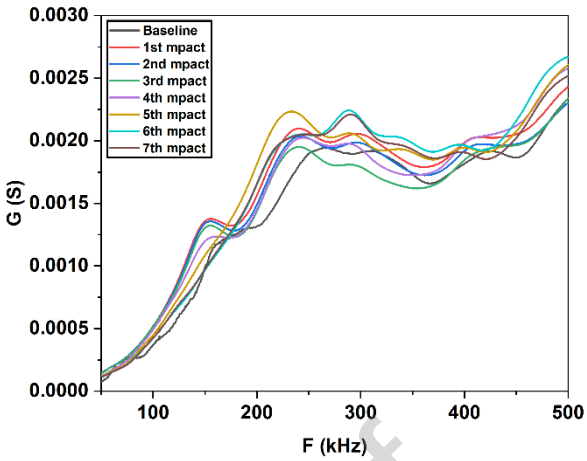



*Figure 18 Conductance signature for JKTPS at 100°C*

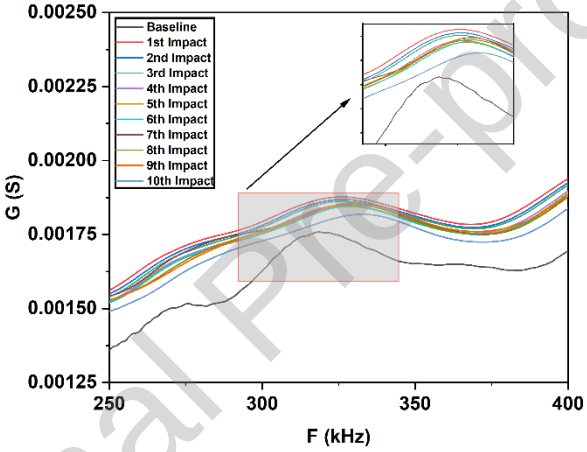



*Figure 19 Conductance signature for JKTPS at 150°C*

### 3.2.3 Damage assessment of concrete under the impact loading at varying temperature using SBPS sensors.

Figure 20 to Figure 22 depicts the graph between *G* in function of *f* for SBPS sensors at 3 m height of impact loading up to the failure under the three different temperature variation of 50 °C, 100 °C and 150 °C. For increase in temperature 50 to 150 °C, SBPS follow the same pattern As the JKTPS i.e. the shifting of the conductance signature were first in upward direction from pristine stage to first impact after that the conductance signature shifted downwards until failure.

Journal Pre-proof

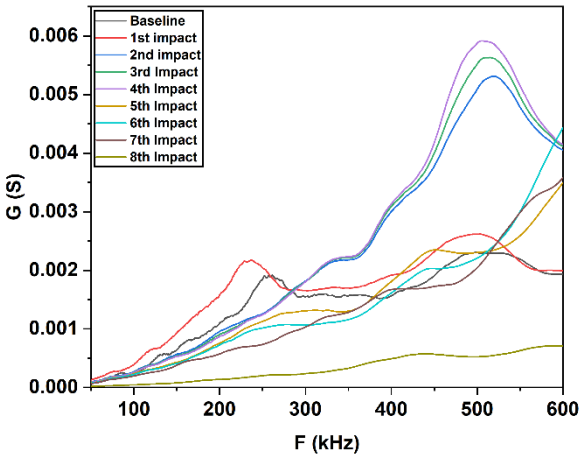



*Figure 20 conductance signature of SBPS at 50$^{o}$C*

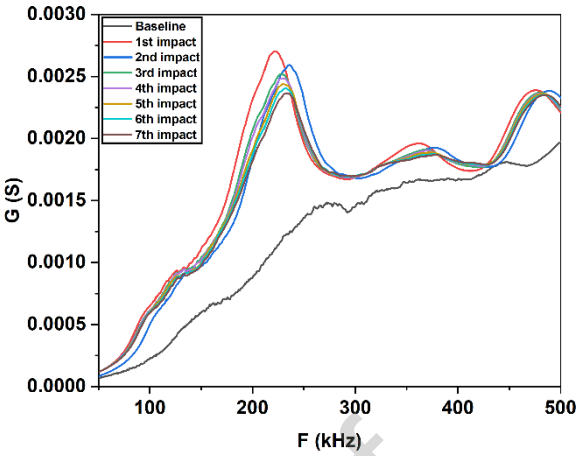



*Figure 21 conductance signature of SBPS at 100$^{o}$C*

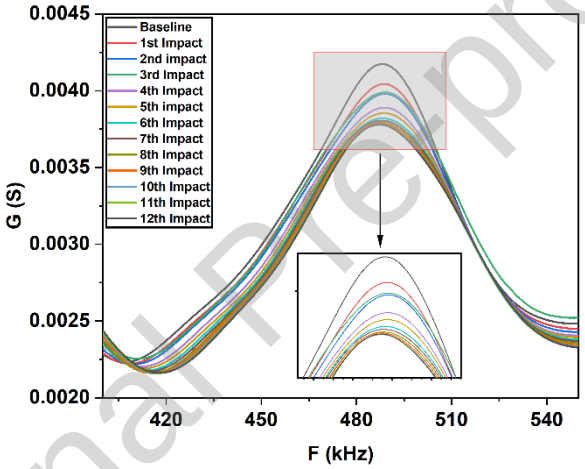



*Figure 22 Conductance signature of SBPS at 150$^{o}$C*

Journal Pre-proof

### 3.2.4 Damage assessment of concrete under the impact loading at varying temperature using NBPS sensors.

Figure 23 to Figure 25 depicts the graph between *G* in function of *f* for NBPS sensors at 3 m height of impact loading up to the failure under the three different temperature variation of 50 $^{o}$C, 100 $^{o}$C and 150 $^{o}$C and they were follows the same pattern as the two previous sensors JKTPS and SBPS.

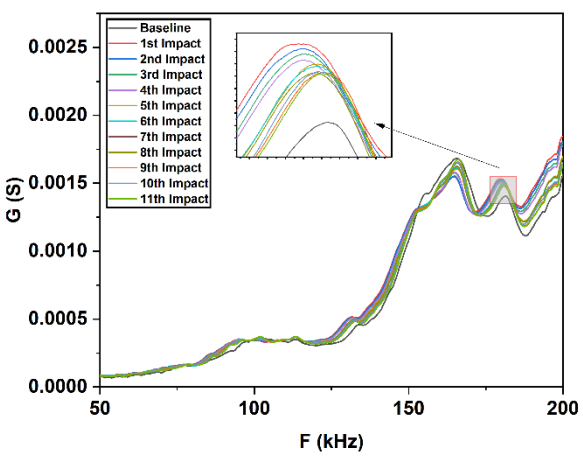



*Figure 23 Conductance signature for NBPS at 50°C*

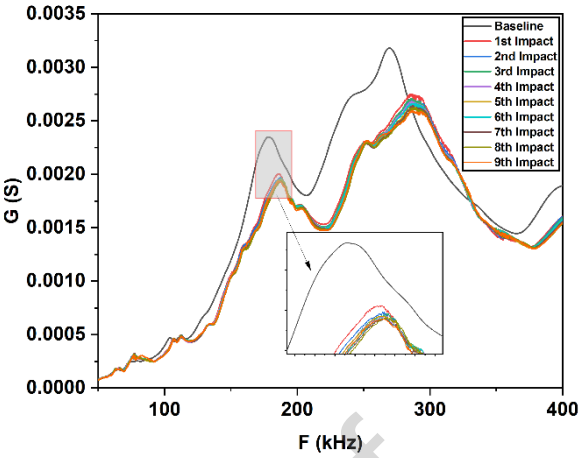



*Figure 24 Conductance signature for NBPS at 100°C*

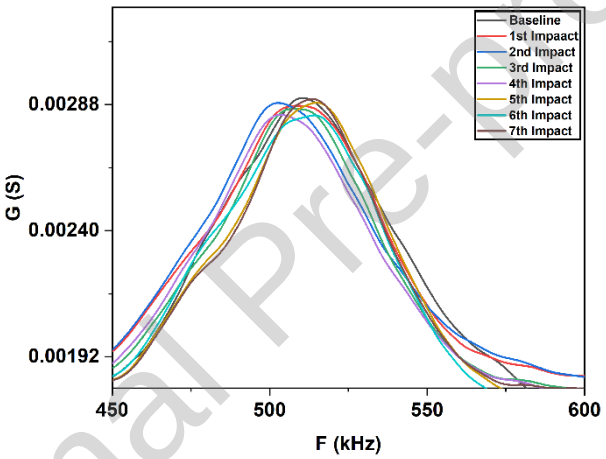



*Figure 25 Conductance signature for NBPS at 150°C*

## 3.3 Damage quantification using statistical technique

Signature evaluation was done by two different types of numerical indices equation**.** These are the root mean square deviation (RMSD) and mean absolute percentage deviation (MAPD) as equation (6)-(7). The RMSD and MAPD indices were used to quantify how much of the conductance signature of all three JKTPS, NBPS and SBPS has been altered due to impact damage at varying temperatures, as defined by,

$$RMSD = \left( \frac{\sum_{k=1}^{N} [Re(Z_k)_j - Re(Z_k)_i]^2}{\sum_{k=1}^{N} [Re(Z_k)_i]^2} \right)^{\frac{1}{2}} \quad 6$$

$$MAPD = \frac{1}{N} \sum_{k=1}^{N} \left| [Re(Z_k)_j - Re(Z_k)_i] / Re(Z_k)_i \right| \quad 7$$

Here, $Re\ (Z_k)_i$ is the reference impedance signature and $Re(Z_k)_j$ is the corresponding signature number of impedance signatures is represented by $N$. The RMSD and MAPD variations for each various sensor configuration are calculated independently under the 3 m height of impact loading at four different temperatures of ambient, 50,100, and 150 $^oC$ while keeping the pristine state as a baseline structure. Figure 26-31 depicts the RMSD and MAPD variation due to the combined effect of impact and temperature for each sensor configuration separately. Figure 26 depicts the fluctuation of the RMSD value for the JKTPS sensors from ambient to 150 $^oC$ temperatures. The RMAD value for JKTPS at ambient temperature follow an increasing pattern ranges from 2.17 to 10.37 % for complete failure up to the seven impacts. These variation goes increasing for the temperature of 50 to 150 $^oC$. For 50 $^oC$ RMSD variation varies from 11.479 to 15.182 %. However, the variation for 100 $^oC$ temperature were inconsistent. For 150 $^oC$ temperature variation were shown the downward pattern i.e. 12.38 to 7.86 %.

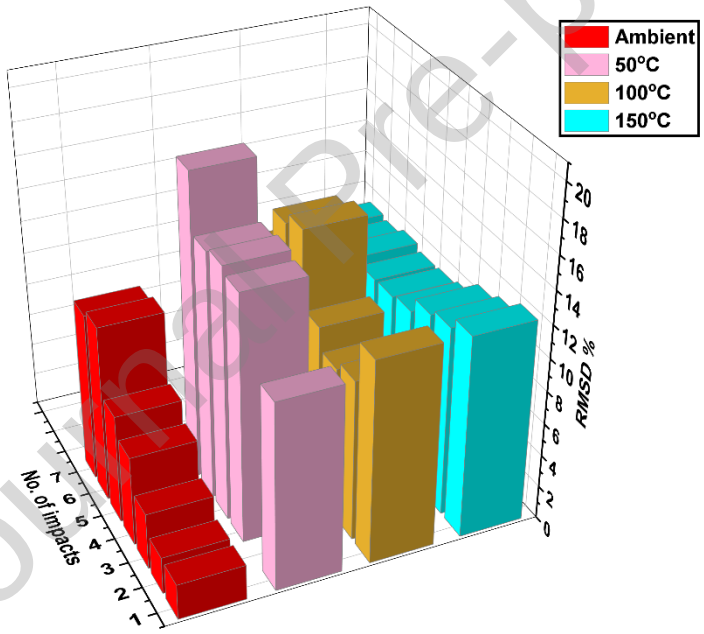



*Figure 26 RMSD value for JKTPS with increasing impact number for different temperature a) ambient b) 50 $^oC$ C) 100 $^oC$ d) 150 $^oC$*

Figure 27 depicts the fluctuation of the RMSD value for the SBPS sensors from ambient to 150 $^oC$ temperatures. For the pristine state they were follow an increasing pattern varied from 0.656 to 2.174 % as per same trend as a JKTPS sensors. However, the variation for the 50 $^oC$ were inconsistent. For 100 and 150 $^oC$ they follow the decreasing pattern from 46.98 to 36.78 % and 78.01 to 67.77 % respectively.

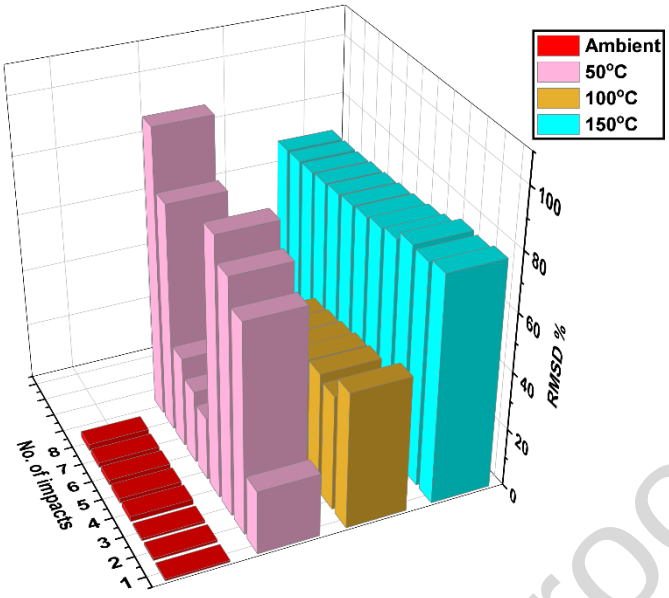
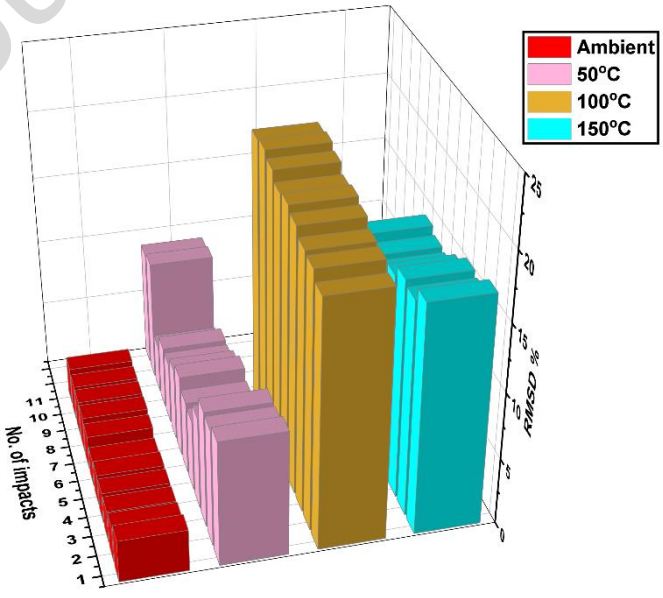



*Figure 27 RMSD Value for SBPS at increasing impact for different temperature conditions ambient, 50 °C, 100 °C and 150 °C*

Figure 28 depicts the fluctuation of the RMSD value for the NBPS sensors from ambient to 150 °C temperatures. The RMSD value variation for the ambient temperature was 3.16 to 2.82 %, they were followed a decreasing pattern from 9.4 % to 3.62 % at 50 °C and an increasing pattern from 18.5 to 20.84 % at 100 °C. Figure 29 depicts the fluctuation of the MAPD value for the JKTPS sensors from ambient to 150 °C temperatures. The MAPD value for ambient temperature varied from 0.4738 to 14.21 % in an increasing pattern, and the same pattern was followed for 50 °C from 13.74 to 16.89 %. MAPD variations at 100 °C are inconsistent, but at 150 °C they follow a decreasing pattern, decreasing from 17.98 to 12.28 %.




Journal Pre-proof

*Figure 28 RMSD Value for NBPS at increasing impact for different temperature conditions ambient, 50 °C, 100 °C and 150 °C*

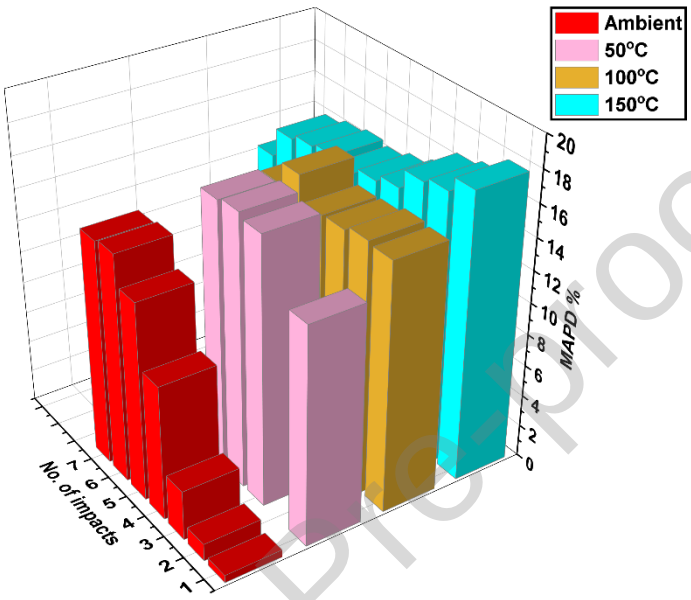



*Figure 29 MAPD Value for JKTPS at increasing impact for different temperature conditions ambient, 50 °C, 100 °C and 150 °C*

Figure 30 depicts the fluctuation of the MAPD value for the SBPS sensors from ambient to 150 °C temperatures. The MAPD value for SBPS sensors ranges from 0.966 to 2.81 % at ambient temperature. They did not follow any pattern at 50 °C, but at 100 and 150 °C, they followed a decreasing trend from 65.24 to 53.56 % and 95.49 to 81.93 %, respectively.

Journal Pre-proof

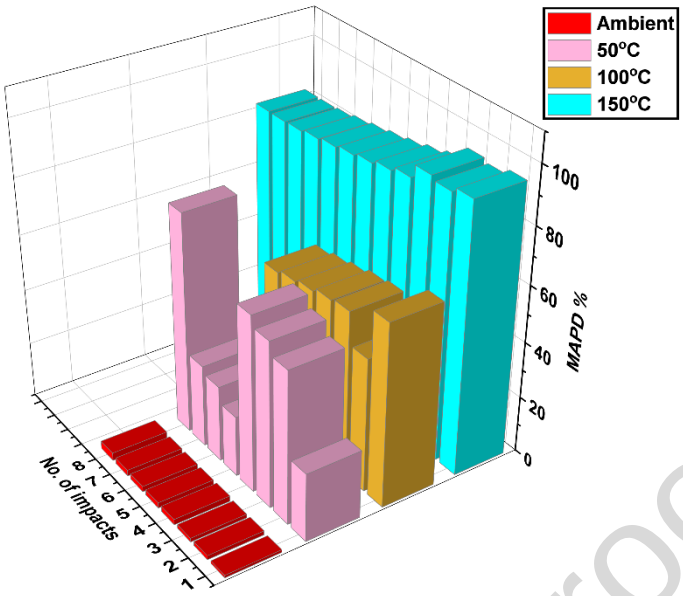



*Figure 30 MAPD Value for SBPS at increasing impact for different temperature conditions ambient, 50 °C, 100 °C and 150 °C*

Figure 31 depicts the fluctuation of the MAPD value for the NBPS sensors from ambient to 150 °C temperatures. The MAPD value has followed an increasing pattern from 1.396 to 3.22 % at ambient temperature and the same increasing pattern has followed for the 100 °C temperature from 14.63 to 16.31%. They followed a decreasing pattern from 10.34 to 4.22 % and 12.95 to 11.22 % for 50 and 150 °C, respectively.

Journal Pre-proof

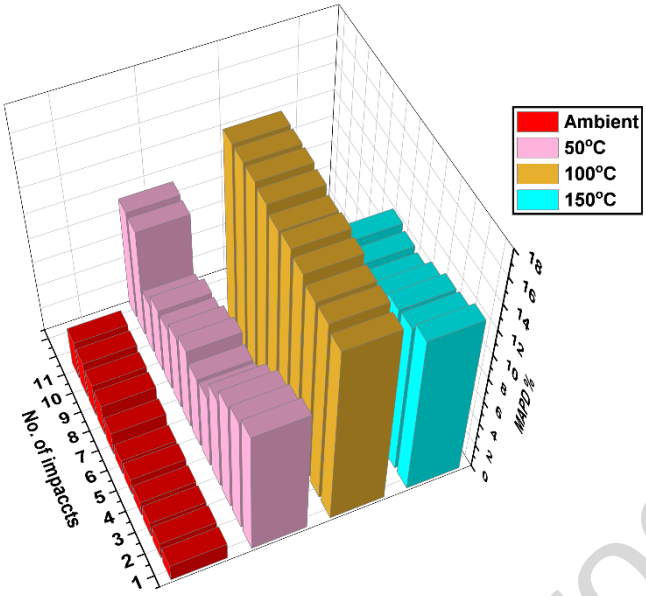



*Figure 31 MAPD Value for NBPS at increasing impact for different temperature conditions ambient, 50 °C, 100 °C and 150 °C*

## 3.4 Normalized frequency shift index (NFSI)

With the increase in the number of impacts, a shift in the resonant frequency peaks of the conductance signatures was detected. A non-dimensional component, normalized frequency shift index (NFSI), can be used to quantify this change. The NFSI was the most relevant parameter in the case of a very uneven change in the resonant frequency from the baseline, and it plays a crucial role in this study because it clearly illustrates how the resonance frequency was moved in the case of the combined impact and temperature effects. For an $n^{th}$ impact, the NFSI $f=f_0$ was calculated as a ratio of the shift in resonant frequency from the base signature to the resonant frequency in the nth impact as shown in Equation 8, where $\Delta f$ signifies the change in resonant frequency for the nth impact from base bas signature and $f_o$ signifies the resonant frequency of the base signature.

$$\boldsymbol{NFSI_{nth}} = \frac{\Delta \boldsymbol{f}}{\boldsymbol{f_o}} \qquad 8$$

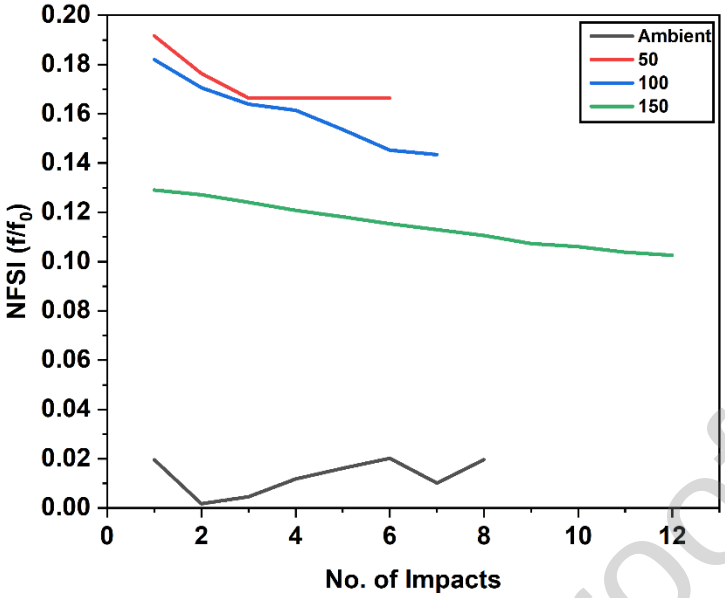



*Figure 32 NFSI Variation with the increasing number of Impacts for NBPS from ambient to 150 $^{o}C$*

Figure 32 depicts a NFSI variation for the NFSI plotted against the number of impacts from ambient to 150 $^{o}$C for the SBPS at an impact height of 3m. The graph depicts the decrease in NFSI value as temperature increases. A decreasing trend of the NFSI value indicates an increase in damage and this is due to the loss in the damping and stiffness of the structure as the number of impacts increased with temperature. The variation in the peak conductance and resonant frequency value were also be analyzed with the increase in number of impacts. Figure 33 shows the variation of peak conductance and resonant frequency in function of incremental of impact load for SBPS sensors at varying temperature. It was observed that the both the resonant frequency and peak conductance has follow the same pattern for 50 $^{o}$C, 100 $^{o}$C and 150 $^{o}$C. For every increase in temperature 50 to 150 $^{o}$C, the resonant frequency decreased from the pristine state to the first impact and then increased after the first impact, whereas the peak conductance increased from the pristine state to the first impact and then decreased after the first impact. The highly sensing behavior of PZT sensors against temperature causes the increasing or decreasing value of peak conductance and resonant frequency from pristine to first impact. The decreasing behaviour of the peak conductance after first impact was caused by an increase in crack width and depth as the number of impacts increased. The loss of stiffness causes an increase in crack depth and width. The loss of damping in the cube caused the resonance peak to rise in nature after the first impact as well as same as Negi and Bhalla et.al[58].

Journal Pre-proof

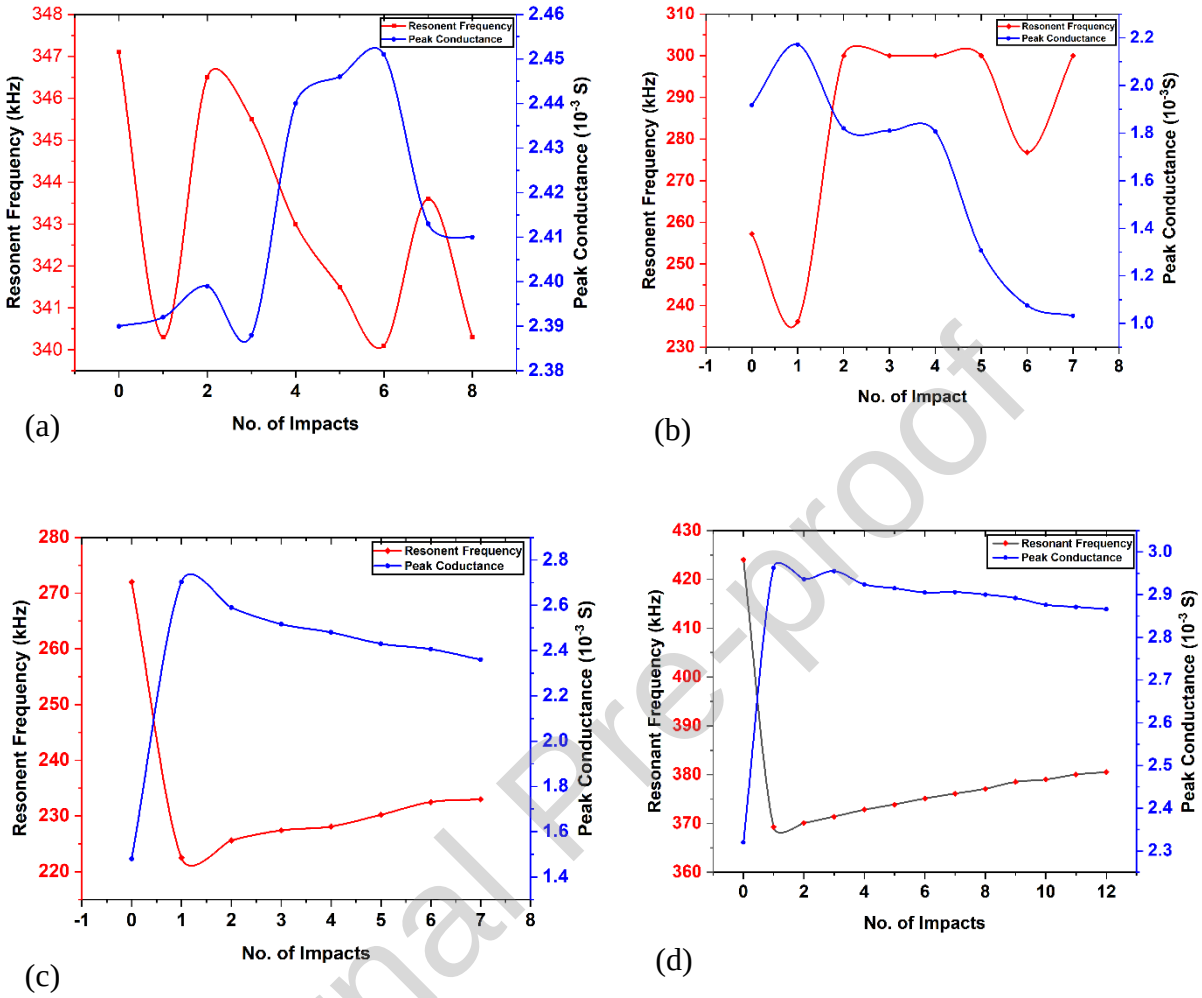



*Figure 33 Variation of Resonant frequency and peak conductance with the increase in impact for SBPS at varying temperature a) ambient b) 50 $^{o}C$ c) 100 $^{o}C$ d) 150 $^{o}C$*

### 3.5 Extraction of Equivalent Stiffness Parameters

Impedance-based discovery has become more useful in detecting damage in recent years. In the field of EMI technique, this is now a well-established concept that has proven to be more accurate than simply observing the change in conductance signature[59].A well establish relation was derived based on the identified equivalent system in between the changes in conductance signature that occur during the damage in concrete and a corresponding change in equivalent stiffness parameter (ESPs) such as stiffness, mass, and damping. From equation 1, Bhalla and Soh [37] derive the mechanical impedance ($Z_{s,eff}=x+yj$) for the monitored structure at a specific frequency ($\omega=2\pi f$). In this analysis for all cases of impact loading under temperature variation, signatures consisting of *G* and *B* plots are retrieved in the frequency range of 50 kHz-500 kHz, plot the graph of experimental plots in between *x* in function of *f* and *y* in function of *f*. Again, the experimental plots of *x* and *y*

were compared to the Hixon [18] equivalent system model. Figure 34 and Figure 35 represents the experimental plot of *x* and *y* with the frequency respectively at incremental damage state under the 3 m height of impact for SBPS at 150 $^{o}$C temperature. After examining the graphs of *x* in function of *f* and y in function of *f* through the equivalent plot, a dashpot*(c)*, spring*(k)*, and mass*(m)* equivalent system in series as depicted in Figure 36 (b) were found to be similar.

Figure 36 depicts the usual variation of *x* and *y* with the frequency for the identified series *m-c-k* system. The ESPs values were obtained using experimental *x* and *y* values within the frequency range, which were then entered into the (9) (10) (11) equation to calculate the m-c-k values. These absolute values of *m-c-k* are derived by rearranging the equations (12) and (13) with some boundary conditions.

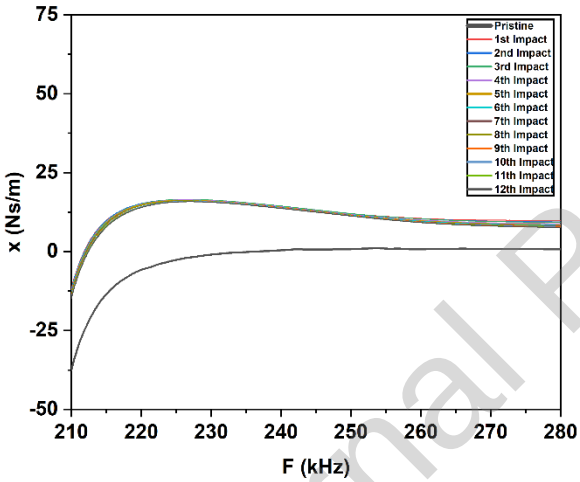



*Figure 34 mechanical impedance variation of x vs f*

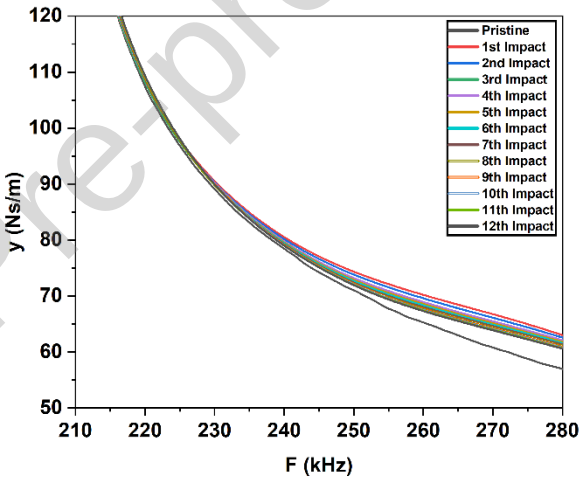


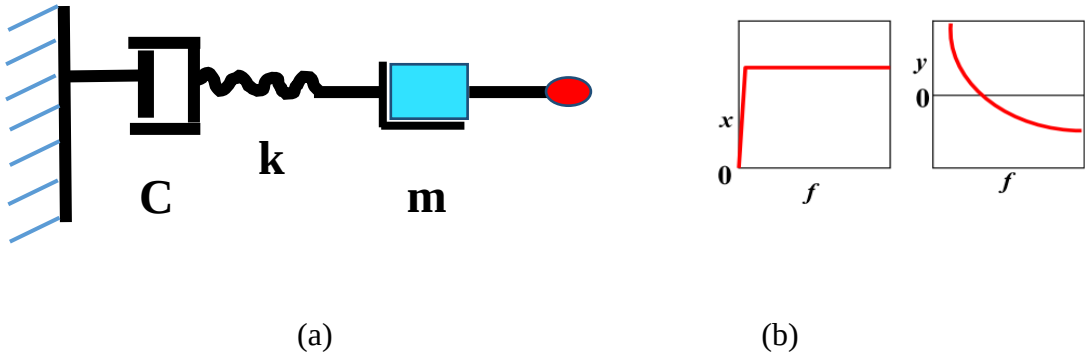

*Figure 35 mechanical impedance variation of y vs f*

$$m = \frac{(\omega_o{}^2 - \omega^2)(x^2 + y^2)}{\omega \omega_o{}^2 y} \qquad 9$$

$$k = \frac{(\omega_o{}^2 - \omega^2)(x^2 + y^2)}{\omega y} \qquad 10$$

$$c = \frac{(x^2 + y^2)}{x} \qquad 11$$

Where $\omega = \omega_o$ at $y = 0$. This Solution is valid for y>0 and $|\omega| < \omega_o$ or y<0 and $|\omega| > \omega_o$.After determining the average values of the m-c-k, the equivalent values of x and y are calculated using equations 12 and 13.




(a) (b)

*Figure 36 (a) Equivalent mechanical system consist of spring, mass and damper in series (b) variation of x and y for the identified series m-c-k system*

$$x = \frac{c^{-1}}{c^{-2}+\left(\frac{\omega}{k}-\frac{1}{\omega m}\right)^2} \qquad 12$$

$$y = \frac{-\left(\frac{\omega}{k}-\frac{1}{\omega m}\right)^2}{c^{-2}+\left(\frac{\omega}{k}-\frac{1}{\omega m}\right)^2} \qquad 13$$

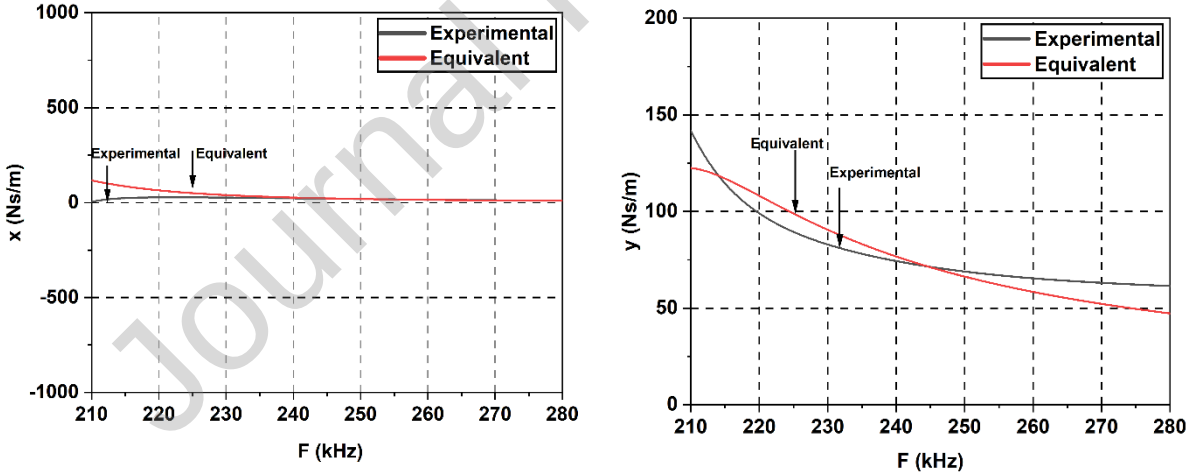



*Figure 37 Comparison of experimental and equivalent plots of x vs f and y vs f*

Table 4 shows the average values of m, c, and k for different sensor configurations for a typical example of cube at 3m height of impact force under 150 $^{o}$C temperature. Based on the match between the experimental and analytical impedance plots presented in Figure 36 (b), a series combination of spring-mass damping (*k-m-c*) system in the frequency range of 210–280 kHz was chosen as an equivalent system for the examination of ESPs.

Journal Pre-proof

Figure 37 shows the reasonable agreement between the experimental and equivalent plot of *x* and *y* extracted for the JKTPS at 150 $^{o}$C temperature.

*Table 4 Average values of m-c-k for different sensors configuration at 3 7m impact height under 150 $^{o}$C*

| **Sensors Configuration** | **Height** | **Temperature** | **m (kg)** | **c (Ns/m)** | **k (N/m)** |
|---|---|---|---|---|---|
| JKTPS | 3m | 150 $^{o}$C | $2.69x10^{-2}$ | $8.516x10^{3}$ | $4.08x10^{4}$ |
| SBPS | | | $2.75x10^{-2}$ | $6.28x10^{-2}$ | $4.23x10^{4}$ |
| NBPS | | | $2.77x10^{-2}$ | $9.26x10^{-2}$ | $4.23x10^{4}$ |

Figure 38 depicts the variation of the Equivalent stiffness value with the number of impacts for various sensor configurations under150 $^{o}$C temperature. Equivalent stiffness values follow the same pattern for all sensor configurations as the number of impacts increases. Except for pristine to first impact, the equivalent stiffness value decreased as the number of impacts increased up to failure, the decrease of the equivalent stiffness value with the increasing impacts clearly indicates the increase in damage in the sample. The value of the increasing damage was due to the loss of its strength and the value of increasing stiffness for first impact was due to the PZT's high temperature sensitivity.

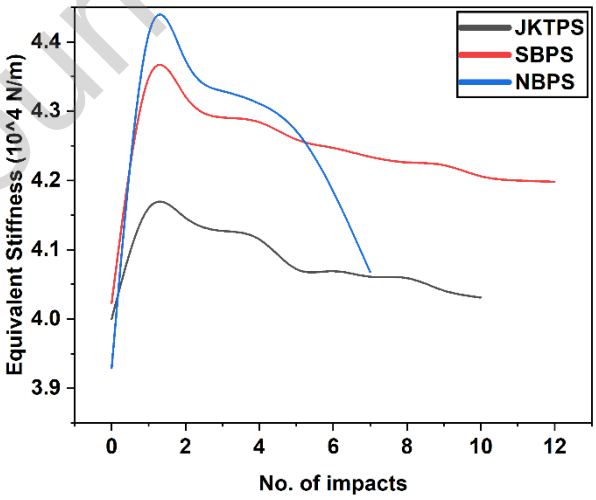



*Figure 38 Variation of Equivalent Stiffness at 150 $^{o}$C for a) JKTPS b) SBPS c) NBPS*

Journal Pre-proof

## 4 Remaining life estimation using equivalent structural parameters

For establishing remaining life model, further use the equivalent stiffness parameter i.e. already calculated in the previous section. A non-dimensional stiffness loss parameter *Δs* was used to quantify the changes in the stiffness caused by incremental damage. *Δs* was defined as given below

$$\Delta s = \left|\frac{\Delta k}{k}\right| \quad 14$$

Where, $\Delta k$ change in equivalent stiffness compared to $k$ and $k$ is the original equivalent stiffness in pristine stage. As indicated below, another non-dimensional parameter, $L$, was included to denote the structure's remaining life.

$$L = 1 - \frac{N}{N_0} \quad 15$$

Where $N_0$ signifies the total number of impacts applied till failure and $N$ denotes the number of impacts applied until a specific stage. Figure 39 shows the variation of the *L in function of Δs* for the JKTPS, NBPS and SBPS. Each sensors has shown the similar pattern of the *L in function of Δs* graph i.e with the increase in the value of *Δs*, remaining life of structure get reduce.

$$L = 440.96978\,(\Delta s)^2 - 47.24186\,(\Delta s) + 1.22576 \quad 16$$

$$L = 517.18207\,(\Delta s)^2 - 90.72141\,(\Delta s) + 3.96587 \quad 17$$

$$L = -72.63299\,(\Delta s)^2 + 4.51\,(\Delta s) + 0.52883 \quad 18$$

A empirical equation was also develop based on the *L in function of Δs* graph for predicting the remaining life of structure under this experiment in terms of *Δs* using best fit curve [60], [61]. Empirical equation for each sensors for JKTPS, SBPS and NBPS are mentioned from equation 16 to 18 respectively.

Journal Pre-proof

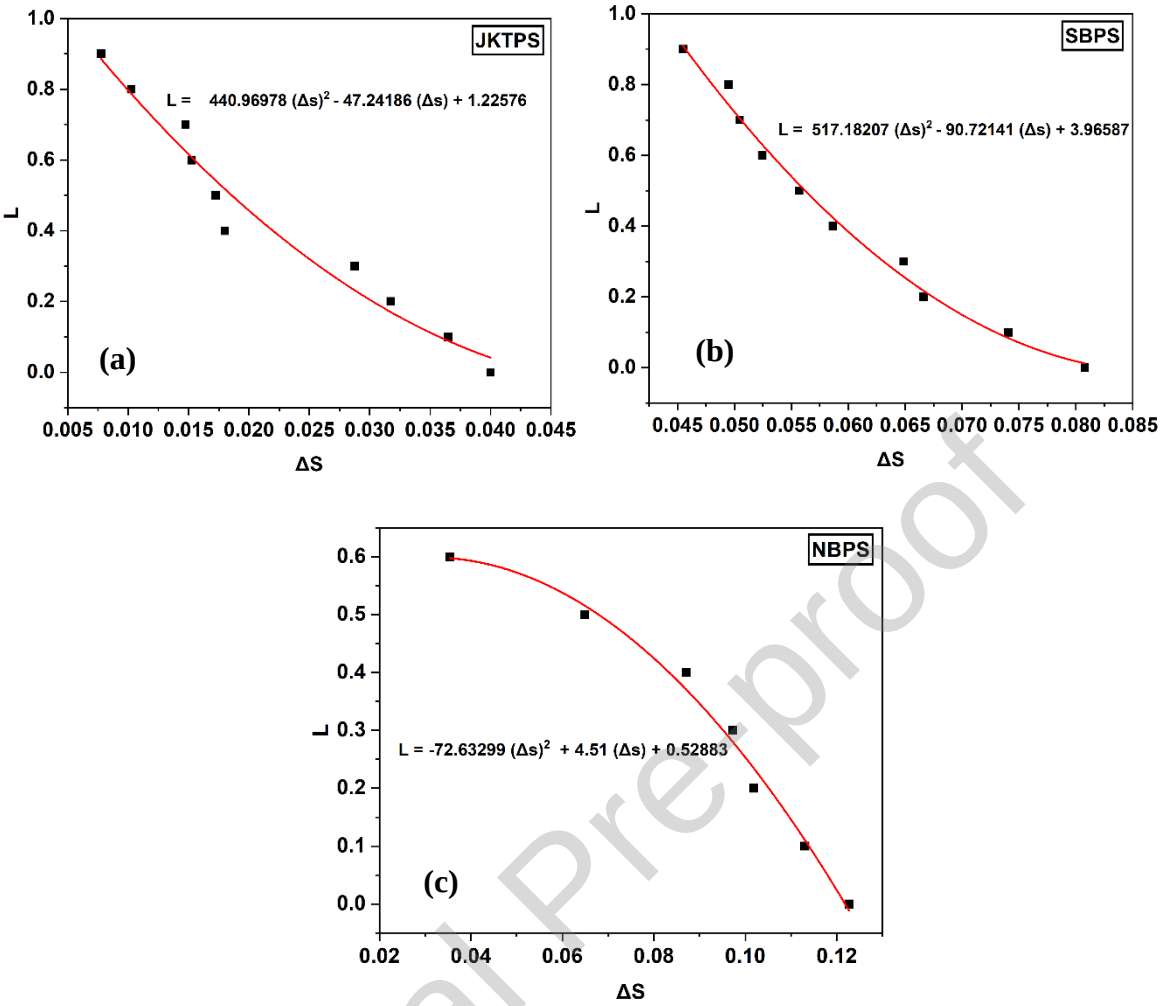



*Figure 39 Variation of L vs Δs for a) JKTPS b) SBPS c) NBPS*

## 5 Conclusion

This study uses the EMI technique to demonstrate the feasibility of three different types of piezo sensor configurations for detecting damage in concrete under the effect of impact loading at varying temperatures. The conclusions discussed here are based on the above-mentioned experimental results. On the basis of the evaluation of the RMSD and MAPD value both are the reliable statistically tool for calculating the incipient and progressive damage in concrete under the effect of impact loading at varying temperature. Due to its lack of direct contact with the host structure, the NBPS recorded lower RMSD and MAPD values than the SBPS and JKTPS for all temperature cases. Both the RMSD and MAPD indices followed a distinct pattern at higher temperatures of 150 $^{o}$C for all sensor configurations, whereas MAPD values followed a more consistent pattern than RMSD values.

Journal Pre-proof

NFSI has been followed decreasing pattern for the increasing value of impact and as well as increasing value of the temperature also this is due to the loss of the strength of the concrete cube. The resonant frequency and peak conductance exhibited an inverse relationship with increasing impact load, and this relationship was maintained at higher temperatures. The extracted equivalent stiffness with increasing impact number clearly indicates damage propagation in concrete sample for different sensor configuration and also showed satisfactory agreement between the experimental and equivalent plot of x and y. ESPs were used to successfully develop an empirical model for predicting the remaining life of structures. This model was created by analyzing the obtained data under the best fit curve, and it shows that the remaining life of the structure decreases as the loss of equivalent stiffness increases. All three sensor configurations, JKTPS, SBPS, and NBPS, perform well in predicting damage when impact and temperature are combined. Despite the fact that piezo sensors have a very high sensing behavior against temperature, the JKTPS performed well with the fewest false alarms when compared to the SBPS and NBPS. Also confirmed that the sensors' reusable tendency after the cube's complete failure, and we discovered that the majority of NBPS and JKTPS are in operational mode. The authors' future work will examine the feasibility and reliability of the EMI technique for structural elements such as beams and columns when subjected to the combined effects of impact and temperature. Furthermore, the research will be extended to correlate the damage state, such as the length and amount of crack, by calibrating the observed signature with the measured length and amount of crack in order to improve the capability of the electromechanical impedance (EMI) technique.

**Funding Statement**

Authors would like to express their sincere gratitude to Delhi Technological University, Delhi, India for providing research facilities and funding to conduct the research work.

**Acknowledgments:**

The authors acknowledge the Department of Civil Engineering and Delhi Technological University, Delhi for providing the research facilities for conducting the tests and financial support for the first author under institutional fellowship for Ph.D.

**Conflict of Interest:**

The authors declare that there are no conflicts of interest.

Journal Pre-proof

**References**


[1] R. Tawie and H. K. Lee, "Monitoring the strength development in concrete by EMI sensing technique," *Constr. Build. Mater.*, vol. 24, no. 9, pp. 1746–1753, 2010, doi: 10.1016/j.conbuildmat.2010.02.014.

[2] P. Negi, T. Chakraborty, N. Kaur, and S. Bhalla, "Investigations on effectiveness of embedded PZT patches at varying orientations for monitoring concrete hydration using EMI technique," *Constr. Build. Mater.*, vol. 169, pp. 489–498, 2018, doi: 10.1016/j.conbuildmat.2018.03.006.

[3] S. W. Shin and T. K. Oh, "Application of electro-mechanical impedance sensing technique for online monitoring of strength development in concrete using smart PZT patches," *Constr. Build. Mater.*, vol. 23, no. 2, pp. 1185–1188, 2009, doi: 10.1016/j.conbuildmat.2008.02.017.

[4] C. Zuo, X. Feng, and J. Zhou, "A three-dimensional model of the effective electromechanical impedance for an embedded PZT transducer," *Math. Probl. Eng.*, vol. 2013, no. December, 2013, doi: 10.1155/2013/218026.

[5] P. Liu, W. Wang, Y. Chen, X. Feng, and L. Miao, "Concrete damage diagnosis using electromechanical impedance technique," *Constr. Build. Mater.*, vol. 136, pp. 450–455, 2017, doi: 10.1016/j.conbuildmat.2016.12.173.

[6] D. Wang and H. Zhu, "Monitoring of the strength gain of concrete using embedded PZT impedance transducer," *Constr. Build. Mater.*, vol. 25, no. 9, pp. 3703–3708, 2011, doi: 10.1016/j.conbuildmat.2011.04.020.

[7] N. Banthia, S. Mindess, A. Bentur, and M. Pigeon, "Impact testing of concrete using a drop-weight impact machine," *Exp. Mech.*, vol. 29, no. 1, pp. 63–69, 1989, doi: 10.1007/BF02327783.

[8] P. H. Bischoff and S. H. Perry, "Impact Behavior of Plain Concrete Loaded in Uniaxial Compression," *J. Eng. Mech.*, vol. 121, no. 6, pp. 685–693, 1995, doi: 10.1061/(asce)0733-9399(1995)121:6(685).

[9] J. H. Haido, A. A. Abdul-Razzak, M. M. Al-Tayeb, B. H. Abu Bakar, S. T. Yousif, and B. A. Tayeh, "Dynamic response of reinforced concrete members incorporating steel fibers with different aspect ratios," *Adv. Concr. Constr.*, vol. 11, no. 2, pp. 89–98, 2021, doi: 10.12989/acc.2021.11.2.089.

[10] S. Bright Singh and M. Murugan, "Performance of carbon fibre-reinforced pervious concrete (CFRPC) subjected to static, cyclic and impact loads," *Int. J. Pavement Eng.*, vol. 0, no. 0, pp. 1–16, 2021, doi: 10.1080/10298436.2021.1883017.

[11] M. Loganaganandan, G. Murali, and M. Parasuraman, "Experimental Study on GFRP Strips Strengthened New Two Stage Concrete Slabs under Falling Mass Collisions," 2020, doi: 10.1007/s12205-020-0331-8.

[12] P. Satya, T. Asai, M. Teshigawara, Y. Hibino, and I. Maruyama, "Impact of drying on structural performance of reinforced concrete beam with slab," *Materials (Basel).*, vol. 14, no. 8, 2021, doi:

10.3390/ma14081887.

[13] R. Smith-Gillis, R. Lopez-Anido, T. S. Rushing, and E. N. Landis, "Development of thermoplastic composite reinforced ultra-high-performance concrete panels for impact resistance," *Materials (Basel).*, vol. 14, no. 10, pp. 1–16, 2021, doi: 10.3390/ma14102490.

[14] A. L. Mina, M. F. Petrou, and K. G. Trezos, "Resistance of an optimized ultra-high performance fiber reinforced concrete to projectile impact," *Buildings*, vol. 11, no. 2, pp. 1–18, 2021, doi: 10.3390/buildings11020063.

[15] S. J. Yoo, T. F. Yuan, S. H. Hong, and Y. S. Yoon, "Effect of strengthening methods on two-way slab under low-velocity impact loading," *Materials (Basel).*, vol. 13, no. 24, pp. 1–16, 2020, doi: 10.3390/ma13245603.

[16] Y. Al, Y. Temsah, O. Baalbaki, A. Jahami, and M. Darwiche, "Experimental investigation on the effect of impact loading on behavior of post- tensioned concrete slabs," *J. Build. Eng.*, vol. 31, no. September 2019, p. 101207, 2020, doi: 10.1016/j.jobe.2020.101207.

[17] A. Bentur, S. Mindess, and N. Banthia, "The behaviour of concrete under impact loading: Experimental procedures and method of analysis," *Mater. Struct.*, vol. 19, no. 5, pp. 371–378, 1986, doi: 10.1007/BF02472127.

[18] S. M. Soleimani, N. Banthia, and S. Mindess, "Behavior of RC beams under impact loading: Some new findings," *Proc. 6th Int. Conf. Fract. Mech. Concr. Concr. Struct.*, vol. 2, no. Figure 1, pp. 867–874, 2007.

[19] T. Pham and H. Hao, "Effect of the Plastic Hinge and Boundary Conditions on the Impact Behavior of Effect of the Plastic Hinge and Boundary Conditions on the Impact Behavior of Reinforced Concrete Beams," no. December, 2016, doi: 10.1016/j.ijimpeng.2016.12.005.

[20] Y. Fu *et al.*, "Investigating the failure behaviors of RC beams without stirrups under impact loading," *Int. J. Impact Eng.*, vol. 137, no. November 2019, p. 103432, 2020, doi: 10.1016/j.ijimpeng.2019.103432.

[21] L. Jin, R. Zhang, G. Dou, J. Xu, and X. Du, "Experimental and numerical study of reinforced concrete beams with steel fibers subjected to impact loading," *Int. J. Damage Mech.*, vol. 27, no. 7, pp. 1058–1083, 2018, doi: 10.1177/1056789517721616.

[22] A. Savva, P. Manita, and K. K. Sideris, "Influence of elevated temperatures on the mechanical properties of blended cement concretes prepared with limestone and siliceous aggregates," vol. 27, pp. 239–248, 2005, doi: 10.1016/j.cemconcomp.2004.02.013.

[23] H. Yang, H. Zhao, and F. Liu, "Residual cube strength of coarse RCA concrete after exposure to elevated temperatures," *Fire Mater.*, vol. 42, no. 4, pp. 424–435, 2018, doi: 10.1002/fam.2508.

[24] H. W. Ye *et al.*, "Research on fire resistance of ultra-high-performance concrete," *Adv. Mater. Sci. Eng.*, vol. 2012, no. 2, 2012, doi: 10.1155/2012/530948.

[25] M. Potha Raju, K. Srinivasa Rao, and P. S. N. Raju, "Compressive strength of heated high-strength

concrete," *Mag. Concr. Res.*, vol. 59, no. 2, pp. 79–85, 2007, doi: 10.1680/macr.2007.59.2.79.

[26] A. Powezka, J. Szulej, and P. Ogrodnik, "Effect of high temperatures on the impact strength of concrete based on recycled aggregate made of heat-resistant cullet," *Materials (Basel).*, vol. 13, no. 2, pp. 1–21, 2020, doi: 10.3390/ma13020465.

[27] M. S. Khan, J. Prasad, and H. Abbas, "Effect of High Temperature on High-Volume Fly Ash Concrete," *Arab. J. Sci. Eng.*, vol. 38, no. 6, pp. 1369–1378, 2013, doi: 10.1007/s13369-013-0606-1.

[28] A. Lau and M. Anson, "Effect of high temperatures on high performance steel fibre reinforced concrete," vol. 36, pp. 1698–1707, 2006, doi: 10.1016/j.cemconres.2006.03.024.

[29] L. Larson, R. Payton, and D. Blank, "United states department of labor mine safety and health administration," 1999.

[30] Y. Xia, H. Hao, G. Zanardo, and A. Deeks, "Long term vibration monitoring of an RC slab: Temperature and humidity effect," *Eng. Struct.*, vol. 28, no. 3, pp. 441–452, 2006, doi: 10.1016/j.engstruct.2005.09.001.

[31] Y. Wang, G. Yuan, Z. Huang, J. Lyv, Z. Q. Li, and T. yan Wang, "Experimental study on the fire behaviour of reinforced concrete slabs under combined uni-axial in-plane and out-of-plane loads," *Eng. Struct.*, vol. 128, pp. 316–332, 2016, doi: 10.1016/j.engstruct.2016.09.054.

[32] Y. Jiao, H. Liu, X. Wang, Y. Zhang, G. Luo, and Y. Gong, "Temperature effect on mechanical properties and damage identification of concrete structure," *Adv. Mater. Sci. Eng.*, vol. 2014, 2014, doi: 10.1155/2014/191360.

[33] V. K. R. Kodur, "Innovative strategies for enhancing fire performance of high-strength concrete structures," *Adv. Struct. Eng.*, vol. 21, no. 11, pp. 1723–1732, 2018, doi: 10.1177/1369433218754335.

[34] W. Khaliq and Taimur, "Mechanical and physical response of recycled aggregates high-strength concrete at elevated temperatures," *Fire Saf. J.*, vol. 96, no. May 2017, pp. 203–214, 2018, doi: 10.1016/j.firesaf.2018.01.009.

[35] L. Jin, J. Bai, R. Zhang, L. Li, and X. Du, "Effect of elevated temperature on thelow-velocity impact performances of reinforced concrete slabs," *Int. J. Impact Eng.*, vol. 149, no. November 2020, p. 103797, 2021, doi: 10.1016/j.ijimpeng.2020.103797.

[36] L. Jin, R. Zhang, G. Dou, and X. Du, "Fire resistance of steel fi ber reinforced concrete beams after low-velocity impact loading," *Fire Saf. J.*, vol. 98, no. November 2017, pp. 24–37, 2018, doi: 10.1016/j.firesaf.2018.04.003.

[37] C. K. Soh and S. Bhalla, "Calibration of piezo-impedance transducers for strength prediction and damage assessment of concrete," *Smart Mater. Struct.*, vol. 14, no. 4, pp. 671–684, 2005, doi: 10.1088/0964-1726/14/4/026.

[38] Y. F. Su, G. Han, A. Amran, T. Nantung, and N. Lu, "Instantaneous monitoring the early age properties of cementitious materials using PZT-based electromechanical impedance (EMI) technique," *Constr. Build. Mater.*, vol. 225, pp. 340–347, 2019, doi: 10.1016/j.conbuildmat.2019.07.164.

[39] A. Narayanan, A. Kocherla, and K. V. L. Subramaniam, “Understanding the coupled electromechanical response of a PZT patch attached to concrete: Influence of substrate size,” *Meas. J. Int. Meas. Confed.*, vol. 124, no. March, pp. 505–514, 2018, doi: 10.1016/j.measurement.2018.04.055.

[40] S. W. Shin, A. R. Qureshi, J. Y. Lee, and C. B. Yun, “Piezoelectric sensor based nondestructive active monitoring of strength gain in concrete,” *Smart Mater. Struct.*, vol. 17, no. 5, 2008, doi: 10.1088/0964-1726/17/5/055002.

[41] T. Jothi Saravanan, K. Balamonica, C. Bharathi Priya, N. Gopalakrishnan, and S. G. N. Murthy, “Piezoelectric EMI–Based Monitoring of Early Strength Gain in Concrete and Damage Detection in Structural Components,” *J. Infrastruct. Syst.*, vol. 23, no. 4, p. 04017029, 2017, doi: 10.1061/(asce)is.1943-555x.0000386.

[42] V. G. M. Annamdas, Y. Yang, and C. K. Soh, “Impedance based Concrete Monitoring using Embedded PZT Sensors,” *Int. J. Comput. Civ. Struct. Eng.*, vol. 1, no. 3, pp. 414–424, 2010, doi: 10.6088/ijcser.00202010033.

[43] V. G. M. Annamdas and P. Rizzo, “Monitoring concrete by means of embedded sensors and electromechanical impedance technique,” *Sensors Smart Struct. Technol. Civil, Mech. Aerosp. Syst. 2010*, vol. 7647, p. 76471Z, 2010, doi: 10.1117/12.847687.

[44] Q. Kong, S. Hou, Q. Ji, Y. L. Mo, and G. Song, “Very early age concrete hydration characterization monitoring using piezoceramic based smart aggregates,” *Smart Mater. Struct.*, vol. 22, no. 8, 2013, doi: 10.1088/0964-1726/22/8/085025.

[45] T. J. Saravanan, K. Balamonica, C. B. Priya, A. L. Reddy, and N. Gopalakrishnan, “Comparative performance of various smart aggregates during strength gain and damage states of concrete,” *Smart Mater. Struct.*, vol. 24, no. 8, p. 85016, 2015, doi: 10.1088/0964-1726/24/8/085016.

[46] D. Wang, H. Song, and H. Zhu, “Embedded 3D electromechanical impedance model for strength monitoring of concrete using a PZT transducer,” *Smart Mater. Struct.*, vol. 23, no. 11, 2014, doi: 10.1088/0964-1726/23/11/115019.

[47] X. Lu, Y. Y. Lim, and C. K. Soh, “A novel electromechanical impedance–based model for strength development monitoring of cementitious materials,” *Struct. Heal. Monit.*, vol. 17, no. 4, pp. 902–918, 2018, doi: 10.1177/1475921717725028.

[48] Y. Yang, B. S. Divsholi, and C. K. Soh, “A reusable PZT transducer for monitoring initial hydration and structural health of concrete,” *Sensors*, vol. 10, no. 5, pp. 5193–5208, 2010, doi: 10.3390/s100505193.

[49] S. Moharana and S. Bhalla, “Development and evaluation of an external reusable piezo-based concrete hydration-monitoring sensor,” *J. Intell. Mater. Syst. Struct.*, vol. 30, no. 18–19, pp. 2770–2788, 2019, doi: 10.1177/1045389X19873414.

[50] C. Bharathi Priya, T. Jothi Saravanan, K. Balamonica, N. Gopalakrishnan, and A. R. M. Rao, “EMI based monitoring of early-age characteristics of concrete and comparison of serial/parallel multi-sensing technique,” *Constr. Build. Mater.*, vol. 191, pp. 1268–1284, 2018, doi:

10.1016/j.conbuildmat.2018.10.079.

[51] C. P. Providakis, E. V. Liarakos, and E. Kampianakis, "Nondestructive Wireless Monitoring of Early-Age Concrete Strength Gain Using an Innovative Electromechanical Impedance Sensing System," *Smart Mater. Res.*, vol. 2013, pp. 1–10, 2013, doi: 10.1155/2013/932568.

[52] X. Lu, Y. Y. Lim, I. Izadgoshasb, and C. K. Soh, "Strength development monitoring and dynamic modulus assessment of cementitious materials using EMI-Miniature Prism based technique," *Struct. Heal. Monit.*, vol. 19, no. 2, pp. 373–389, 2020, doi: 10.1177/1475921719848087.

[53] G. Park, K. Kabeya, H. H. Cudney, and D. J. Inman, "Removing Effects of Temperature Changes from Piezoelectric Impedance-Based Qualitative Health Monitoring," *Smart Struct. Mater. 1998 Sens. Phenom. Meas. Instrum. Smart Struct. Mater.*, vol. 3330, no. March, pp. 103–114, 1998, doi: 10.1117/12.316963.

[54] S. Bhalla, A. S. K. Naidu, and C. K. Soh, "Influence of structure-actuator interactions and temperature on piezoelectric mechatronic signatures for NDE," *Smart Mater. Struct. Syst.*, vol. 5062, p. 263, 2003, doi: 10.1117/12.514762.

[55] T. Bansal, V. Talakokula, and K. Mathiyazhagan, "Equivalent structural parameters based non-destructive prediction of sustainable concrete strength using machine learning models via piezo sensor," *Meas. J. Int. Meas. Confed.*, vol. 187, no. September 2021, p. 110202, 2022, doi: 10.1016/j.measurement.2021.110202.

[56] T. Bansal and V. Talakokula, "Deterioration of structural parameters due to corrosion in prestressed concrete identified by smart probe-based piezo sensor," *Eng. Res. Express*, vol. 3, no. 1, 2021, doi: 10.1088/2631-8695/abded9.

[57] Keysight (Agilent) Technologies, "E4980A Precision LCR Meter 20 Hz to 2 MHz," pp. 1–38, 2014, doi: https://www.keysight.com/in/en/assets/7018-01355/data-sheets/5989-4435.pdf.

[58] P. Negi, R. Chhabra, N. Kaur, and S. Bhalla, "Health monitoring of reinforced concrete structures under impact using multiple piezo-based configurations," *Constr. Build. Mater.*, vol. 222, pp. 371–389, 2019, doi: 10.1016/j.conbuildmat.2019.06.100.

[59] N. K. S. Bhalla, S. Moharana, V. Talakokula, "Piezoelectric Materials Applications in SHM, Energy Harvesting & Biomechanics," *Athena Acad. Wiley, London*, vol. 44, no. May, 2017.

[60] S. Bhalla, P. A. Vittal, and M. Veljkovic, "Piezo-impedance transducers for residual fatigue life assessment of bolted steel joints," *Struct. Heal. Monit.*, vol. 11, no. 6, pp. 733–750, 2012, doi: 10.1177/1475921712458708.

[61] A. Dixit and S. Bhalla, "Prognosis of fatigue and impact induced damage in concrete using embedded piezo-transducers," *Sensors Actuators, A Phys.*, vol. 274, pp. 116–131, 2018, doi: 10.1016/j.sna.2018.03.005.

Journal Pre-proof

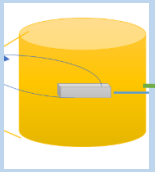
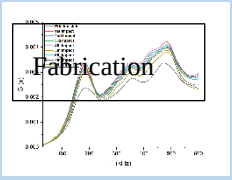
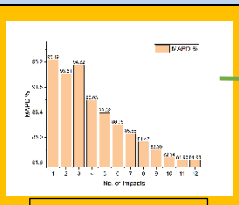
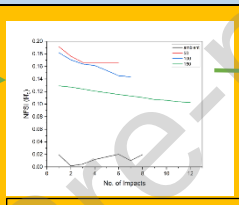
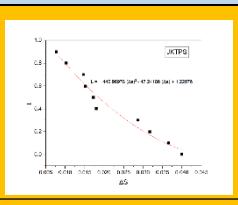
# Authors' biographies

**Mr. Indrajeet Singh** is currently a Research Scholar in the Department of Civil Engineering, Delhi Technological University, Delhi. Mr. Singh completed his Bachelor of Technology in Civil Engineering from the Inderprastha Engineering College, Ghaziabad, in 2013 and compleated his Masters in Structural Engineering from MDU Rohtak, in 2017, His main areas of interest are smart structures, structural health monitoring, electromechanical impedance (EMI) technique, energy-efficient buildings, finite element modelling and geopolymer concrete.

**Dr. Nirendra Dev** is currently Professor at the Department of Civil Engineering, Delhi Technological University, Delhi. His main areas of interest are smart structures, Structural Engineering., Structural Dynamics, Finite Element Methods, Disaster Management, High Strength Nano Composites, and Concrete Technology. Dr. Dev has published over 20 papers in international journals, and over 40 in international/ national refereed conferences/ workshops. Dr. Dev has three patents and has applied for two more.

**Dr. Shilpa Pal** is currently Associate Professor at the Department of Civil Engineering, Delhi Technological University, Delhi. Her main areas of interest are 2-D & 3-D Seismic Analysis of Dams and Seismic Slope Stability Analysis, Damage detection in buildings, Biofortified Concrete, Green Materials in Construction. Dr. Pal has published over 35 papers in international and national journals, and over 60 in international/ national refereed conferences/ workshops.

Journal Pre-proof

| | |
|---|---|
| **Title** | **Damage assessment of concrete under the combined effect of impact and Temperature using different piezo sensors** |
| **Methodology** |   Fabrication → Installation → Testing → Data Extraction |
| **Findings** |  RMSD/MAPD →  NFSI/Peak conductance →  Remaining life |
| **Conclusion** | All three sensor configurations, JKTPS, SBPS, and NBPS, perform well in predicting damage when impact and temperature are combined. Despite the fact that piezo sensors have a very high sensing behaviour against temperature, the JKTPS performed well with the fewest false alarms when compared to the NBPS and SBPS |

Journal Pre-proof

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.

**Highlights of review**

1. Damage identification under the effect of impact loading at varying temperatures.
2. Study the behavior of different piezo sensors configuration.
3. Establish the remaining life model for the specimen.
4. Extraction of equivalent stiffness parameter as useful in damage detection in SHM.

2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) | 978-1-6654-1120-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAECT54875.2022.9808083

# Implementation of Inductive Wireless Power Transmission system for Battery Charging applications

Abhishek Shakya
Department of Electrical Engineering
Delhi Technological University Delhi, India
abhishekshakya101@gmail.com

Mayank Kumar
Department of Electrical Engineering
Delhi Technological University Delhi, India
mayankkumar@dtu.ac.in

***Abstract*— Wireless power transfer systems are gaining popularity in the modern period due to their wide range of applications, such as some devices which can be charged by wireless charging like cell phones, laptops, electric vehicles, underwater vehicles, and implants. Wireless power transfer technology known as inductive power transfer (IPT) is a sort of technology which is known as wireless power transfer technology. The goal of this work is to construct an electric vehicle wireless power transfer charger based on IPT. A series–series (SS) compensated IPT system has been developed for this purpose. To begin, the SS compensated IPT system's designing formulas are derived. Based on the derived formulas of design, values are calculated for deriving a 2 KW wireless power charger and the same is implemented using Simulink/MATLAB.**

***Keywords—Inductive Power Transfer (IPT), Wireless Charging System (WCS), Wireless Power Transfer (WPT), Electric Vehicle (EV), Coupling.***

## I. INTRODUCTION

There has been a growing desire for the advancement of EVs that are battery-powered to decrease air pollution and to use as fossil-fuel replacement since a century ago [1]. The battery charger is critical to the advancement of electric vehicles (EVs). EV battery chargers are essential for high-efficiency and high-power density applications. As a result, chargers having high quality, high efficiency and power density are essential for sustaining and extending battery lifetime while also minimizing charging time [2][3].

The rapid growing use of combustion engines (mostly in vehicles) is increasing as the population grows. Resulting, the usage of fossil fuels is expanding bit by bit, putting a strain on fossil fuel resources while also contributing to pollution and global warming. The adoption of EVs is a viable solution to this challenge. The use of wireless charging technology can make charging these EVs more suitable. Wireless charging system (WCS) based on Inductive power transmission for EVs is more suitable, adaptable, dependable, and safer than conductive charging, and also it removes the need for any worker.

The two coils are required for charging system, one of which is put in the parking garage (primary coil/transmitter coil) and the other of which is located inside the EV (secondary coil/receiver coil). In order to be effective, the EV side coil should be tiny and light in weight. The compensation approach is a crucial part of constructing an inductive base WCS because it boosts the power transfer ability and aids in establishing smooth switching of power electronics equipment. The main requirements for a battery charger in EV applications are high efficiency, minimal size, and high dependability [5][6].

EV owners are interested in the WPT technology, which can eliminate any charging hassles. Charging becomes the simplest operation when energy is transferred wirelessly to the EV. The drivers of car in case of stationary WPT system simply park their cars at the charging station. In the case of a dynamic WPT system, this means that the EV can be powered while travelling; the EV can run indefinitely without stopping. In addition, as compared to EVs that use conductive charging, the battery capacity of EVs with wireless charging could be reduced by 20% or less [7][9].

## II. INDUCTIVE COUPLING

Inductive coupling is described as the coupling of two LC circuits with the same resonant frequency. It works by utilizing magnetic field induction, which is a natural byproduct motion of current through wire. For example, a varying magnetic field can be produced by alternating current in a primary coil connected to a source which induces a voltage across the terminals at the receiver side of the secondary coil. In inductive coupling, the primary and secondary coils are two separate coils. Each of them was connected wirelessly, and inductive coupling has become a common technology for transferring electricity without wires due to its simplicity, convenience, and safety. Various types of electrical devices have previously been created using this technological application. As a result, it has been effectively

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:42:59 UTC from IEEE Xplore. Restrictions apply.

applied to a variety of items, including EV, medical implants, a toothbrush, and a charging pad for a phone or laptop [8].

When two coils are steadily separated from each other, and if their alignment is not ideal in the case of inductive coupling, leads to gradually declination of power transfer. These kinds of issues usually emerge when it isn't used carefully. It works best when the charging node of the device and power receiving node are in close proximity, usually less than a coil diameter; for example, the range can be measured in centimeters, and the direction of charging must be aligned [4].

## III. System Modelling

It is clearly observed from Fig. 1 that the WCS has two parts, primary coil and the secondary coil. The primary side has DC supply and inverter section and the receiver side has rectifier and load part. The primary is supplied with grid power and its setup is situated at the charging stations/parking area of vehicles, whereas the setup of receiver is installed inside the EVs. The configuration must be operated at the resonating frequency to obtain the transfer of power from the primary to secondary coil to its maximum limit.

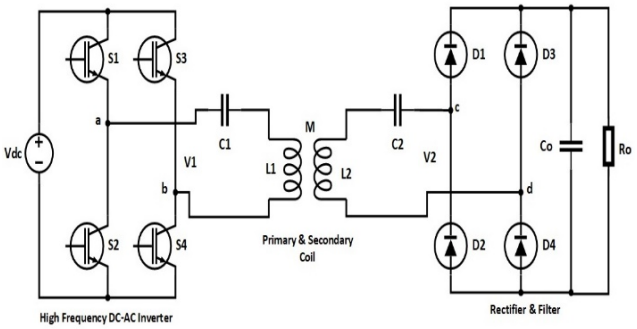



Fig. 1. Circuit Diagram of IPT system

To compensate the losses and to achieve the resonance condition in the primary and secondary coil, capacitor C1 and C2 are used.

L1 and L2 denotes the self-inductances of primary and secondary coil respectively, C1 and C2 respectively denotes the series compensation capacitances of primary and secondary coil.

Most of the IPT systems needs flow of power in only one direction, which can be achieved by using an AC source at the input side, whereas the AC source is comprising of a high frequency converter at the primary side, and a simple diode bridge at the secondary side. The series compensation type of circuit is shown in above fig. 1, having simple design and low cost.

Fig. 2 displays the equivalent circuit diagram of IPT system which helps in deriving the formulas for electrical parameters designing of inductive WPT system.

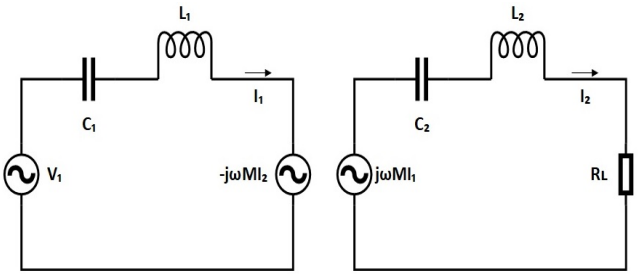



Fig. 2. Equivalent Circuit Diagram of IPT system

Lithium-ion batteries are used in EVs which are available in the range of 300-600 V.

Taking output voltage as 450 V and input supply as 240 V, the load resistance will be:

$$Ro = \frac{Vo^2}{Po} = \frac{450 \times 450}{2000} = 101.25\,\Omega \quad (1)$$

Equivalent resistance as seen from terminals CD in fig 1:

$$R_L = \frac{8}{\pi^2} \times \frac{Vo^2}{Po} = 82\,\Omega \quad (2)$$

The secondary side rms voltage can be expressed as

$$V2\text{rms} = \frac{2\sqrt{2}Vo}{\pi} = \frac{2\sqrt{2} \times 450}{\pi} = 405.35\,V \quad (3)$$

Self-inductance of secondary side coil is as follows:

$$L2 = Qs \times \frac{R_L}{\omega o} = 4 \times \frac{82}{2\pi \times 85000}$$

$$= 6.14 \times 10^{-4} H \quad (4)$$

Where, Qs is the quality factor of secondary side.

I2rms represent the rms current of secondary side and can be expressed as

$$I2rms = \frac{V2rms}{R_L} = \frac{405.35}{82} = 4.94\,A \quad (5)$$

I1rms represent the rms current of primary side and can be expressed as

$$I1rms = \frac{Po}{V1rms} = \frac{2000}{240} = 8.33\,A \quad (6)$$

Mutual inductance between primary and secondary coil can be expressed as

$$M = I2rms \times \frac{R_L}{I1rms \times \omega o}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:42:59 UTC from IEEE Xplore. Restrictions apply.

$$= 4.94 \times \frac{82}{8.33 \times 2\pi \times 85000}$$

$$- 9.1 \times 10^{-5} H \quad (7)$$

The coefficient of coupling can be calculated as below:

$$Kc = \frac{1}{Qs} \times \sqrt{1 - \left(\frac{1}{4 \times Qs^2}\right)}$$

$$- \frac{1}{4} \times \sqrt{1 - \left(\frac{1}{4 \times 4^2}\right)} - 0.248 \quad (8)$$

Self-inductance of primary side coil is as follows:

$$L1 = \frac{M^2}{L2 \times Kc^2} = \frac{(9.1 \times 10^{-5})^2}{6.14 \times 10^{-4} \times (0.248)^2}$$

$$= 219.28 \times 10^{-6} H \quad (9)$$

Series compensation capacitance of primary side coil can be expressed as

$$C1 = \frac{1}{L1 \times \omega o^2} = \frac{1}{219.28 \times 10^{-6} \times (2\pi \times 85000)^2}$$

$$= 1.6 \times 10^{-8} F \quad (10)$$

Series compensation capacitance of secondary side coil can be calculated as

$$C2 = \frac{1}{L2 \times \omega o^2} = \frac{1}{6.14 \times 10^{-4} \times (2\pi \times 85000)^2}$$

$$= 5.71 \times 10^{-9} F \quad (11)$$

## IV. Simulation Results

The WPT system as shown in Fig. 1 was simulated in MATLAB/Simulink with the mentioned/calculated parameters in the below Table I. Firstly, the output current, voltage and power is observed with 100 Ω load across capacitor Co without adding boost converter on load side.

Again, the output voltage, current and power is analyzed with 100 Ω load with adding a boost converter on load side. The output of the designed WPT system can be controlled by taking help of boost converter and by varying the duty cycle, variation in output voltage can be observed.

It is observed from the simulation that was carried out for transmission of power wirelessly with 240 V dc input supply. This input supply is inverted using high frequency DC-AC inverter then fed to the primary/transmitting coil. By creating resonance between primary and secondary coil, the high frequency AC is supplied to the secondary side by the help of mutual induction. The AC obtained from secondary coil is first rectified before supplied to the load (battery) of 100 Ω resistance.

The high frequency DC-AC inverter formed with IGBT inverted the dc supply into high frequency AC and then this inverted signal/high frequency AC is fed to primary side coil. The two gate pulses are provided to high frequency inverter part for its operation to produce high frequency AC. To get the required high DC voltage at the output, DC-DC boost converter is added which helps to obtain the required output voltage by varying/changing the duty cycle. In boost converter MOSFET is used as a switch with 25KHz as switching frequency.

TABLE I. System Parameters

| **Parameters** | **Designator** | **Value** |
|---|---|---|
| Input Voltage | V1 | 240 V |
| Load resistance | Ro | 100 Ω |
| Output Voltage | Vo | 135 V, 450 V |
| Switching Frequency | Fo | 85 KHz |
| Inductance of primary coil | L1 | $219.28\ \mu H$ |
| Inductance of secondary coil | L2 | $6.1 \times 10^{-4} H$ |
| Mutual Inductance | M | $9.1 \times 10^{-5} H$ |
| Coupling Coefficient | $Kc$ | 0.248 |
| Primary Coil Capacitance | C1 | $1.6 \times 10^{-8} F$ |
| Secondary Coil Capacitance | C2 | $5.7 \times 10^{-9} F$ |
| Output Power | Po | 2 KW |

The above inductive WPT system as shown in the above Fig. 1 was simulated using MATLAB/Simulink. The simulation was carried out by the help of derived system

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:42:59 UTC from IEEE Xplore. Restrictions apply.

parameters which are shown in the above Table I. All the system parameters are calculated and are displayed by the above equations from (1) to (11).

Fig. 3 and Fig. 4 shows the primary side tank capacitor voltage of 240 AC with opposite polarity and current value respectively after passing through high frequency DC-AC inverter.

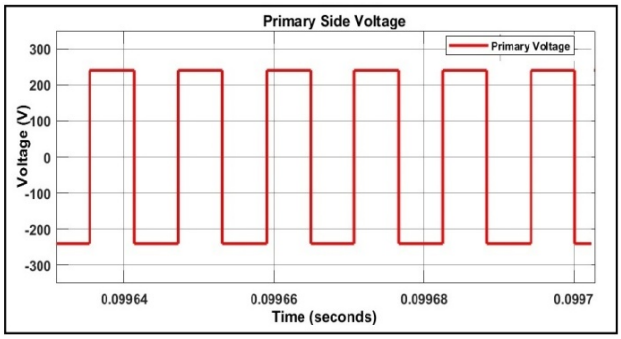



Fig. 3. Primary side tank capacitor voltage

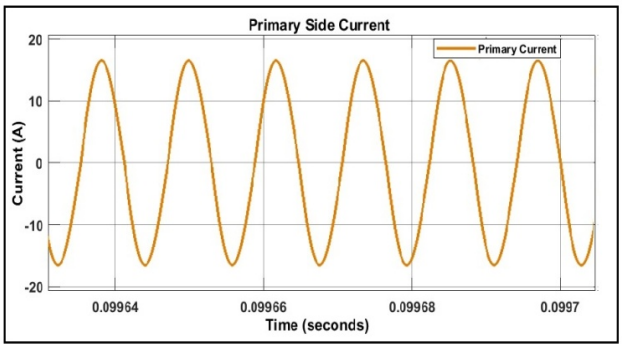



Fig. 4. Primary side tank capacitor current

Fig. 5 and Fig. 6 shows the secondary side tank capacitor voltage of 136 V AC and current of value 27 A respectively.

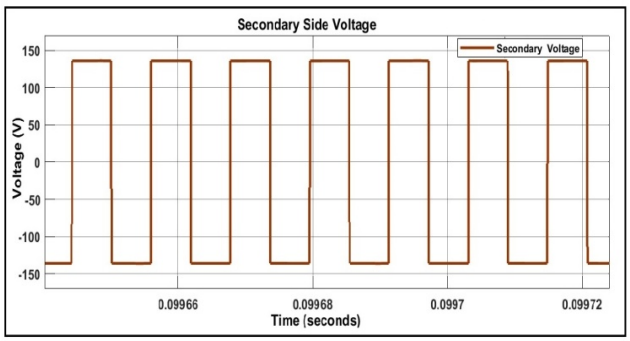



Fig. 5. Secondary side tank capacitor voltage

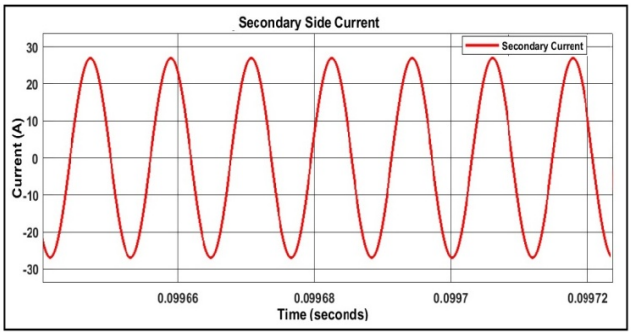



Fig. 6. Secondary side tank capacitor current

Fig. 7 and Fig. 8 shows the output voltage of 135 V DC and current of value 27 A respectively with 100 Ω load. These output voltage and current values are first calculated without adding the boost converter on the load side.

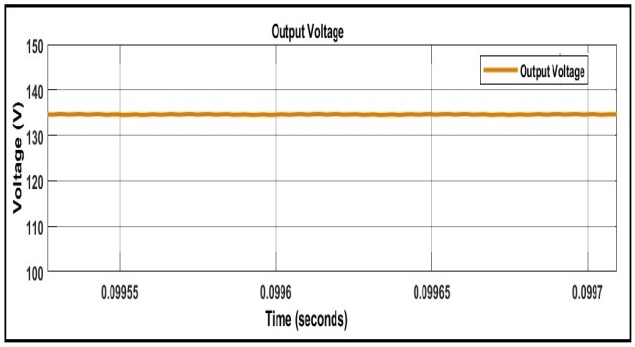



Fig. 7. Output voltage (without boost converter)

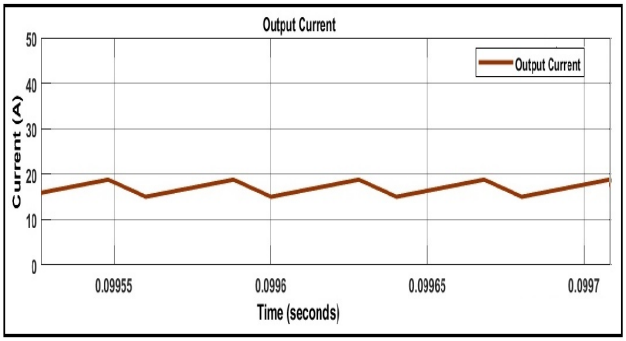



Fig. 8. Output current (without boost converter)

Fig. 9 and Fig. 10 shows the output voltage of 450 V DC and current of value 4.5 A respectively with 100 Ω load. These output voltage and current values are again calculated with adding a boost converter on the load side.

It is observed from Fig. 9 and Fig. 10 that both output voltage and current waveforms are smooth, so it becomes possible to get the smoothed output power of the system.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:42:59 UTC from IEEE Xplore. Restrictions apply.

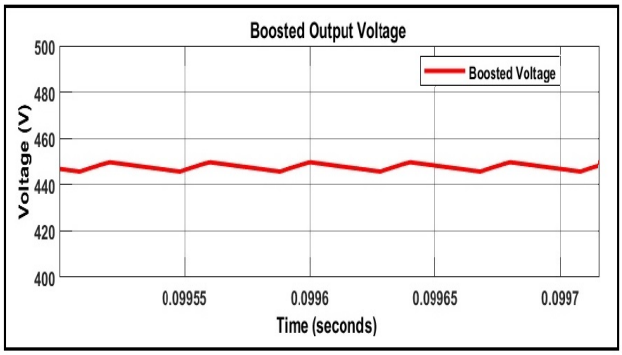



Fig. 9. Output voltage (with boost converter)

The boost converter used has to be implemented inside the EV, this leads to the additional weight on the EV.

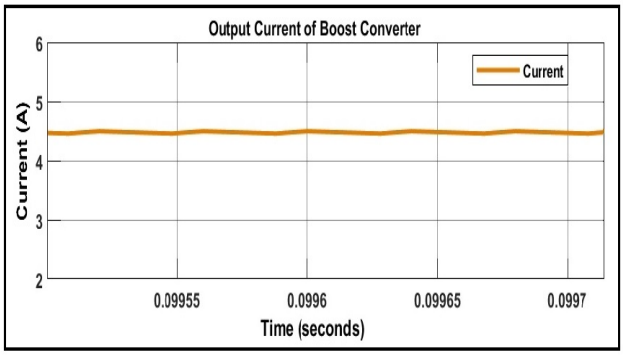



Fig. 10. Output current (with boost converter)

Fig. 11 shows the output power of 2 KW with or without adding the boost converter on the load side with 100 Ω load.

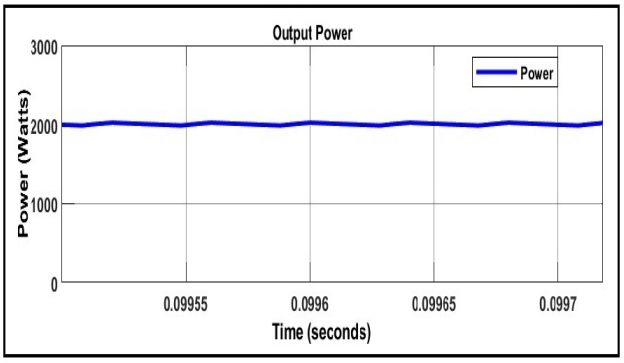



Fig. 11. Output power

## V. Conclusion

In this paper, by the help of Inductive coupling, a WPS system has been designed and the system was implemented/Simulated using the MATLAB/Simulink. The paper is a review of wireless charging of EVs. The simulation results are observed with or without adding the boost converter on the load side. The output power observed with or without boost converter is 2 KW with 100 Ω load. It is also observed that by using the boost converter, the output waveforms become smoother, which is very needed for battery charging applications. With the upcoming technology development, this system of wireless charging of batteries of EVs can be brought to reality and also it requires further studies/research.

## VI. References

[1] S. Schilling, “Ensuring Lead-Acid Battery Performance with Pulse Technology”, 1100 South Kimball Avenue, South Lake Texas, 1999.

[2] L. R. Chen, "A Design of an Optimal Battery Pulse Charge System by Frequency-Varied Technique," IEEE Transactions on Industrial Electronics, Vol. 54, No.1, February 2007, pp.398.

[3] L. R. Chen, “Design of duty-varied voltage pulse charger for improving Li-ion battery-charging response,” IEEE Trans. Ind. Electron, Vol. 56, No.2, Feb. 2009, pp.480-487.

[4] F. Musavi, W. Eberle, “Overview of wireless power transfer technologies for electric vehicle battery charging,” IET Power Electron., 2014, pp.1755-4535.

[5] C. Mi. M. A. Masrur, and D.W. Gao, “Hybrid Electric Vehicles: Principles and Applications with Practical Perspectives,” Hoboken, NJ, USA: Wiley, 2011.

[6] X. Zhang, and C.Mi, “Vehicle Power Management Modeling, Control and Optimization,” Berlin, Germany: Springer-Verlag, 2011.

[7] Saikot Baroi, Md. Shahidul Islam and Shawon Baroi, "Design and Simulation of a Wireless Charging System for Electric Vehicles," 2nd International Conference on Electrical & Electronics Engineering (ICEEE), 19-21 December 2017, RUET, Rajshahi, Bangladesh.

[8] Mohammad Shidujaman, Hooman Samani, Mohammad Arif, “Wireless Power Transmission Trends,” 3rd International Conference on Informatics, Electronics & Vision 2014.

[9] Siqi Li, and Chunting Chris Mi, “Wireless Power Transfer for Electric Vehicle Applications,” IEEE Journal of Emerging and selected topics in Power Electronics, Vol. 3, No. 1, March 2015.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:42:59 UTC from IEEE Xplore. Restrictions apply.

2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) | 978-1-6654-1120-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAECT54875.2022.9807975

# Implementation of Z-Source Resonant Converter for Power Factor Correction & Wireless Charging Applications

Ayush Kumar Srivastava
Department of Electrical Engineering
Delhi Technological University
Delhi, India
ayusriv08@gmail.com

Narendra Kumar
Department of Electrical Engineering
Delhi Technological University
Delhi, India
narendrakumar@dtu.ac.in

***Abstract*— The Z-source converter is discussed in this study in the context of Wireless Power Charging (WPC). The concept is illustrated using the Z-source resonant converter, which is a Wireless power Transmission) device for charging electric vehicle batteries. The Z-Source Resonant Converter (ZSRC) implements Power Factor Correction (PFC) and controls the voltage at the output side automatically that too simultaneously, thanks to the Z-Source Inverter (ZSI), without going for any control circuitry or extra semiconductor devices, as traditional PFC converters do. To put it another way, the ZSN is a single-stage PFC converter family. Furthermore, because the ZSN is not affected by shoot-through states, it is appropriate for high-power applications, increasing system stability and adding a boost function. This study describes and analyses the ZSRC-based WPT system's operational principle.**

***Keywords—LLC Resonant Converter, Half Bridge Inverter, Z-Source Inverter (ZSI), Z-Source Network (ZSN), Full Bridge Diode Rectifier, Wireless Power Transfer (WPT), Total Harmonic Distortion (THD), Power Factor Correction (PFC).***

## I. INTRODUCTION

Wireless power transfer (WPT) for electric vehicle (EV) battery chargers is now being researched for convenience, reliability, and environmental adaptation [1]. The WPT charger technology permits the person who is driving vehicle to position and charge the battery without coming out of the car in a motionless application. EVs could also be charged while driving through the embedded transmit coils along the lines of a Dynamic Wireless Power Transfer charging system. (Fig. 1) [2]. A properly-designed storage system for energy and charging has a potential to reduce strength of an electric car battery by 20% [3], resulting into decreased weight and cost.

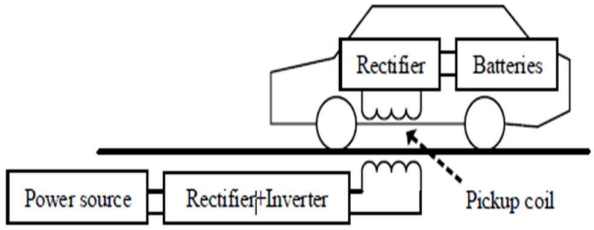



Fig. 1. Dynamic Wireless Power Transfer architecture

As shown in Fig. 2, a typical battery charger of On-Board Type (OBC) has a structure with two level containing PFC unit on front side and a high frequency transformer with a DC-DC Converter. Although each phase can be tweaked independently, the two-stage structure has a high overall cost and complexity.

OLPT applications have also been investigated using single-stage OBC converters. They maintain the voltage at the output side of the system while also improving the power factor (PF). Single phase single stage ($S^2$PFC) topologies have been studied in the past [4,5]. Mostly such topologies possess one switch only, ensuing in reduced prices as well as convenience of use. Because these converters descend from the flyback, they all place a great deal of strain on the transistor or semiconductor devices, necessitating the employment of a superior voltage handling capacity with a lot of transposition disadvantages. Consequently, the $S^2$PFC technique is suitable where reduced-power is utilized (i.e., shorter than 100W) where affordability is a primary consideration [6].

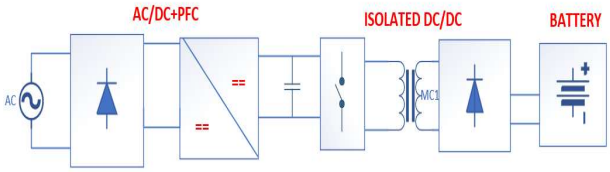



Fig. 2 Function block diagram of a Typical On-board Charger

The forward-unit PFC converter is combined with a resonant converter in other topologies. A boost converter and a half-bridge LLC converter are used as an example of this architecture in [7]. This topology's power specification is restricted.

Many different topologies have been investigated [1,8,9], however the ZSI will be chosen as a benchmark in this work.

## II. BATTERY CHARGER FOR ONBOARD USE

### *A. ZSRC WPT SYSTEM*

This study proposes the Z-source based Resonant converter as a revolutionary PFC Conversion Topology. The ZSI fundamental topology was created in [11] and is mostly used for governing voltage at the output end. For On Board Charging requirements, the Z-Source Inverter can be positioned in the Power Factor Correction level, followed by a secluded DC/DC converter, i.e a typical Series Resonant Converter. The Z-source resonant converter is the OBC-designed architecture (ZSRC).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:44:33 UTC from IEEE Xplore. Restrictions apply.

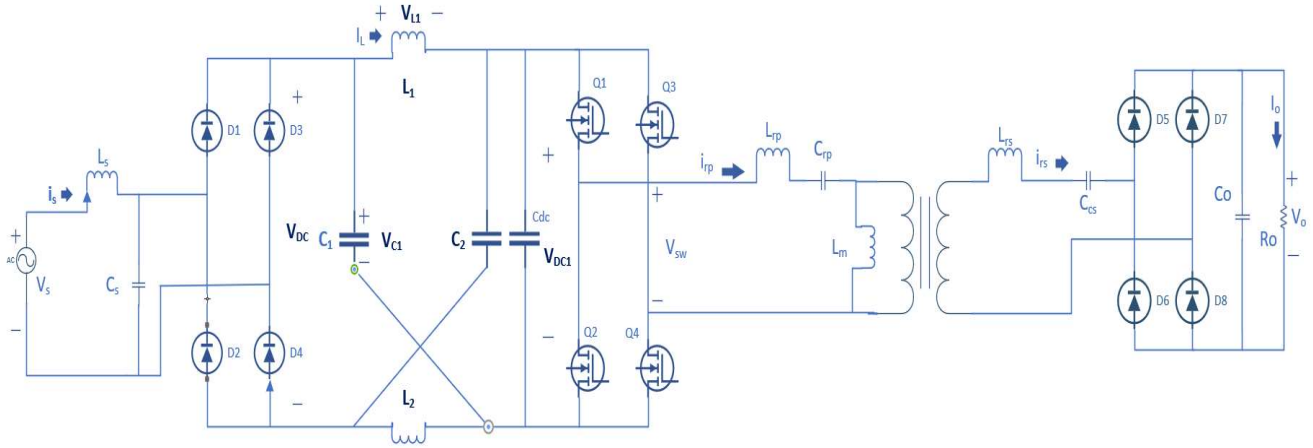



Fig. 3 Z-Source based Resonant Converter

Inherent PFC is a unique proposition dealt by ZSN present in ZSRC, without going for additional devices which is not available in typical PFC converters. It does so by preventing shoot-through states in the H-bridge inverter topology and introducing a control variable (shoot-through duty cycle parameter ($D_{st}$)) that could be used as a prototype for shaping the current at input end as a typical sine wave i.e in coherence with the voltage at the input end. This variable enhances the system as well, which is why it's frequently used for voltage regulation [11]. The devised ZSN-based OBC, on the other hand, controls the output voltage with the active state duty cycle parameter ($D_{act}$), a typical parameter for controlling in Series Resonant Converters. The Z-Source Inverter Topology has exceptional upsides in comparison with various PFC Converter Topologies like boost converter based WPT. We'll go through these in the next section.

### *B. Comparison of conventional & suggested PFC converters in general*

The Z-Source Inverter lowers the count of various power semiconductors devices in the system while slowly raising the count of passive nature components in contrast with a boost converter topology based PFC. The leading detriment of this is that the system at the primary end may be larger and bulkier. However, because this side is not at the bottom of the EV, its size and weight aren't as important in OBC applications. It's important to note that this is a hypothetical comparison. In convention, the boost PFC converter will be bulkier and more uneconomical due to the supplementary heat radiator for the power semiconductors devices and supplementary governing integrated circuits for the switch [12].

Capacitors and semiconductor devices, according to [13], have the highest failure and degradation rates of any component in a power electronic system (30 % and 21 % , respectively). Consequently, the presence of such devices in a circuit is less trustworthy than that of capacitors. The ZSN has two capacitors and no semiconductor devices, in comparison with it two capacitors and two semiconductor devices are used in a conventional PFC boost converter. The ZSN has a lower failure and degradation rate than a standard boost PFC converter, according to the research, resulting in a longer lifetime.

Ultimately, because the ZSN can manage shoot-through conditions, adding it to the OBC improves system reliability. As a result, random EMI shoot-through will have no effect on system functioning, however shoot-through states will be a worry for typical PFC converters. The notion of operation, evaluation, simulation, and observational evidence of the schemed Z-Source Inverter based Wirelss Transmission system are described in the subsequent segments.

## III. OVERVIEW OF Z-SOURCE BASED RESONANT CONVERTER

In comparison to DC/AC applications, Z-Source Resonant Converters have more states in a single switching cycle. All such states must be elucidated in order to perceive the ZSRC. Amidst such states, the boost type ratio of the Z-Source Inverter is still influenced by and large by the shoot through state duty cycle. The upcoming section explains the ZSRC operation principle.

### *A. ZSRC WPT system operation principle*

For conventional analysis, the AC voltage source and input rectifier were processed as a DC voltage source with a value of |Vac| (absolute value of AC input voltage), as illustrated in Fig. 4. Considered the fact that the resonant network serves as a type of filter permitting the current lying within a certain range or band of frequencies to pass, thereby flowing current at the resonant frequency only. Assuming, that the Z-Source Inverter (ZSI) components in Fig. 4 are symmetrical ($C_1 = C_2 = C$ and $L_1 = L_2 = L$), we have: $V_{C1} = V_{C2} = V_C$ and $V_{L1} = V_{L2} = V$. Furthermore, ZSN's switching frequency is ten times lower than L and C's resonant frequency. As a result, for a single switching cycle, the Z-Source Inverter (ZSI) inductor current and Z-Source Inverter (ZSI) capacitor voltage are deemed steady.

The principal output factor of the H-bridge type inverter topology is in coherence with the current at primary end ($i_{rp}$) of LLC Resonant Converter at resonance frequency [6], and the overall duration of shoot through period (Tst) is uniformly disseminated throughout one switching period.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:44:33 UTC from IEEE Xplore. Restrictions apply.

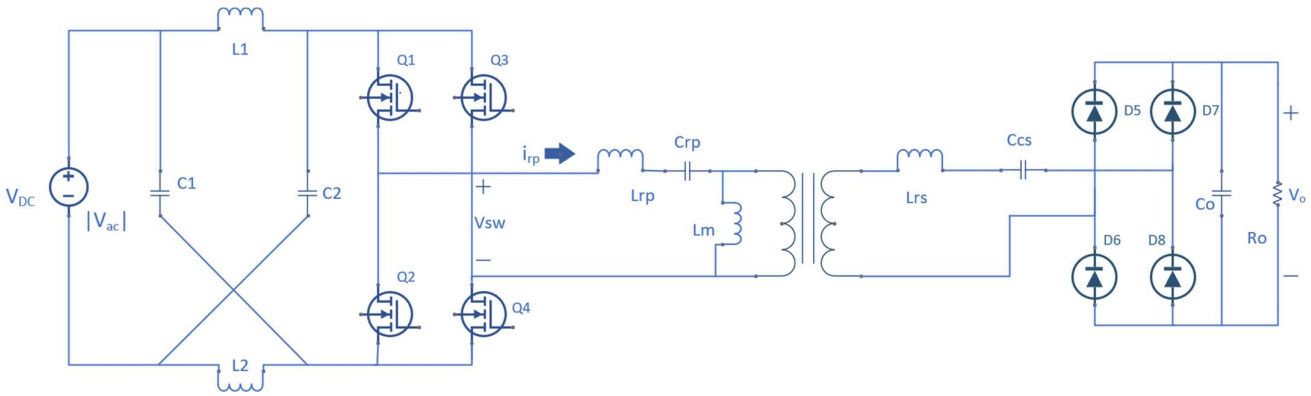



Fig. 4 Equivalent DC Model of ZSRC

**1. Active State:**

The diagonal switches are ON. Both the ZSN inductor and the capacitor supply current to the resonant network. The dissimilarity amid load current (irp) and ZSN inductor current is originated by connecting the two ZSN capacitors in series with the DC source ($I_L$). Only load current travels via the switches (irp). For this time interval, the ZSN inductor voltage is as follows:

$$V_L = 0.5(|v_{ac}| - v_z) = 0.5(|\widehat{v_{ac}} \sin \omega t| - v_z) \quad (1)$$

Where $\widehat{v_{ac}}$ is the AC input voltage maximum value and ω is the line's angular frequency.

**2. Shoot Through State:**

Three switches are turned on. The load current is carried by two horizontal switches, while the Z-Source Inverter (ZSI) inductor current is carried by two switches in one phase leg. As a result, the two currents are carried by a single switch. Because the Z-Source Network current of inductor is always unipolar and the current of load is bidirectional, these two currents whether subtract or add, assisting to the sum with their definite values. In the shoot through state, phase shift based control mechansim permits only various polarity currents to pass through the same switch. The Z-Source Inverter (ZSI) capacitors charge the Z-Source Network inductors here (this is how the Z-Source Resonance Circuit may enhance the voltage). For this time interval, the Z-Source Network inductor voltage is as follows:

$$V_L = V_C \quad (2)$$

**3. Zero State:**

Two horizontal switches are ON during the time interval of zero state. The Z-Source Network is not connected to the load in any way. The load current is freewheeling, and the Z-Source Network capacitors are charged by the inductors. For this time interval, the ZSN inductors' voltage is represented as:

$$V_L = V_{DC} - V_C \quad (3)$$

Using voltage second balance across the Z-Source Inverter (ZSI) inductor throughout one switching period by combining (1), (2), and (3), we have:

$$\frac{T_{act}}{T_s}(V_{DC} - V_C) + \frac{T_{st}}{T_S}V_C + \left(1 - \frac{T_{act}}{T_S} - \frac{T_{st}}{T_S}\right)(V_{DC} - V_C) = 0 \quad (4)$$

$$V_C = \frac{1 - D_{st}}{1 - 2D_{st}} V_{DC} \quad (5)$$

$$D_{st} = \frac{V_Z - V_{DC}}{2V_Z} = \frac{V_C - V_{DC}}{2V_C - V_{DC}} \quad (6)$$

$$V_Z = \frac{V_{DC}}{1 - 2D_{st}} \quad (7)$$

The PFC is governed by the shoot-through duty cycle, which is another control variable in the system. $D_{act} + D_{st} + D_{zer} = 1$ should always be met by its management, where $D_{zer} = \frac{T_{zer}}{T_s}$ is the conventional zero state duty cycle.

The voltage on the ZSN capacitor is unrelated to the active state duty cycle Dact, according to this relationship in (6).

All of ZSRC's conceivable states are represented by these three states. Varying load regulation characteristics would result from different allocations of these three states during the course of a switching period.

TABLE I. SWITCH SELECTION DURING DIFFERENT STATES

| STATES | INVERTER SWITCHES ON |
|---|---|
| ZERO | Q2,Q4 or Q1,Q3 |
| SHOOT THROUGH | Q1,Q2,Q4 or Q1,Q3,Q4 or Q1,Q2,Q3 or Q2,Q3,Q4 |
| ACTIVE | Q1,Q4 or Q2,Q3 |

Fig. 5 Choice of Switch

*B. Overview of LLC Resonant Converter*

Figure 3 illustrates a full-Bridge LLC resonant DC-DC converter with resonant capacitor $C_r$, resonant inductor $L_r$, and magnetising inductor $L_m$. It consists of three elements:

a. Square Wave source: Power switches Q1, Q2, Q3, Q4, which are generally MOSFETs, are arranged which act as a type of square wave prime mover. This generator

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:44:33 UTC from IEEE Xplore. Restrictions apply.

yields a bipolar square-wave voltage ($V_{sw}$) by driving switches Q1 and Q4 & Q2 and Q3 alternately with pulse frequency modulation approach that uses a fluctuating frequency clock to propel the power switches and deliver energy from the input to the output.

b. Resonant Network: Resonant capacitor $C_r$, resonant inductor $L_r$, and magnetising inductor $L_m$ comprise the resonant network. The resonant network circulates the energy and delivers it to the load via the transformer. A bipolar square wave is received at the transformer's primary side and passed via the secondary winding. The transformer also provides electrical isolation and the proper turn ratio for the needed output voltage.

c. Rectifier Network: The rectifier network converts AC voltage to DC voltage using diodes and capacitors. The rectifier network is made up of a complete bridge rectifier with a capacitive output filter.

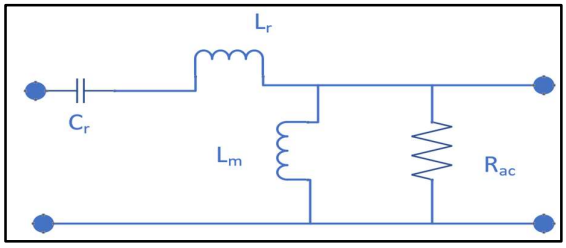



Fig. 6 Equivalent model of LLC resonant converter

## IV. SIMULATION RESULTS

To validate the analysis, simulations of the proposed system were run at (192 W). The simulation parameters and component values are listed in Table I. The ZSN design was based on [12, 14] and was targeted at decreasing the twice-line frequency ripple (twice ripple) passing through it. The compensating capacitors were set so that the band pass filter's resonant frequency equalled the switching frequency in the SRC design.These systems have switching frequencies ranging from 100 to 130 kHz [17]. The frequency of 110 kHz was chosen for this study.

TABLE II. PARAMETERS VALUES

| **Parameters** | **Values** |
|---|---|
| Input Voltage($V_{ac}$) | 220 V RMS |
| Line Frequency | 50 Hz |
| Switching Frequency | 117 KHz |
| Resonant Frequency | 110 KHz |
| Voltage at Output End ($V_O$) | 24 V |
| Current at Output End($I_O$) | 8 A |
| Turns Ratio | 12.08:1 |

TABLE III. COMPONENTS VALUES

| **Components** | **Values** |
|---|---|
| $C_1$,$C_2$ | 4.7 mF |
| $L_1$,$L_2$ | 1 µH |
| $L_{rp}$ | 53.87 µH |
| $C_{rp}$ | 38.86 nF |
| $L_m$ | 188.54 µH |
| $C_O$ | 0.1 mF |
| $R_O$ | 3 Ω |

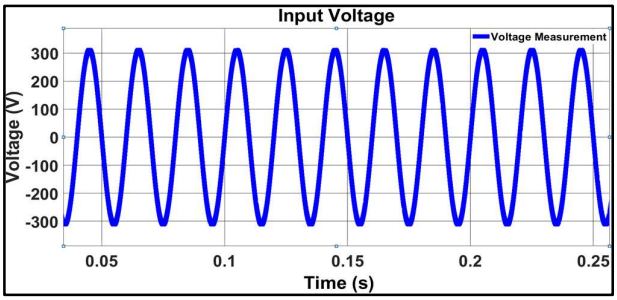



Fig. 7 Voltage Waveform at the Input Side

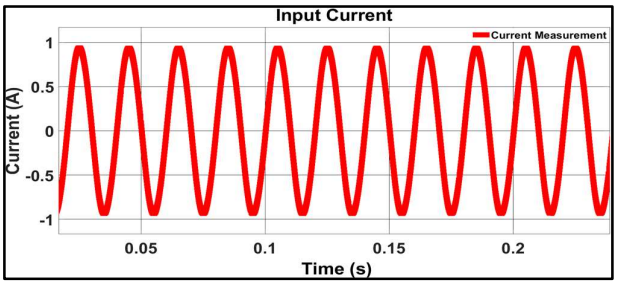



Fig. 8 Current Waveform at the Input Side

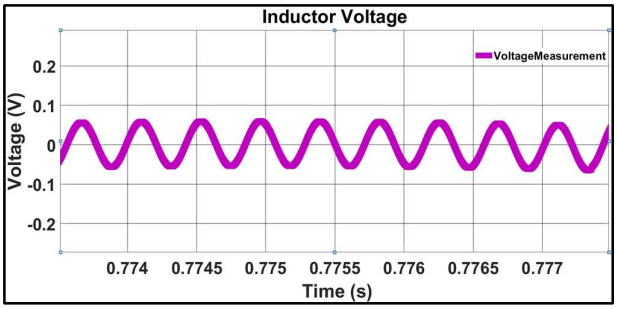



Fig. 9 ZSN Inductor Voltage Waveform

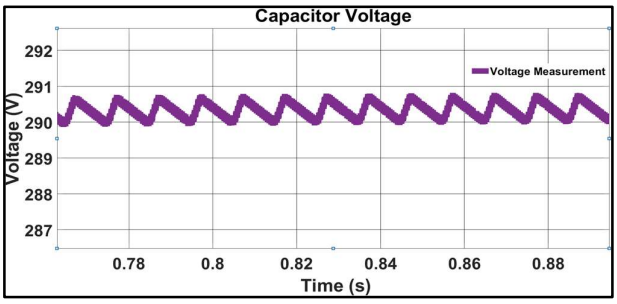



Figure 10 ZSN Capacitor Voltage Waveform

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:44:33 UTC from IEEE Xplore. Restrictions apply.

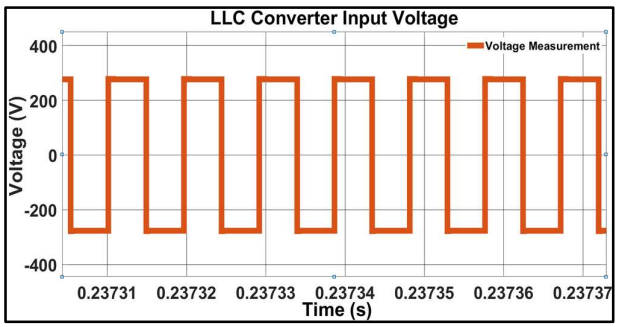



Fig. 11 Voltage Waveform of LLC Resonant Converter

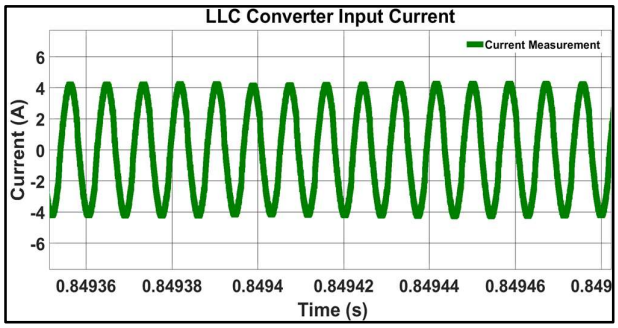



Fig. 12 Current Waveform of LLC Resonant Converter

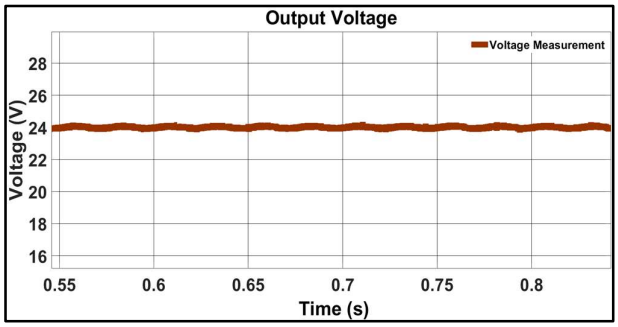



Fig. 13 Voltage Waveform at the Output Side

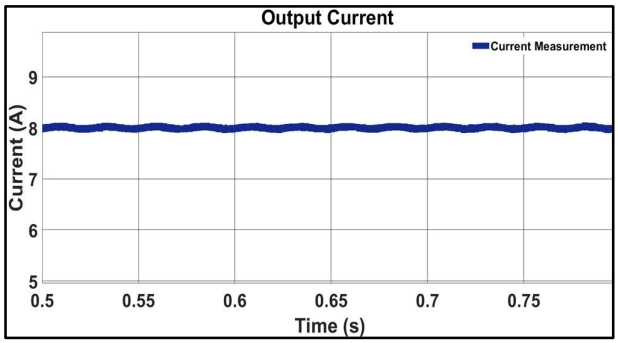



Fig. 14 Current Waveform at the Output Side

## V. CONCLUSION

This research study used a Z-Source Resonance Converter WIRELESS POWER TRANSFER system for electric vehicle battery charging as an example to demonstrate the ZSN for Wireless Charging applications. Considering the engagement of the Z-Source Inverter (ZSI), the Z-Source Resonant Type Converter becomes resistant to shoot-through period disturbance, thereby increasing system reliability while also introducing a new control parameter (shoot-through period duty cycle) that can be used for Power Factor Correction (PFC). In this study, the proposed Z-Source Network-based On Board Charger was analysed and compared to the usual boost-converter-based OBC.

This project used simulations to validate the theoretical analysis & could be used as an On-Board Battery Charger of 24V & 8A. The devised single stage PFC architecture when simulated on MATLAB achieved a PF of 0.9981 and a T.H.D of 1.33 % thereby settling the system output voltage to 24V.

Future work will include optimising the governing methodology to reduce harmonic content and integrating soft switching to improve system efficiency so as to meet the set standards. Also, while implementing this research study into a hardware model, we need to use a Mutual Inductor in place of Transformer for executing Wireless Power Transmission.

## REFERENCES

[1] S Yilmaz, M.; Krein, P.T., "Review of Battery Charger Topologies, Charging Power Levels, and Infrastructure for Plug-In Electric and Hybrid Vehicles," in Power Electronics, IEEE Transactions on, vol.28, no.5, pp.2151-2169, May 2013.

[2] Modern Trends in Inductive Power Transfer for Transportation Applications- Covic, G.A.; Boys, J.T., "Modern Trends in Inductive Power Transfer for Transportation Applications," in Emerging and Selected Topics in Power Electronics, IEEE Journal of, vol.1, no.1, pp.28-41, March 2013.

[3] Musavi, F.; Edington, M.; Eberle, W., "Wireless power transfer: A survey of EV battery charging technologies," Energy Conversion Congress and Exposition (ECCE), 2012 IEEE, vol., no., pp.1804, 1810, 15-20 Sept. 2012.

[4] Ghanbari, A.R.; Moghani, J.S.; Abdi, B., "Single-stage soft-switching PFC converter based on DCVM buck and flyback converters," in Power Electronics, Drive Systems and Technologies Conference (PEDSTC), 2013 4th, vol., no., pp.218-223, 13-14 Feb. 2013.

[5] ShiguoLuo; WeihongQiu; Wenkai Wu; Batarseh, I., "Flyboost power factor correction cell and a new family of single-stage AC/DC converters," in Power Electronics, IEEE Transactions on, vol.20, no.1, pp.25-34, Jan. 2005.

[6] Zhang, Junming, Hulong Zeng, and Ting Jiang. "A primary-side control scheme for high-power-factor LED driver with TRIAC dimming capability." Power Electronics, IEEE Transactions on 27.11 (2012): 4619-4629.

[7] Lai, C.M.; Shyu, K.K., "A single-stage AC/DC converter based on zero voltage switching LLC resonant topology," in Electric Power Applications, IET, vol.1, no.5, pp.743-752, Sept. 2007.

[8] Musavi, F.; Edington, M.; Eberle, W.; Dunford, W.G., "Evaluation and Efficiency Comparison of Front-End AC-DC Plug-in Hybrid Charger Topologies," in Smart Grid, IEEE Transactions on, vol.3, no.1, pp.413-421, March 2012.

[9] Vyshakh, A.P.; Unni, M.R., "Performance enhanced BLIL PFC boost converter for PHEV battery chargers," in Emerging Research Areas: Magnetics, Machines and Drives (AICERA/iCMMD), 2014 Annual International Conference on, vol., no., pp.1-6, 24-26 July 2014.

[10] Cha, Honnyong, Fang Z. Peng, and DongWookYoo. "Z-source resonant DC-DC converter for wide input voltage and load variation." Power Electronics Conference (IPEC), 2010 International. IEEE, 2010.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:44:33 UTC from IEEE Xplore. Restrictions apply.

[11] Fang ZhengPeng, "Z-source inverter," Industry Applications, IEEE Transactions on, vol.39, no.2, pp.504,510, Mar/Apr 2003.
[12] M. Shen. “Z-source inverter design, analysis, and its application in fuel cell vehicles.” Ph.D. dissertation, Michigan State University, East Lansing, Michigan, 2006.
[13] Shaoyong Yang; Dawei Xiang; Bryant, A.; Mawby, P.; Ran, L.; Tavner, P., "Condition Monitoring for Device Reliability in Power Electronic Converters: A Review," in Power Electronics, IEEE Transactions on, vol.25, no.11, pp.2734-2752, Nov. 2010.
[14] MiaosenShen; Joseph, A.; Yi Huang; Peng, F.Z.; ZhaomingQian, "Design and Development of a 50kW Z-Source Inverter for Fuel Cell Vehicles," in Power Electronics and Motion Control Conference, 2006. IPEMC 2006. CES/IEEE 5th International, vol.2, no., pp.1-5, 14-16 Aug. 2006.
[15] Hulong Zeng; Shuitao Yang; FangzhengPeng, "Wireless power transfer via harmonic current for electric vehicles application," in Applied Power Electronics Conference and Exposition (APEC), 2015 IEEE, vol., no., pp.592-596, 15-19 March 2015.
[16] Huh, J.; Lee, S.W.; Lee, W.Y.; Cho, G.H.; Rim, C.T., "Narrow-Width Inductive Power Transfer System for Online Electrical Vehicles," in Power Electronics, IEEE Transactions on, vol.26, no.12, pp.3666-3679, Dec. 2011.
[17] A Unity-Power-Factor IPT Pickup for High-Power Applications-Keeling, N.A.; Covic, G.A.; Boys, J.T., "A Unity-Power-Factor IPT Pickup for High-Power Applications," in Industrial Electronics, IEEE Transactions on, vol.57, no.2, pp.744-751, Feb. 2010.
[18] Electrical engineering: principles and applications AR Hambley, N Kumar, AR Kulkarni Pearson Education (New Jersey), 2011.
[19] Power Electronics Devices MH Rashid, N Kumar, A Kulkarni - Circuits and Applications (4th Edition), England, 2014.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:44:33 UTC from IEEE Xplore. Restrictions apply.

2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) | 978-1-6654-1120-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAECT54875.2022.9807996

# Importance of Initialization in K-Means Clustering

Anubhav Gupta
*Dept. of Software Engineering*
*Delhi Technological University*
Delhi, India
anubhavg0407@gmail.com

Antriksh Tomer
*Dept. of Software Engineering*
*Delhi Technological University*
Delhi, India
antriksh12off@gmail.com

Sonika Dahiya
*Dept. of Software Engineering*
*Delhi Technological University*
Delhi, India
sonika.dahiya@dtu.ac.in

***Abstract*—Data clustering is a method of visualizing the data in such a way that enables the researcher to see similar patterns formed in the data and these lead to conclusions that can be helpful to interpret the data and could be further used for other research purposes. In this paper the focus would be on the initialization technique used and would present how an improper initialization of centroid could lead to bad or unfruitful results, not only this the complexity of the overall algorithm depends upon the type of initialization used. Thus, study compares various initialization techniques and their respective research work to come upon a study that would help the researcher to get an insight of the available techniques and thus choose the one suitable. This research would focus on the nature of the data presented and would see how different types of Datasets get affected by the choice of initialization. Along with this would also analyze the impact of repeating the K-Means Clustering Algorithm on the results.**

***Keywords – Centroid Index, Sum-of-Squared Errors, Initialization, Time Complexity, Clustering***

## I. INTRODUCTION

### A. Overview

Clustering is the process of converting unlabelled data into meaningful data that can be organized into clusters.

K-Means Clustering Algorithm

It is an unsupervised learning technique. Its popularity is well deserved because of the way it produces such significant results even after utilizing very less T.C. and S.C. and since the traditional method uses random initialization of centroid it becomes simpler in comparison to others but yes very effective at the same time, and the biggest reason for the same is it's easy to implement process.

The steps followed by the algorithm is been demonstrated below:

1. Randomly first select any K number of data points from the dataset.
2. Then for each data point calculate the distances between all the current centroids and then assign that data point to the cluster whose centroid is closest to it.
3. Next after the assignment of clusters now calculate the new mean / centroid for each and every cluster using the new set of assigned data points.
4. Until the convergence is not reached keep on repeating the Steps 2 and 3, convergence is the scenario when no data point needs to change its cluster as its closer to its previous centroid only.

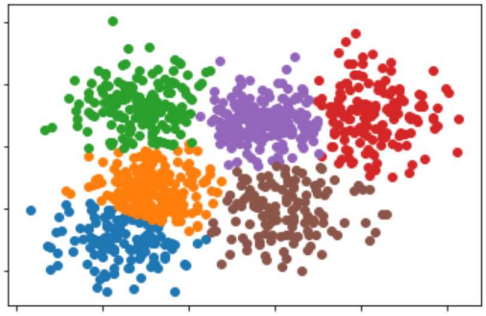


Fig. 1. Clustering.

Each iteration of K-Means uses O(kN) Time Complexity. K-means clustering is more popular since it converges quickly. Several factors have contributed to its appeal. To begin with, it is conceptually basic and straightforward to implement. It is included in almost every data mining software package. Second, it is adaptable, meaning that nearly every part of the algorithm (initialization, distance function, termination criterion, and so on) may be changed.

Applications – The algo is widely used in Pattern Recognition Systems, Image Classifications, Information Retrieval in Web Resources, Biology etc.

The algorithm is not so straightforward though because it has some drawback or you can say weakness [12,14] as most of the researchers call it.

Its weaknesses are first, it requires the user to provide the K parameter, so in some cases this can be teasing or difficult for the user to determine it and may lead to confusions.
Second, since it deals with Euclidean Dist. in its working, a group or some significant numbers of outlier or noisy points can affect the overall clustering hence tamper with the results of the clusters formed.
Third, high dependence on just 1 step i.e., dependent on initialization of centroids, if not initialized properly it can lead to bad local minima which means that it won't give its best classifications.

All these are some drawbacks that are encountered when using K-Means Algorithm, but in this study the main focus is going to be to address the 3rd weakness which is the initialization of centroids.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:41:15 UTC from IEEE Xplore. Restrictions apply.

*B. Motivation*

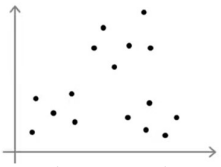


Fig. 2. Data Points.

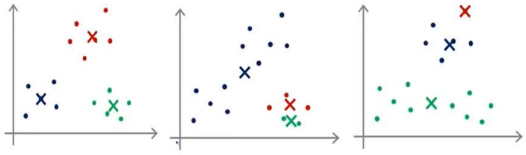


Fig. 2(a). Fig. 2(b). Fig. 2(c).

Given K=3 in this case. Now in the above figures a very interesting thing happens i.e., all the 3 diagrams below here are the 3 different outputs for the same data given in Figure-2 and the explanation of this result is the weakness of the algorithms discussed above.

All 3 of them had random centroids and hence resulted in 3 different results whereas by naked eye one can see that 2(a) should be the most optimal result but yet K-means gave 2(b) and 2(c) as the results which clearly shows them incorrect or can also be said that 2(b) and 2(c) unfortunately got stuck in a bad local minimum due to their poor initialization. This explains how much centroid explanation is important and this very first step can prove out to be a game changing for this algorithm. The biggest reason is that centroids are unable to move when there are stable clusters between them or if the distance is big. So, this gave a basis to research and work upon this topic to compare and give a survey on various centroid initialization techniques used for K-means Algorithm.

## II. Related Work

Researchers are constantly working to improve the K-means algorithm's accuracy and efficiency. There are numerous numbers of methods [13] but this section discusses some of the most popular and common initialization techniques proposed.

Here are some Initialization Techniques Proposed-

1. Forgy's Approach

This technique [4] is still used by K-Means. Forgy proposed to randomly initialize the centroids, to explain this he said that it would be more probable that the selected centroid is near to centroid of the cluster because a high margin of data point would be present there so the probability would increase.

2. MacQueen Method

In this method [5] K initial points are taken as centroid and when next point comes their distances are calculated with the centroids and the centroid with the lowest distance is chosen and its cluster id is given to that data point now a mean is taken for all the previous data points in that cluster previously along with the newly added cluster point and hence the new centroid is generated for that particular cluster. So, this method is updating the centroid for each and every incoming data points and stops when all the data points are assigned their respective clusters.

3. Bradley and Fayyad Method

They proposed [2] a method in which they divide their dataset into L subset. Then the K-Means with Forgy's approach is applied to each of these L subsets and as a result for this would obtain K centroids for each and every L subsets thus resulting in a total L*K centroids. Then the K-Means is run L number of times initialized with K number of centroids obtained int the previous results. The results returned from these are then used as the initializations. This method has a very high computational cost when compared to others, but still is one of the likes for the initialization centroids methods.

## III. Methodology

*A. Evaluation Methods*

1. Sum-of-Squared Error (SSE)

$$\text{SSE} = \sum_{i=1}^{N} ||x_i - c_j||^2 \quad (1)$$

Suppose there are N data points, then it is the sum of square of distance between each and every data point from its closest assigned centroid. Lesser the SSE better the performance. Also called Cost or Distortion Function.

SSE is also sometimes referred to as local minimum as each iteration can have its own local minimum value and the lowest possible value is often called global local minimum. This method is widely used as an evaluation method [16].

2. Centroid Index (CI)

Proposed in [3] is a different method of analyzing result obtained from K-Means. This says that SSE is a very raw metric to measure because one cannot judge about the clustering from some random values and does not give some practical information on how good or accurate the proposed method is. Now, this is where this metric comes into play the value CI denotes how many centroids in the cluster are wrongly placed. C=0 is the best-case scenario possible which would mean 0 clusters are misplaced.

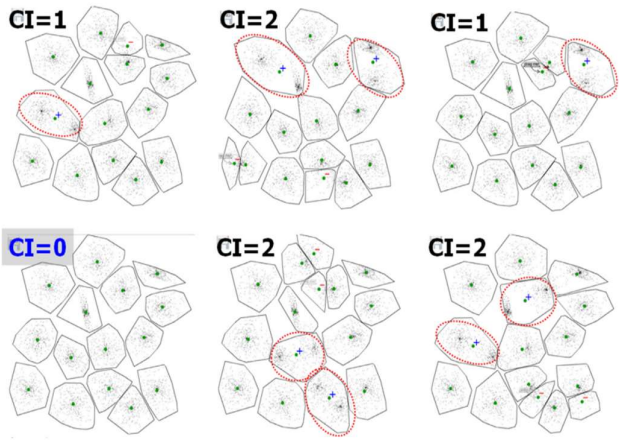



Fig. 3. Demonstration of Centroid Index.

Here (+) denotes that cluster is missing a centroid and (-) denote that centroid is incorrectly assigned to the cluster.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:41:15 UTC from IEEE Xplore. Restrictions apply.

*B. Factors affecting Data*

The following factors have been taken into account for the study of the behavior they impose when tested with the different initialization methods in K-Means clustering.

1. Cluster Overlap – This defines how much are the different clusters overlapping in the data, that is the data can be highly overlapped or can also have well separated clusters.

2. No. of Clusters – This defines the no. of clusters present inside the data.

3. Dimension – This defines dimension of the data.

4. Unbalance – This defines whether the data is highly imbalanced or well balanced.

This research focusses more on the Cluster Overlap and the Unbalance factor than the other 2 in this paper.

*C. Dataset Used*

The dataset used here is the clustering basic benchmark [6]. Remember each of the dataset has its own significance i.e., each of them is related to 1 of the 4 factors discussed above. The dataset is available here: http://cs.uef.fi/sipu/datasets/

1. For Cluster Overlap study uses the Dataset S which is further in 4 forms and lie in the order of S1, S2, S3 and S4 where S1 is well separated whereas S4 is highly overlapped.
   N=5000 data points and k=15 clusters with different degrees of cluster overlap.

2. Dataset A further divided to A1(k=20), A2(k=35) & A3(k=50) where each cluster has 150 points.

3. Unbalance Dataset to study the behavior based upon the balance of data (this dataset is imbalanced in nature). N=6500 data points and k=8 Clusters.

*D. Algorithms*

1. Repeated K-Means

This algorithm is not something on its own but is just an advancement of K-Means Clustering Algorithm [1] and the only difference in this is that here the code initializes the K-means with different values and then taking the one among all the repetitions that has the lowest value of SSE.

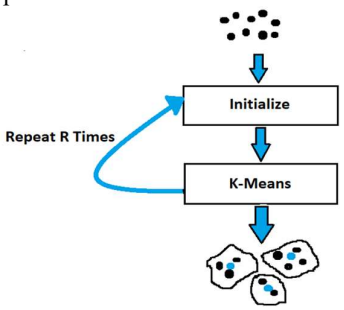



Fig. 4. Demonstration of Repeated K-Means.

The thing to see here is that this repetition comes at a cost as the overall complexity is now multiplied with the number of times the algorithm is repeated. Hence some researchers do not prefer to use it at all [15].
It is useful only when the initialization technique used has some randomness because if there is no randomness at initialization, every time K-Means is run would generate almost the similar result and hence repeating would be of no use and an unnecessary increase in complexity.

2. Random Centroid

This method is also popularly known as Forgy's Initialization [4]. Classical K-means Clustering Algorithm uses the same in its working due to its simple complexity.
If the number of partitions is K, then this method randomly assigns any K data points from the given data as centroids. It makes sure that no data point is repeated for the candidacy of being a centroid.
To explain this, he said that it would be more probable that the selected centroid is near to centroid of the cluster because a high margin of data point would be present there so the probability would increase.

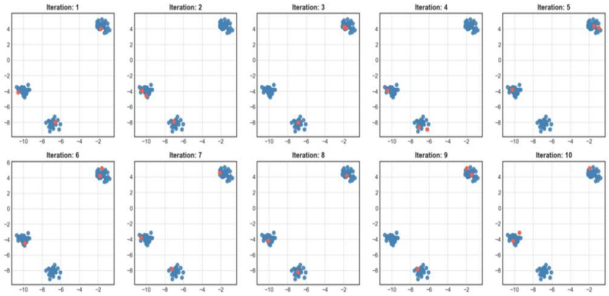


Fig. 5. Random Centroid.

Iteration here should not be confused with K-Means Iteration it just denotes the different result obtained just after initialization and show the initial k centroids. It can be seen 4, 5, 9, 10 have very badly placed initial centroids and this would result in bad final result or bad local minimum.

3. Random Partition

All the data point are randomly assigned to one of the K clusters [10]. Now, for each cluster, average of all the data point lying in that only particular cluster is calculated and hence this average value as the output acts as the initial Centroid for the algo i.e., is actually one of the K centroids formed. This algo works poorly with well separated data.
Although there is randomization but it is still more deterministic as centroid are by the nature of algorithm are produced in the center of the dataset.

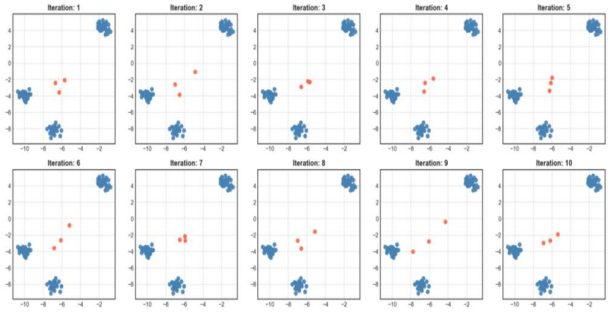


Fig. 6. Random Partition.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:41:15 UTC from IEEE Xplore. Restrictions apply.

It can be seen all initializations have all the 3 centroids in the center and that is because algorithm has assigned the centroids by taking the average of a cluster of data points.

4. K-Means++

It is also known as the randomized version of Maxmin which is another Initialization Technique [9]. It is the most popular among all of the initialization technique as it lowers the likelihood of a bad cluster and speeds up convergence.

Steps involved to find K clusters-

1) It randomly chooses any 1 point from the dataset.
2) Now for each of the remaining data points find only the 1 distance between itself and its nearest centroid.
3) Now assign higher probability to those data point whose distances is greater which also means that data point would have a higher probability of being selected as one of the K centroids.
4) Now, repeat the steps 2 & 3 until get K initial Centroids.

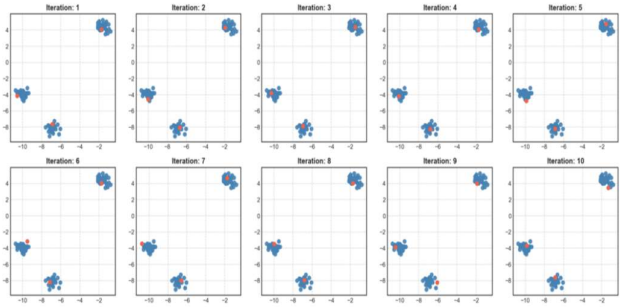


Fig. 7. K-Means++.

Here, each and every run has very good centroid initialization and this is also evident why this is one of the most popular algorithms for centroid initializations, running K-means now for every iteration would give much better results when compared to the above 2 methods.

5. Maxmin

The algorithm [11] it uses comes from its name itself. First it selects any arbitrary point as the very first centroid and then selects the next centroid whose distance from its closest(min) existing centroid is the largest (max) [7]. The K-Means++ and Maxmin are very similar and uses the similar concept to get the first k initial centroids.

6. Sorting-Heuristic

As the name says here, first sort the data point but this sorting is used differently by different researchers, here the method proposed in [1,8] is used, which tells to sort the data points according to their distance to the center of the data and then choose every N/kth occurring in the sorted order of data points until there are a total of k centroids.

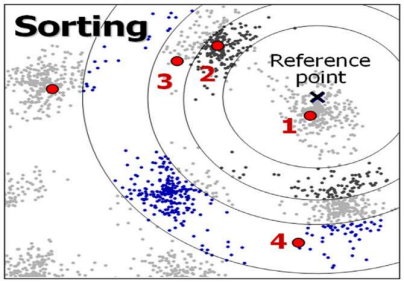



Fig. 8. Sorting Initialization Technique.

But it is also observed that actually in practice the centroids selected are just some of the random points in the dataset. The approach can be visualized by the figure shown above after applying the Sorting-Heuristic Initialization.

TABLE 1. Initialization Methods Used

| Initialization Method | Complexity | Randomized |
|---|---|---|
| **Random Centroid / Forgy** | O(N) | Yes |
| **Random Partition** | O(N) | Yes |
| **K-Means++** | O(kN) | Yes |
| **Maxmin** | O(kN) | Modified |
| **Sorting Heuristic** | O(N logN) | Modified |

This also mentions whether it uses random initialization because that would be helpful in the case when analyzing the repeated K-Means.

## IV. Comparisons and Results

### *A. Cluster Overlap*

It can be assumed logically that well separated data would be easy for clustering algorithms to cluster the data but actually it's not. For K-Means, the algorithm performs worse when there are well separated clusters.

It can be seen that for the most well separated dataset S1, it is clear that for K-Means++, Maxmin & Sorting Heuristic the SSE & the CI-Values are distinguishably lower then both the values of Random Centroid and Random Partition. This proves why a good initialization method is important. But, on moving from S2 to S4 the difference in both SSE and CI-Values between the 5 algorithms become closer and closer, and for S4 they are very similar.

This leads to a conclusion that initialization plays a major role when the data is well separated but is not important when the data is highly overlapped, which means using any of the 5 algorithms would yield similar results for highly overlapped data because of the K-Means Iterations.

Further to this extension in [1] it said that in well separated clusters K-Means does not play a major role and the whole algorithm depends merely upon the initialization method used i.e., the results depend mostly on initialization method used. Whereas for highly overlapped data whatever the initialization is used K-Means would step in and generate good results for all of the initialization methods anyways.

So, a choice of good initialization method becomes a priority when dealing with well separated data, and since K-Means Iterations does very little work after initialization, some other algorithm can be used instead to generate more better results.

### *B. No. of Clusters*

In the Table 4. the SSE increased when the dataset had a greater number of clusters present in it similar

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:41:15 UTC from IEEE Xplore. Restrictions apply.

TABLE 2. Shows the results for Cluster Overlap dataset S. It shows the average SSE values obtained from 20 different runs. The SSE was divided by 10^8 due to large values.

| Initialization Method | Sum-of-Squared Errors | | | | CI-Values | | | |
|---|---|---|---|---|---|---|---|---|
| | *S1* | *S2* | *S3* | *S4* | *S1* | *S2* | *S3* | *S4* |
| **Random Centroid / Forgy** | 198163 | 153356 | 195251 | 177573 | 1.8 | 1.4 | 1.3 | 0.9 |
| **Random Partition** | 235342 | 188865 | 187541 | 175475 | 3.3 | 0.6 | 1.2 | 0.4 |
| **K-Means++** | 116526 | 142792 | 186736 | 168766 | 1.0 | 0.9 | 1.0 | 0.8 |
| **Maxmin** | 119548 | 140758 | 174892 | 175899 | 0.7 | 1.0 | 0.7 | 1.0 |
| **Sorting Heuristic** | 132341 | 161897 | 180148 | 179784 | 1.3 | 1.1 | 1.0 | 0.7 |

TABLE 3. Shows the results for Cluster Overlap dataset S. It shows the most optimal SSE obtained from 10 repeated K-Means. CI-Values are obtained after 100 repeats.

| Initialization Method | Sum-of-Squared Errors | | | | CI-Values | | | |
|---|---|---|---|---|---|---|---|---|
| | *S1* | *S2* | *S3* | *S4* | *S1* | *S2* | *S3* | *S4* |
| **Random Centroid / Forgy** | 135526 | 102792 | 118903 | 120033 | 0.1 | 0 | 0 | 0 |
| **Random Partition** | 143489 | 128835 | 120855 | 115896 | 1.4 | 0 | 0 | 0 |
| **K-Means++** | 89176 | 112791 | 127903 | 117039 | 0 | 0 | 0 | 0 |
| **Maxmin** | 98548 | 104758 | 125896 | 123896 | 0 | 0 | 0 | 0 |
| **Sorting Heuristic** | 112341 | 121897 | 125874 | 113894 | 0 | 0 | 0 | 0 |

TABLE 4. Shows the results for No. of Clusters dataset A. It shows the average SSE values obtained from 20 different runs.

| Initialization Method | Sum-of-Squared Errors | | | CI-Values | | |
|---|---|---|---|---|---|---|
| | *A1* | *A2* | *A3* | *A1* | *A2* | *A3* |
| **Random Centroid / Forgy** | 153 | 277 | 432 | 2.5 | 4.5 | 6.6 |
| **Random Partition** | 315 | 597 | 985 | 6 | 10.7 | 17.9 |
| **K-Means++** | 121 | 228 | 314 | 1.5 | 2.9 | 4.2 |
| **Maxmin** | 133 | 231 | 325 | 1 | 2.6 | 2.9 |
| **Sorting Heuristic** | 129 | 255 | 495 | 1.5 | 3.6 | 5.5 |

CI-Values used in the tables are obtained from [1].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:41:15 UTC from IEEE Xplore. Restrictions apply.

is the case with CI-Values for ex. using K-Means++ values 121, 228 & 314 SSE's were obtained for k=20, 35 & 50 respectively. The reason for the same is that since there are more numbers of clusters or can say algorithm now needs to deal with higher number of clusters, this leads to an increase in SSE and the CI-Values because obviously the clustering now would not be as good as when there were a smaller number of clusters present in the dataset, so the initialization technique would not affect this. But still a good initialization technique must be used as it lowers the SSE as well as CI-Values as seen between Random Partition and K-Means++.

In [1] further proposed/concluded that the performance of the K-Means Algorithm would have a linear dependency on k i.e., greater the number of clusters worser the algorithm performs regardless of the initialization method.

The study concludes that initialization technique has no effect if the data is varied by the number of clusters, but would still help overall to get better results than the naïve initialization techniques like Random Centroid and Random Partition.

*C. Dimension*

Dimensions do not have a direct impact. The only indirect impact it imposes is that when the dimensions are increased for a dataset, then by nature the dataset becomes more separated and as discussed above due to this separation of data the performance of K-Means deteriorates.

*D. Unbalance*

To explain the unbalance factor in the dataset using 2 initialization techniques i.e., Random Centroid & K-Means++ with the help of an unbalanced dataset.

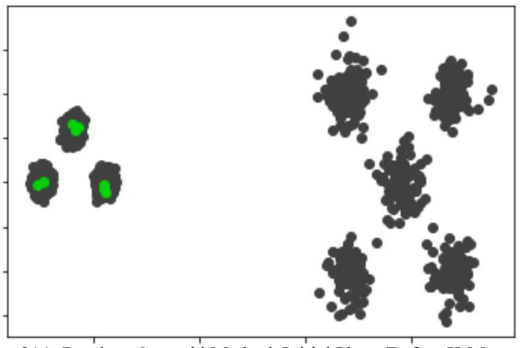


Fig. 9(a). Random Centroid Method, Initial Phase/Before K-Means.

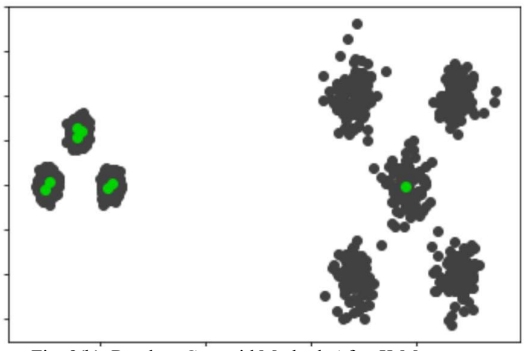


Fig. 9(b). Random Centroid Method, After K-Means.

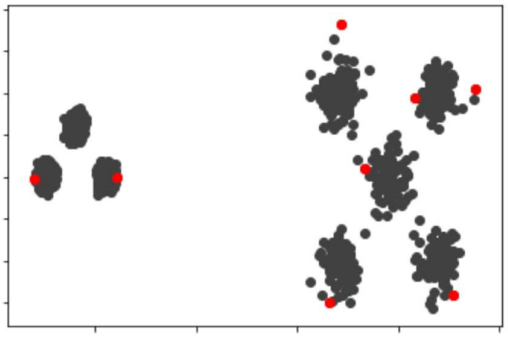


Fig. 10(a). K-Means++ Method, Initial Phase/Before K-Means.

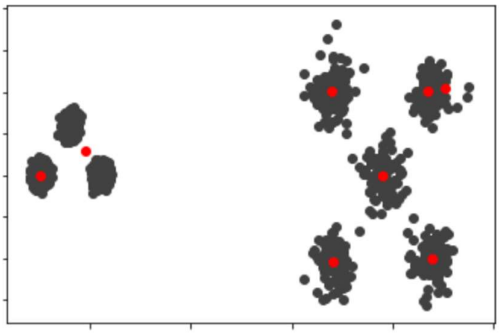


Fig. 10(b). K-Means++ Method, After K-Means.

Now to explain this it can be seen that there are 8 clusters in the dataset out of which 5 are sparse and 3 are dense. Now if Random Centroid is used for initialization as seen in Fig. 9(a), it shows that all the 8 initial centroids lie in the dense region and now running the K-Means generates the results shown in Fig. 9(b), which clearly shows how badly the Clustering algorithm performed but this is not due to K-Means, improper initialization is the reason behind this. Since no centroid was selected in the sparse area only 1 of the centroids was able to move out into the sparse area and the 4 more centroids that were required in sparse region remain inside the dense region.

Whereas in the case of K-Means++ the initial centroids were well placed both in dense as well as sparse region as shown in Fig 10(a), because of intelligence of K-Means++ initialization good results were generated as shown in Fig. 10(b) only 1 of the centroids was misplaced and all the others were placed fine. Further [20] showed that small clusters were been merged wrongly and on the other hand the larger ones split wrongly.

*E. Repeated K-Means*

From Table 3. and by definition [1] it can be interpreted that using Repeated K-Means would give the best result because the minimum value of SSE and CI-Value is chosen among all the repeats, for most of the datasets and initialization methods the study obtained CI=0 which is the best possible case for the algorithm. But as said above this privilege comes at a high cost and hence, must be used only in limit as it affects the complexity of the algorithm by a direct multiple. Further, it was seen in study that for some datasets 100 repeats can even go up to 30 minutes.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:41:15 UTC from IEEE Xplore. Restrictions apply.

## V. CONCLUSIONS

Finally, this research shows how much K-Means initialization is an important part of the K-Means Clustering Algorithm. It is now clear form above readings that how poor initialization can lead to very poor results.
It is seen that when Clusters are well separated the whole K-Means Algorithm depends completely on the initialization method used and hence find out that methods like Maxmin and K-Means++ gave much better results than Random Centroid and Random Partition. If the clusters are highly overlapping then the initialization method doesn't play a major role and the K-Means Algorithm would give good results for all the methods, and hence in these cases one can use Random Partition or Random Centroid method as well.

Further the study showed that K-Means has linear dependency with the number of clusters inside the dataset and deteriorates if there are high number of clusters independent of the type of initialization method used.

Dimensions of the dataset play an indirect role on K-Means because increasing the dimension separates the data further and hence, a good initialization method must be used now.

Unbalance in the dataset is bad for all the initialization methods but if using a good initialization technique, the algorithm can still produce very good results as compared to others.

To sum up the study concludes that no matter how the dataset is, one should use a good initialization like K-Means++ or Maxmin for K-Means Clustering Algorithm and along with it use the repeated K-Means to get more enhanced results but also make sure that too many repetitions do not deteriorate the time complexity of the overall Clustering Algorithm.
Various Standalone Algorithms like Luxburg [17] and Split [18,19] have also been developed which generate similar results to K-Means Clustering Algorithm [1].

## REFERENCES

.


[1] Pasi Fränti, Sami Sieranoja, "How much can k-means be improved by using better initialization and repeats?" Pattern Recognition, Volume 93, 2019, Pages 95-11.

[2] Bradley PS, Fayyad UM. "Refining initial points for k-means clustering". InICML 1998 Jul 24 (Vol. 98, pp. 91-99).

[3] Fränti, Pasi, Mohammad Rezaei and Qinpei Zhao. "Centroid index: Cluster level similarity measure." Pattern Recognit. 47 (2014): 3034-3045.

[4] Forgy EW. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. biometrics. 1965;21:768-9.

[5] MacQueen J. "Some methods for classification and analysis of multivariate observations." InProceedings of the fifth Berkeley symposium on mathematical statistics and probability 1967 Jun 21 (Vol. 1, No. 14, pp. 281-297).

[6] Fränti, P., Sieranoja, S. "K-means properties on six clustering benchmark datasets." Appl Intell 48, 4743–4759 (2018).

[7] J.M Peña, J.A Lozano, P Larrañaga, "An empirical comparison of four initialization methods for the K-Means algorithm," Pattern Recognition Letters, Volume 20, Issue 10, 1999, Pages 1027-1040.

[8] Hartigan, J. A., and M. A. Wong. "Algorithm AS 136: A K-Means Clustering Algorithm." Journal of the Royal Statistical Society. Series C (Applied Statistics) 28, no. 1 (1979): 100–108.

[9] Arthur D, Vassilvitskii S. k-means++: "The advantages of careful seeding." Stanford; 2006 Jun 7: 1027–1035.

[10] Lloyd S. Least squares quantization in PCM. IEEE transactions on information theory. 1982 Mar;28(2):129-137.

[11] Gonzalez, Teofilo F.. "Clustering to Minimize the Maximum Intercluster Distance." Theor. Comput. Sci. 38 (1985): 293-306.

[12] S. Sujatha and A. S. Sona, "New Fast K-Means Clustering Algorithm Using Modified Centroid Selection Method," International Journal of Engineering Research & Technology (IJERT), Vol. 2, No. 2, 2013, pp. 1-9.

[13] Kodinariya, Trupti. (2014). "Survey on Exiting Method for Selecting Initial Centroids in K-means Clustering," INTERNATIONAL JOURNAL OF ENGINEERING DEVELOPMENT AND RESEARCH. 2. 2865-2868.

[14] Singh, Harmanpreet & Kaur, Kamaljit. (2013). "New Method for Finding Initial Cluster Centroids in K-means Algorithm." International Journal of Computer Applications. 74. 27-30. 10.5120/12890-9837.

[15] Marco Capó, Aritz Pérez, Jose A. Lozano, "An efficient approximation to the K-means clustering for massive data," Knowledge-Based Systems, Volume 117, 2017, Pages 56-69,

[16] Celebi, M. Emre & Kingravi, Hassan & Vela, Patricio. (2013). "A Comparative Study of Efficient Initialization Methods for the K-Means Clustering Algorithm," Expert Systems with Applications. 40. 200–210. 10.1016/j.eswa.2012.07.021.

[17] U.V. Luxburg, Clustering stability: an overview, Found. Trends Mach. Learn. 2 (3) (2010) 235–274.

[18] P. Fränti, T. Kaukoranta, O. Nevalainen, On the splitting method for VQ code- book generation, Opt. Eng. 36 (11, November) (1997) 3043–3051.

[19] X. Wu, K. Zhang, A better tree-structured vector quantizer, in: IEEE Data Com- pression Conference, Snowbird, UT, 1991, pp. 392–401.

[20] L. Morissette, S. Chartier, "The k-means clustering technique: general consider- ations and implementation in Mathematica," Tutor. Quant. Methods Psychol. 9 (1) (2013) 15–24.


Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:41:15 UTC from IEEE Xplore. Restrictions apply.

Journal Pre-proof

# Improved resistive switching of RGO and $SnO_2$ based resistive memory device for non-volatile memory application

Km Komal, Govind Gupta, Mukhtiyar Singh, Bharti Singh

PII: S0925-8388(22)02587-7

DOI: https://doi.org/10.1016/j.jallcom.2022.166196

Reference: JALCOM166196

To appear in: *Journal of Alloys and Compounds*

Received date: 14 April 2022
Revised date: 21 June 2022
Accepted date: 3 July 2022

Please cite this article as: Km Komal, Govind Gupta, Mukhtiyar Singh and Bharti Singh, Improved resistive switching of RGO and $SnO_2$ based resistive memory device for non-volatile memory application, *Journal of Alloys and Compounds,* (2022) doi:https://doi.org/10.1016/j.jallcom.2022.166196

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.

© 2022 Published by Elsevier.

Journal Pre-proof

# Improved resistive switching of RGO and $SnO_2$ based resistive memory device for non-volatile memory application

Km. Komal[1], Govind Gupta[2], Mukhtiyar Singh[1], Bharti Singh[1*]

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

[2]Sensor Devices & Metrology, CSIR-National Physics Laboratory (CSIR), Delhi-110012, India

Corresponding Author e-mail addresses: bhartisingh@dtu.ac.in

mailto:komal_phd2k18@dtu.ac.in

## ABSTRACT

Resistive switching memory, which combines simple architectures with the ability of high density, high switching speed, and low power consumption, has sparked a lot of attention in the field of non-volatile memory devices. In the present work, the resistive switching performance of reduced graphene oxide (RGO) and tin oxide ($SnO_2$) based nanocomposite have been carried out. RGO, $SnO_2$, and RGO-$SnO_2$ nanocomposite were synthesized via hydrothermal technique and the formation of nanocomposite is confirmed by X-ray diffraction and Raman spectroscopy techniques. We herein present a bipolar resistive switching memory based on $SnO_2$ and RGO-$SnO_2$ fabricated by a simple spin-coating process. The two different memory devices have been fabricated by depositing pure $SnO_2$, and RGO-$SnO_2$ films over the bottom ITO electrode and top Al electrodes deposited through a shadow mask. It was observed that devices with RGO-$SnO_2$ showed improved resistive switching in comparison to pure $SnO_2$ film. This improvement of the composite-based film is recognized in terms of switching parameters, such as a reduction in the device operating voltage from 4.11 V for $SnO_2$ to 1V for RGO-$SnO_2$ composite-based devices. In addition, the resistance ratio in the composite-based films showed significant enhancement. Further, the retention and endurance behavior of the fabricated composite film-based device was tested up to $1*10^3$s and up to 100 cycles respectively which didn't exhibit any degradation. The operating mechanism of the memory device was explained by fitting the current-voltage characteristics in the low-resistance and high-resistance states. It may be highlighted that the present metal oxide and RGO-based devices have great potential for future non-volatile memory devices.

Journal Pre-proof

**Keywords:** Resistive switching, non-volatile memory, metal oxide, reduced graphene oxide, nanocomposite.

## INTRODUCTION

In the recent past year, the changing market trend from low power and data-centric devices has created the opportunity for the identification of various non-volatile memory devices. The major requirement for these non-volatile memories includes the high density of integration, high speed, low cost, simple structure, low operating power, long retention time, and compatibility with the existing CMOS technology [1-3]. It has been found that several candidates are competing for the next-generation memory, including magnetic random access memory (MRAM), ferroelectric random access memory (FRAM), and phase-change memory (PCM) [4-7]. MRAM for example utilizes magnetic tunnel junctions whereas, FRAM works on the principle of reversible polarisation of ferroelectric materials, respectively, and has gained a lot of attention and has seen a lot of development in the previous few years [8]. However, both MRAM and FRAM have serious scaling issues [9, 10]. In this situation, a new possibility emerged: resistance switching random access memory (ReRAM), in which the non-volatile memory effect arises by systematic switching of the memory cell into different resistance states named as high resistance state (HRS) and low resistance state (LRS)[11]. For observing the resistive switching phenomenon, the active insulating material is sandwiched between two metal electrodes in a metal-insulator-metal (MIM) structure. It may be mentioned that initially, the device needs to be activated by applying the initial voltage called electroforming voltage ($V_F$) resulting in a substantial shift in resistance to LRS. After that, the memory cell switches between a high resistance state (HRS) and a low resistance state (LRS) in a repeatable manner (LRS). The 'SET' process is used to switch from the HRS to the LRS, while the 'RESET' process is used to switch back [12], which would be used in RRAM devices. Bipolar resistive switching (RS) devices require both positive and negative voltages to switch the resistance state between HRS and LRS [13]. If just one polarity of the voltage is required to switch between HRS and LRS, the RS is known as unipolar in nature (URS). Various researchers have established the ability to control RS using voltage in oxides[14], nitrides[15], and organic materials[16]. It has been found in the literature that the binary transition metal oxides such as $HfO_2$, $TiO_2$, $SnO_2$, $ZnO$, $ZrO_2$, and carbon-based compounds are most favoured for the ReRAM potential material as compared to ternary or quaternary

oxide semiconductor films due to the ease of fabrication and compatibility with the CMOS industry [17].

Among all, $SnO_2$ (3.6 eV) has drawn much attention in the field of ReRAM devices due to its wide bandgap with visible optical transparency of up to 97 % and has large exciton binding energy (130 meV) [18]. **Over the past few years, $SnO_2$ has gathered the interest of many researchers in different fields of applications, as previously reported applications of $SnO_2$ in perovskite devices, such as photodetectors [19], Perovskite solar cells (PSCs) [20] [21, 22], etc. Due to excellent optoelectronic properties of metal halide perovskites, such as solution-processable property, low cost, long intrinsic carrier diffusion length, small exciton binding energy, tunable bandgap, high molar absorption coefficient, and so on, perovskite solar cells (PSCs) have received a lot of attention in recent years[22]. Zhuang et al. have established a bulk and interface defect passivation technique using the synergistic impact of anion and cation at the $SnO_2$/perovskite interface, using multifunctional $K_2SO_4$. $K^+$ ions in $K_2SO_4$ diffuse into the perovskite layer and suppress the production of bulk defects in perovskite films, while $SO_4^{2-}$ ions are anchored at the interface by strong chemical interactions with the $SnO_2$ and perovskite layers, respectively[21]. $SnO_2$-based planar PSCs have a validated PCE of 25.5 percent [23], demonstrating that they are highly promising.** In its bulk form, $SnO_2$ has a tetragonal rutile structure [24]. Its crystal structure and stoichiometry are easy to control than multinary compounds in the application of the ReRAM devices. Owing to its interesting electrical and optical properties, it has been widely studied for ReRAM devices. The several nanostructures of $SnO_2$ are nanocrystals, nanowires, nanobelts, and nanotubes. Among these structures, nanocrystals are technologically essential for ReRAM devices owing to their size-dependent features [25]. However, the resistive switching mechanism has not been fully explored and needs further work in the direction to modify the switching parameters. Therefore, one of the ways for the modification is to make nanocomposite based on semiconductor metal oxide and carbon-based material such as graphene, RGO, HBC, etc.

Among the various carbon-based materials, graphene is a fast-developing substance in material science and nanotechnology, due to its large specific surface area, conductivity and remarkable mobility, optical transparency, high mechanical flexibility, and chemical stability [26, 27]. **In literature, several others graphene-based photonic applications, such as imaging sensors, motion detectors, optical interconnectors, and optical modulators,**

**have also been reported [28-30]. Zhenhua et. al presented a systematic review for solar energy conversion by using the photodetection principle in graphene and other 2D-related materials. They have discussed that for increasing the photodetection response it's very important to understand the underlying physical mechanism (photoelectric, photo-thermoelectric, and photo-bolometric regimes) in addition to the device design and materials engineering [29]. Further, Yoo et. al has reported the graphene/P-Si based photodetector, whereby simple chemical doping of the graphene, the dark current was reduced by from orders of magnitude from 980 nA to 219 pA [28].** Several reports are available where pure reduced graphene oxide is used as active material in ReRAM and the device displays good retention time, and fast switching speed, but the current ratio is small in comparison to their metal oxide counterpart[31]. In this direction, the nanocomposites combining the properties of inorganic metal oxides and graphene-like 2D materials have been explored in literature, showing enhanced properties of the resultant device [32-34]. Along with this, it has been seen that several reports are available for metal oxide (MO) and RGO-based nanocomposite for photocatalytic application, energy storage, and gas sensing but there are few reports available for RGO-MO based nanocomposite for ReRAM application and as a result, motivates the present piece of work, where $SnO_2$ and RGO based nanocomposite film have been used an active material in MIM matrix. Several experimental techniques, including hydrothermal, chemical deposition, aerosol spraying, microwave synthesis, solvothermal, precipitation processes, and sol-gel auto combustion synthesis, have been intensively studied for binding of $SnO_2$ with RGO [35, 36]. Among all, hydrothermal synthesis is regarded as the most favourable of these procedures as it is cost-effective, uses mild synthesis conditions, and consumes little energy with basic equipment to generate homogeneous $SnO_2$ nanoparticles over RGO nanosheets, [37, 38]. $SnO_2$ and RGO-$SnO_2$ nanocomposites were synthesized by hydrothermal technique and turned into devices by uniformly spin coating on ITO substrate in the present study. Furthermore, the charge transport mechanism in the resultant Al/$SnO_2$/ITO and Al/RGO-$SnO_2$/ITO nanocomposite films have been thoroughly investigated in preparation for their use in memory devices.

## 2. EXPERIMENTAL

### 2.1 Chemical reagents

The chemicals used in the synthesis are tin (IV) chloride pentahydrate ($SnCl_4.5H_2O$), oleic acid, oleylamine, toluene, graphite powder, concentrated sulphuric acid ($H_2SO_4$), hydrazine

hydrate ($N_2H_4.H_2O$), hydrogen peroxide ($H_2O_2$), hydrogen chloride, potassium permanganate ($KMnO_4$), orthophosphoric acid ($H_3PO_4$) and ethanol ($C_2H_5OH$). Sigma Aldrich and Fisher Scientific are the suppliers of all the chemicals. All of them are of analytical grade quality and can be used without further purification.

### 2.2 Materials Synthesis

Fig. 1 shows a flow diagram for material synthesis used in the present work. A modified Hummers process was used to synthesize graphene oxide powder (GO) [39]. Furthermore, 1.8 g of $SnCl_4.5H_2O$ was dissolved in 60 mL oleic acid and 7.5 mL oleylamine for the formation of RGO-$SnO_2$ nanocomposite. After that, 3 mg of RGO (in 3.0 mL $H_2O$) and 30 mL ethanol was added and stirred for 5 hours at room temperature. The above mixture was further placed in a stainless-steel autoclave lined with Teflon and heated at 180°C for 8 hours. Moreover, the autoclave was allowed to come down to ambient temperature. After that, the product was rinsed with toluene and removed the toluene through a rotary evaporator, and then centrifuged (3 times). Then, the collected paste was dried in the oven at 50°C for 4 hours and after that, the sample is again dried for 24 hours at 400°C to obtain RGO-$SnO_2$ nanocomposite powder [40].

### 2.3 Film fabrication

The thin film of RGO-$SnO_2$ nanocomposite was produced on rectangular indium-tin-oxide (ITO) coated glass substrates. Initially, cleaning of the ITO substrates was done using a regular wet cleaning process. After the cleaning, the substrates were dried for 20 minutes at 80 degrees Celsius on a hot plate. The spin-coating deposition was done by the spin-coater at ambient temperature (25°C) which includes the following steps: (1) 50 mg of RGO-$SnO_2$ powder was dissolved in 2.5 ml ethanol solution in an ultrasonic bath for 30 minutes and then four drops of RGO-$SnO_2$ solution were dropped onto the ITO-glass substrate (2cm x 1cm) and then spun at 1250 rpm for 20 seconds [41]. (2) The coated sample was then heated on the hot plate at 80°C for 5 minutes. This step was repeated 7 times and we get a uniform layer of RGO-$SnO_2$. Finally, a 100 nm thick top Al electrode was deposited through the shadow mask with the help of the thermal evaporation technique under the pressure of $4x10^{-5}$ bar. The Schematic representation of the fabrication of the Al/RGO-$SnO_2$/ITO- based MIM device is given in Fig. 2.

## 3. RESULTS AND DISCUSSION

All the prepared samples are systematically characterized using Raman Spectrometer (WiTec alpha 300RA) with a laser source of wavelength 530 nm. An X-ray diffractometer study has been used (Bruker D8 Advance) to study the crystallinity of the RGO, $SnO_2$, and RGO-$SnO_2$ nanocomposite, with a Cu Kα radiation source (1.54 Å). The surface morphologies of the RGO, $SnO_2$, and RGO-$SnO_2$ nanocomposite were characterized by scanning electron microscopy (SEM; LYRA3 TESCAN). The bandgap of the as-prepared samples was analyzed using UV-vis absorption spectra (PerkinElmer Lambda 750 UVvis/NIR spectrophotometer) and photoluminescence spectra were carried out using a spectrofluorophotometer (RF-5301PC). Interfacial investigation of RGO, $SnO_2$, and RGO-$SnO_2$ nanocomposite structures was carried out by X-ray photoelectron spectroscopy (XPS) technique Perkin Elmer ultra-high vacuum XPS system (Model 1257 PHI) using a monochromatic Al Kα source. To fabricate the memory device, thin films were deposited by using the spin-coater, and Al electrodes of thickness ~100 nm were deposited using the thermal evaporation technique at a chamber pressure of $4x10^{-5}$ bar. Finally, the (I–V) measurements were carried out for all the prepared thin film-based memory devices (Al/$SnO_2$/ITO, Al/RGO-$SnO_2$/ITO) by using Keithley 2450 source meter.

### 3.1 X-Ray Diffraction (XRD)

The crystalline phases and crystallite size of pure RGO, pure $SnO_2$, and RGO-$SnO_2$ nanocomposite were studied using the XRD technique. XRD spectra confirm the presence of RGO and $SnO_2$ in the nanocomposite of RGO-$SnO_2$, as shown in Figure 3. The XRD spectra of pure $SnO_2$ confirm the presence of tetragonal rutile structure of $SnO_2$ with the occurrence of peaks at $26^o$, $34^0$, $52^0$, and $65^o$, which correspond to (110), (101), (211), and (202) planes. Similarly, the peak observed at $26^o$ corresponding to the (002) plane confirms the presence of RGO. Furthermore, the RGO-$SnO_2$ nanocomposite confirms the existence of the tetragonal rutile phase of $SnO_2$ with the presence of peaks at $26^0$, $34^o$, and $52^o$ attributes to (110), (101), and (211) planes. In the case of RGO-$SnO_2$ nanocomposite, the peaks of pure RGO are not observed owing to the low crystallinity of RGO. In the composite sample, the RGO peak observed at $26^o$ (002) is dominated by the diffraction peak of pure $SnO_2$ corresponding to the (110) plane. It may be attributed to the conversion of $Sn^{4+}$ into $SnO_2$ nanoparticles in the $SnO_2$-RGO sample **[42]**. The broadening in the diffraction peaks of $SnO_2$ was observed with the presence of RGO. It confirms the high crystalline nature of $SnO_2$ particles with the presence of RGO. As crystallinity depends on crystallite size, the crystallite size was

calculated by using the Debye-Scherer formula for pure RGO, pure $SnO_2$, and RGO-$SnO_2$ nanocomposite [41, 43];

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

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 calculated average crystal size for pure $SnO_2$, pure RGO, and RGO-$SnO_2$ nanocomposite are found to be 12.6 nm, 14.7 nm, and 8.65 nm, respectively. It has been observed that the average crystallite size of RGO- $SnO_2$ nanocomposite decreases owing to the growth and nucleation of $SnO_2$ crystals, which was hindered by RGO [44, 45].

### 3.2 Raman Spectroscopy

Raman scattering was used to investigate the structure of RGO-$SnO_2$ nanocomposites. The Raman spectra of pure $SnO_2$, pure RGO, and RGO-$SnO_2$ composites are shown in Fig. 4. The Raman spectra of the pure RGO display two major peaks at 1344 $cm^{-1}$ and 1574 $cm^{-1}$, corresponding to the characteristic D-band and G-band respectively. The G-band is induced by vibrations of $sp^2$ hybridized carbon atoms, while the D-band is driven by structural instability at defect sites. The other less intense peak was observed at 2689 $cm^{-1}$, which corresponds to the second-order D band. In the case of $SnO_2$, three peaks are present at 475.9 $cm^{-1}$, 631 $cm^{-1}$, and 773.3 $cm^{-1}$, which corresponds to the $E_g$, $A_{1g,}$ and $B_{2g}$ modes of $SnO_2$. Similarly, in the case of RGO-$SnO_2$ nanocomposite, a shift of 556 $cm^{-1}$, 629.9 $cm^{-1}$, and 773.5 $cm^{-1}$ in the peak position at 1356 $cm^{-1}$ and 1606 $cm^{-1}$ is found for D and G band, respectively. The shift in the peak position of nanocomposite corresponds to the interactions between RGO and $SnO_2$ nanoparticles during the hydrothermal synthesis process [45]. Furthermore, the intensity ratio ($I_D/I_G$) of D and G bands is 1.02 for pure RGO which decreases to 0.83 in the RGO-$SnO_2$ nanocomposite.

### 3.3 UV-Visible Spectroscopy

UV-Vis spectroscopy was employed to measure the degree of oxidation for synthesized samples. Fig. 5 depicts the UV-Vis absorption spectra of pure RGO, pure $SnO_{2,}$ and RGO-$SnO_2$ nanocomposite. The absorption peak for RGO was found at 271 nm. This absorption

peak arises due to the π-π* transitions of C-C bonds present in graphitic material [39]. Furthermore, in the case of pure $SnO_2$, the characteristic absorption peak was found at 258 nm. Similarly, the UV-Vis spectrum of RGO-$SnO_2$ nanocomposite revealed an absorption peak at 270 nm. The absorption peak at 270 nm is ascribed to the RGO-$SnO_2$ nanocomposite, which confirms the presence of the nanocomposite.

Furthermore, the bandgap of the synthesized samples was evaluated by Tauc's plot;

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

Where A is a constant, $E_g$ is the bandgap and α is the absorption coefficient. The optical bandgap for pure $SnO_2$, pure RGO, and RGO-$SnO_2$ nanocomposite was calculated by the plot between $(\alpha h\nu)^2$ vs hν [46].

The calculated bandgap of RGO and $SnO_2$ was 1.90 eV, and 3.63 eV, respectively. Similarly, the bandgap increases to 4.07eV for RGO-$SnO_2$ nanocomposite. The increase in the bandgap of RGO-$SnO_2$ nanocomposites is due to the incorporation of RGO which limits the radial growth of $SnO_2$ [40],[47].

### 3.4 Scanning Electron Microscopy (SEM)

SEM was used to examine the morphology of the pure RGO, pure $SnO_2$, and RGO-$SnO_2$ nanocomposite. The SEM image of RGO exhibits wrinkled and layered flakes on its surface as shown in fig. 6a. It also exhibits randomly aggregated RGO sheets. Similarly, fig. 6b shows the SEM image of nanocrystalline $SnO_2$. It shows the homogeneous and uniform distribution of $SnO_2$ nanocrystallites. Also, the SEM image of the RGO-$SnO_2$ nanocomposite demonstrates that $SnO_2$ nanoparticles are connected to the surface of RGO sheets, as shown in fig. 6c. It indicates that hydrothermal treatment of RGO and $SnCl_4$ solution is a successful approach for the synthesis of the RGO-$SnO_2$ nanocomposite [39].

### 3.5 Photoluminescence Spectra (PL)

Fig. 7 depicts the photoluminescence spectra of pure RGO, pure $SnO_2$, and RGO-$SnO_2$ nanocomposites. It has been observed Graphite (semiconducting material) does not own luminescence properties. Quantum mechanics explains the opening of the bandgap. The luminescence phenomenon occurs due to nano-carbon clusters formed in RGO sheets. These RGO sheets' oxide functional group and carbon vacancies are responsible for this [46]. In the case of pure RGO, excitation peaks are observed at 467, 482, 493, and 605nm ($\lambda_{em}$ = 400

nm), as shown in Fig. 7a. Similarly, four emission peaks are observed at 400, 465, 481, and 490 nm ($\lambda_{ex}$ = 605 nm), as shown in Fig. 7b [46]. Reduction of GO leads to a decrease in oxide functional group and expansion of sp2 carbon clusters. Excitation and emission spectroscopy techniques were used to reveal the defect states. **The PL properties of pure RGO are largely dependent on the types and amounts of final oxidation groups on its surface, which are determined by the degree of reduction during the synthesis process [48]. The $\pi$-$\pi^*$ transition of localized electrons trapped in substantially larger-sized sub-domains of $sp^2$-linked carbons of the RGO[49] may be responsible for the green emission at excitation wavelengths in the spectral range 450–700 nm. As a result, the observed behavior of PL emission in RGO at 605 nm excitation wavelengths could be due to localized electron transitions in different sized sub-domains of the $sp^2$ bonded carbons in the sheet with variable highest occupied molecular orbitals to lowest unoccupied molecular orbitals (HOMO-LUMO) gaps [50].** Furthermore, Fig. 7(c) and 7(d) shows the emission spectra and room-temperature excitation spectra of $SnO_2$ nanoparticles. The excitation spectra of $SnO_2$ show strong broadband at 136 nm ($\lambda_{em}$ = 244 nm) is visible, consistent with absorption spectra [42]. One strong band in emission spectra is observed at 25.4 nm ($\lambda_{ex}$= 398nm)[51]. PL spectra of the RGO-$SnO_2$ nanocomposite are shown in fig 7(e) and 7(f). The emission peaks of nanocomposite are observed at 447, 465, 479, and 489 nm ($\lambda_{ex}$ = 369 nm). Similarly, the excitation spectra are observed at 450, 467, 483, and 492 nm ($\lambda_{em}$=376 nm). An enhancement in photoluminescence intensity of RGO-$SnO_2$ nanocomposite is due to photo-generated $SnO_2$ to RGO charge separation and rapid charge transfer. From these results, it can be proposed that the synergetic effect of RGO-$SnO_2$ nanocomposite is due to the favorable $e^-$ transfer across the interface.

### 3.6 X-ray Photoelectron spectroscopy (XPS):

The surface characterization technique used to examine the elemental and compositional information of the samples is XPS. The XPS spectra of pure RGO, pure $SnO_2$ and RGO-$SnO_2$ nanocomposite are shown in Fig. 8. It confirms the presence of C, Sn, and O, no other impurities could be observed.

### Carbon Spectra

Fig. 9a shows the XPS results of the C1s present in pure RGO and RGO-$SnO_2$ nanocomposite. In the case of pure RGO, three characteristic peaks are observed at 284.6 eV, 286.1 eV, and 288.8 eV, corresponding to C-C, C-O, and C=O, respectively [52]. In the case

of RGO-$SnO_2$ nanocomposites, peaks are present at 284.6, 286.1, and 289.1 eV, which correspond to the $sp^2$ C, C–O, and C=O groups, respectively. In the RGO-$SnO_2$ composite, the peak position of the C=O (289.1 eV) group shifted towards the higher binding energy side. After the hydrothermal process, carbon-oxygen bond contributions to the C1s peak are drastically reduced. The shift towards the higher binding energy side corresponds to electronic interaction at the interface of RGO and $SnO_2$ [53].

**Oxygen Spectra**

The oxygen spectrum of pure RGO, pure $SnO_2$, and RGO-$SnO_2$ nanocomposite are shown in Fig. 9b. In the case of RGO, the peak at 531.7 eV is attributable to oxygen-containing functional groups present as a result of the hydrazine reduction phase during RGO synthesis. Similarly, two deconvoluted peaks are fitted for pure $SnO_2$. The peak in the O1s spectra at 531.6 eV corresponds to an extra oxygen species in the tin oxide sample. Another detected peak at 533.2 eV, suggests that the surface is significantly surrounded by hydroxide OH groups and corresponds to non-lattice oxygen [54]. Furthermore, in the case of RGO-$SnO_2$ nanocomposite, the O1s peak of non-lattice oxygen is shifted towards a lower binding energy side with a peak position of 532.8 eV. The shift in the peaks of O1s is shown in Table 3. The shift in the O1s peak shows an interfacial interaction, most likely between carbon from reduced graphene oxide and oxygen species from $SnO_2$ [55]. The O1s (RGO-$SnO_2$) spectra show narrower width of RGO which is exhibited due to the latter, i.e., at 531.6 eV, which might arise from adsorbed oxygen on the sample surface.

**Tin Spectra**

Fig. 9c shows the XPS spectra of Sn 3d for pure $SnO_2$ and RGO-$SnO_2$ nanocomposite. In the case of pure $SnO_2$, two characteristic peaks of Sn $3d_{5/2}$ and Sn $3d_{3/2}$ are observed at 484.5 eV and 493 eV respectively, which can be seen, demonstrating that tin exists in the form of $Sn^{4+}$ [56]. While, in the Sn 3d spectra of RGO-$SnO_2$ nanocomposite, the XPS peaks are observed at 485.4 eV and 493.9 eV, which shows a higher binding energy shift of 1 eV and 0.9 eV for Sn $3d_{5/2}$ and Sn $3d_{3/2}$ peaks respectively. The peak-to-peak spacing is 8.4 eV between the characteristic peaks for $SnO_2$ and is consistent with previous findings. The above-observed shift indicates that the $SnO_2$ nanoparticles are connected to the RGO nanosheets, and the same has been confirmed in a study[57]. The formation of the Sn-C-O connection between $SnO_2$ and RGO indicated the presence of a strong interaction between them, which was crucial in improving resistive switching performance. As a result, the Sn3d peaks from RGO-

$SnO_2$ are moved to the higher binding energy**[58].** Because $SnO_2$ has a smaller work function than RGO, the electrons in RGO-$SnO_2$ hybrids can generally flow ($SnO_2$ to RGO)[59]. As a result, the Sn3d peaks for RGO-$SnO_2$ have been moved to have greater binding energy. The electrostatic repulsive contact between O and a negative charge in the lattice allows Sn to produce more energy. The binding energy of O1s will be affected by the change in electronic orbit for Sn [60].

**Table 1.** Binding energy (eV) of pure RGO, pure $SnO_2$, and RGO-$SnO_2$ nanocomposite.

| **Sample Name** | **C 1s (eV)** | | | **O 1s (eV)** | | **Sn 3d (eV)** | |
|---|---|---|---|---|---|---|---|
| | I | II | III | I | II | I | II |
| RGO | 284.6 | 286.1 | 288.8 | 531.7 | - | - | - |
| $SnO_2$ | - | - | - | 531.6 | 533.2 | 484.5 | 493 |
| RGO-$SnO_2$ | 284.6 | 286.1 | 289.1 | 531.6 | 532.8 | 485.4 | 493.9 |

The potential barrier at the interface plays an important role in the enhancement of memory device properties. Owing to the interfacial potential barrier, the liberated electrons will be transported between $SnO_2$ and RGO. This interaction leads to the modification of the interface and hence enrichment of the memory device properties. The interaction between RGO and $SnO_2$ is inferred by the energy band model.

$SnO_2$ and RGO have work functions of 4.55 eV and 4.75 eV, respectively. Because $SnO_2$ has a lower work function than RGO, the electrons will move from $SnO_2$ to RGO to equalize the Fermi level and produce a potential barrier (0.2 eV) at the interface, as illustrated in Fig. 10. The potential barrier between RGO and $SnO_2$ forms due to fermi levels equilibrium, resulting in a built-in electric field at the interface that favours the separation of charges and hence improves in properties of the memory device.

### 3.6 Switching characteristics of Al/$SnO_2$/ITO and Al/RGO-$SnO_2$/ITO devices

The switching behavior of pure $SnO_2$ and RGO-$SnO_2$ nanocomposite is shown in Fig. 11a-c. In order to obtain the switching behavior of $SnO_2$ and RGO-$SnO_2$, current-voltage (I-V) measurements were carried out by using Keithley source meter 2450. I-V characteristics were attained by the fabrication of metal-semiconductor-metal memory devices, as illustrated in Fig. 2c. To acquire the switching characteristics of fabricated memory devices (Al/$SnO_2$/ITO

and Al/RGO-$SnO_2$/ITO), sweeping the voltage from +15 V to -15 V and +5 V to -5 V, respectively. In the first cycle, the upper and lower electrode of the memory device was connected with negative and positive potential, respectively.

Fig. 11a shows the bipolar resistive switching characteristics of the Al/$SnO_2$/ITO device with a logarithmic scale in the inset. The initial state of $SnO_2$ was a high resistance state (HRS). By sweeping the potential from 0 V to 15 V, the device switch to the low resistance state (LRS) at 11.14 eV. Similarly, when sweeping the potential from -0 V to -15 V, the device switch back to the HRS state at -10.30 V. Therefore, our device exhibits the bipolar resistive switching behaviour. Further, the switching phenomenon of the RGO-$SnO_2$ based nanocomposite is shown in Fig. 11b. The RGO-$SnO_2$ nanocomposite-based device exhibits the transition from HRS to LRS at 4.93 V and LRS to HRS transition at -3.21 V. Similarly, the hysteresis area is higher than that of pristine $SnO_2$. From these results, it is observed that the RGO-$SnO_2$ nanocomposite-based device exhibits an enhancement in switching performance. Moreover, the $I_{ON}/I_{OFF}$ ratio for RGO-$SnO_2$ nanocomposite-based devices is 4 times higher than pure $SnO_2$-based memory devices thus indicating the improved resistive switching characteristics of the nanocomposite device. **The improved resistive switching ratio ($I_{ON}/I_{OFF}$) in the RGO-$SnO_2$ based memory device in comparison to pure $SnO_2$, may be attributed to the interfacial interaction between the RGO and $SnO_2$ nanoparticles. It has been reported in the literature that reduced graphene oxide and metal oxide hybridization greatly influences the oxygen vacancies concentration in the metal oxides [61]. Further, it may be mentioned that the composite formation between graphene oxide and tin oxide also greatly enhances the chemical composition, as shown in the XPS spectra, which also helps in the switching parameter enhancement[62]. Singh et. al reported the dual benefit of RGO based metal oxide nanocomposite for resistive memory application where, it not only acts as oxygen reservoir as well as the blocking layer for preventing the diffusion of oxygen from switching layers, thereby enhancing the switching ratio[55].**

Furthermore, to understand the mechanism of resistive switching in Al/RGO-$SnO_2$/ITO, various conduction models were used to fit the I–V plot. Fig. 11c shows the log (I) vs. log (V) curve of the RGO-$SnO_2$ nanocomposite. In the LRS state, the log (I) vs. log (V) curve indicates ohmic conduction with a slope value of ~1.3, whereas in HRS, at the small voltage the curve exhibits the linear behavior with a slope of ~1.2 and at the higher voltage the curve exhibits the slope of ~1.9, indicating that space charge limited current conduction mechanism

(SCLC) may be responsible for observed resistive switching[55, 63]. The mechanism of SCLC takes place when the recombination rate of electrons is less than implanting electrons inside the conduction band of material. This mechanism of SCLC describes the switching behavior from a current-voltage relationship where the dielectric layer is set off by the insufficiency of oxygen [13]. SCLC mechanism explains that the change of resistance state in the switching device occurs due to the formation of the conduction path. The formation of the conduction path is assisted by the development of electron confinements due to the formation of oxygen vacancies in the RGO-$SnO_2$ layer [63]. These results indicate that the formation of oxygen vacancies near the RGO-$SnO_2$ interface enhances the electric field, which increases the memory device performance in comparison to pure $SnO_2$. Hence, the observed bipolar resistive switching may be attributed to the electric field-assisted trapping and de-trapping of oxygen ions/vacancies at the RGO-$SnO_2$ interface.

To investigate the stability of the device (Al/RGO-$SnO_2$/ITO), endurance tests with multiple cycles were performed. It was found that the performance of the device remains stable after 100 cycles, as shown in fig. 12a. For set and reset, the sweeping voltage of 0 to +5 V and 0 to -5 V was applied, respectively. The voltage of 1 V was applied to measure the ON and OFF resistance at room temperature. The memory window of the device was calculated by $R_{OFF}/R_{ON}$ and it is found to be 5, as shown in Fig. 12b. From these results, it is evident that there is no degradation of the Al/RGO-$SnO_2$/ITO device even after a repetitive cycle. These results confirm the switching between off and on states is reversible and repeatable. For an aggregate wait time of $1*10^3$ s, both HRS and LRS are unaffected at the voltage of 1 V with a steady resistance ratio of LRS/HRS over 100 cycles. The above results confirm that the Al/RGO-$SnO_2$/ITO device exhibits improved switching performance with consistency and durability in switching between HRS and LRS. It may be due to the rise in the oxygen vacancies in the nanocomposite-based structure, as confirmed by XPS. The increased oxygen vacancies in the nanocomposite increase the memory device performance in comparison to pure $SnO_2$.

Furthermore, the switching mechanism of the device is shown in Fig. 13. The low resistance state (LRS) and high resistance state (HRS) mechanism can be explained by the creation and rupture of conducting filaments between the electrodes. In the case of the formation of conducting filament, when an electric potential is applied between the top and bottom electrode, oxygen vacancies present in the pure $SnO_2$ and RGO-$SnO_2$ nanocomposite help in

the formation of filament. The performance of the ReRAM device depends on the dimension of the conduction filaments as well as on the oxygen vacancies present in the insulating layers. The switching behaviour of pure $SnO_2$ and RGO-$SnO_2$ nanocomposite is shown in fig. 13a and 13d. From the Figure, it is clear that the number of oxygen vacancies is directly proportional to the concentration of RGO. Furthermore, in the presence of an electric field, oxygen vacancies are aligned to form conducting filaments with conical shapes as shown in fig. 13b and 13e. When the reverse polarity is applied on the top electrode, the depletion of oxygen vacancies is observed towards the top electrode, which may be due to filament rupturing as shown in fig. 13c and 13f. According to the literature, the size of the filament is directly proportional to the number of oxygen vacancies present inside the insulating layer. A large number of oxygen vacancies implies the formation of a broader filament. Thus, oxygen vacancies play an important role in the performance of resistive switching devices [64, 65] and therefore have motivated to carry out the weight percentage study of RGO in the $SnO_2$ matrix.

## 4. Conclusion

In conclusion, pure $SnO_2$ and RGO-$SnO_2$ nanocomposites were studied for resistive memory device application, where, RGO-$SnO_2$ nanocomposite powder was synthesized using the hydrothermal method. The synthesis of RGO-$SnO_2$ nanocomposite was confirmed by SEM, XRD, and Raman spectroscopy techniques. The shift in C1s, O1s and Sn 3d peaks in the RGO-$SnO_2$ nanocomposite in comparison to pure $SnO_2$ and RGO indicate an interfacial interaction. Reduction in switching parameters, and improvement in the current ratio further confirmed the modified electronic properties in the RGO-$SnO_2$ nanocomposite-based resistive memory device. Multiple electron transport mechanism responsible for current transport also indicates that $SnO_2$, RGO, and $SnO_2$-RGO interface together determine the electrical properties of the device, especially in HRS. However, in LRS the electrical resistance is mostly governed by the oxygen vacancies induced by metallic filaments. The fabricated memory device demonstrated improved switching performance, with roughly 100 cycles of endurance and $10^3$-second stable state retention. The present study demonstrated an interesting application of reduced graphene oxide for improving the switching properties of tin oxide-based resistive memory devices.

## 5. Acknowledgement

The authors thank the CSIR, Govt. of India for providing the fellowship during the research work and also thank Advance Instrumentation Centre, Department of Applied Physics, Delhi Technological University.

**References**

[1] B. Singh, B. Mehta, D. Varandani, A.V. Savu, J. Brugger, CAFM investigations of filamentary conduction in Cu2O ReRAM devices fabricated using stencil lithography technique, Nanotechnology, 23 (2012) 495707.
[2] A. Chen, A review of emerging non-volatile memory (NVM) technologies and applications, Solid-State Electronics, 125 (2016) 25-38.
[3] R. Khosla, S.K. Sharma, Integration of ferroelectric materials: An ultimate solution for next-generation computing and storage devices, ACS Applied Electronic Materials, 3 (2021) 2862-2897.
[4] J. Boukhobza, S. Rubini, R. Chen, Z. Shao, Emerging NVM: A survey on architectural integration and research challenges, ACM Transactions on Design Automation of Electronic Systems (TODAES), 23 (2017) 1-32.
[5] J.S. Meena, S.M. Sze, U. Chand, T.-Y. Tseng, Overview of emerging nonvolatile memory technologies, Nanoscale research letters, 9 (2014) 1-33.
[6] Q. Cao, W. Lü, X.R. Wang, X. Guan, L. Wang, S. Yan, T. Wu, X. Wang, Nonvolatile multistates memories for high-density data storage, ACS Applied Materials & Interfaces, 12 (2020) 42449-42471.
[7] E.C. Ahn, H.-S.P. Wong, E. Pop, Carbon nanomaterials for non-volatile memories, Nature Reviews Materials, 3 (2018) 1-15.
[8] P. Feng, C. Chao, Z.-s. Wang, Y.-c. Yang, Y. Jing, Z. Fei, Nonvolatile resistive switching memories-characteristics, mechanisms and challenges, Progress in Natural Science: Materials International, 20 (2010) 1-15.
[9] R. Guo, L. You, Y. Zhou, Z.S. Lim, X. Zou, L. Chen, R. Ramesh, J. Wang, Non-volatile memory based on the ferroelectric photovoltaic effect, Nature communications, 4 (2013) 1-5.
[10] A. Rani, D.B. Velusamy, R.H. Kim, K. Chung, F.M. Mota, C. Park, D.H. Kim, Non-Volatile ReRAM Devices Based on Self-Assembled Multilayers of Modified Graphene Oxide 2D Nanosheets, Small, 12 (2016) 6167-6174.
[11] P. Noé, C. Vallée, F. Hippert, F. Fillot, J.-Y. Raty, Phase-change materials for non-volatile memory devices: from technological challenges to materials science issues, Semiconductor Science and Technology, 33 (2017) 013002.
[12] M. Ahmad, M. Rafiq, Z. Imran, K. Rasool, R. Shahid, Y. Javed, M. Hasan, Charge conduction and relaxation in MoS2 nanoflakes synthesized by simple solid state reaction, Journal of Applied Physics, 114 (2013) 043710.
[13] D.P. Sahu, S.N. Jammalamadaka, Remote control of resistive switching in TiO 2 based resistive random access memory device, Scientific reports, 7 (2017) 1-8.
[14] S. Long, L. Perniola, C. Cagli, J. Buckley, X. Lian, E. Miranda, F. Pan, M. Liu, J. Suñé, Voltage and power-controlled regimes in the progressive unipolar RESET transition of HfO2-based RRAM, Scientific reports, 3 (2013) 1-8.
[15] C. Pan, Y. Ji, N. Xiao, F. Hui, K. Tang, Y. Guo, X. Xie, F.M. Puglisi, L. Larcher, E. Miranda, Coexistence of grain-boundaries-assisted bipolar and threshold resistive switching in multilayer hexagonal boron nitride, Advanced functional materials, 27 (2017) 1604811.
[16] J. Park, K. Biju, S. Jung, W. Lee, J. Lee, S. Kim, S. Park, J. Shin, H. Hwang, Multibit Operation of $\hbox {TiO} _ {x} $-Based ReRAM by Schottky Barrier Height Engineering, IEEE Electron Device Letters, 32 (2011) 476-478.

[17] C.-H. Wang, Y.-H. Tsai, K.-C. Lin, M.-F. Chang, Y.-C. King, C.J. Lin, S.-S. Sheu, Y.-S. Chen, H.-Y. Lee, F.T. Chen, Three-Dimensional $\hbox {4F}^{2} $ ReRAM With Vertical BJT Driver by CMOS Logic Compatible Process, IEEE transactions on electron devices, 58 (2011) 2466-2472.
[18] S. Nilavazhagan, S. Muthukumaran, M. Ashokkumar, Structural, optical and morphological properties of La, Cu co-doped SnO2 nanocrystals by co-precipitation method, Optical Materials, 37 (2014) 425-432.
[19] H. Wang, P. Zhang, Z. Zang, High performance CsPbBr3 quantum dots photodetectors by using zinc oxide nanorods arrays as an electron-transport layer, Applied Physics Letters, 116 (2020) 162103.
[20] S. Wang, T. Wu, J. Qiu, R. Wang, Z. Zhu, X. Zhang, Y. Hua, Enhanced electron transfer dynamics in perylene diimide passivated efficient and stable perovskite solar cells, EcoMat, 3 (2021) e12146.
[21] Q. Zhuang, H. Wang, C. Zhang, C. Gong, H. Li, J. Chen, Z. Zang, Ion diffusion-induced double layer doping toward stable and efficient perovskite solar cells, Nano Research, (2022) 1-9.
[22] C. Zhang, H. Wang, H. Li, Q. Zhuang, C. Gong, X. Hu, W. Cai, S. Zhao, J. Chen, Z. Zang, Simultaneous passivation of bulk and interface defects through synergistic effect of anion and cation toward efficient and stable planar perovskite solar cells, Journal of Energy Chemistry, 63 (2021) 452-460.
[23] J.J. Yoo, G. Seo, M.R. Chua, T.G. Park, Y. Lu, F. Rotermund, Y.-K. Kim, C.S. Moon, N.J. Jeon, J.-P. Correa-Baena, Efficient perovskite solar cells via improved carrier management, Nature, 590 (2021) 587-593.
[24] M.-M. Bagheri-Mohagheghi, M. Shokooh-Saremi, Electrical, optical and structural properties of Li-doped SnO2 transparent conducting films deposited by the spray pyrolysis technique: a carrier-type conversion study, Semiconductor science and technology, 19 (2004) 764.
[25] N. Van Hieu, Comparative study of gas sensor performance of SnO2 nanowires and their hierarchical nanostructures, Sensors and Actuators B: Chemical, 150 (2010) 112-119.
[26] X. Li, J. Shan, W. Zhang, S. Su, L. Yuwen, L. Wang, Recent advances in synthesis and biomedical applications of two-dimensional transition metal dichalcogenide nanosheets, Small, 13 (2017) 1602660.
[27] J. Qiu, X. Guo, R. Chu, S. Wang, W. Zeng, L. Qu, Y. Zhao, F. Yan, G. Xing, Rapid-response, low detection limit, and high-sensitivity capacitive flexible tactile sensor based on three-dimensional porous dielectric layer for wearable electronic skin, ACS applied materials & interfaces, 11 (2019) 40716-40725.
[28] T.J. Yoo, S.Y. Kim, M.G. Kwon, C. Kim, K.E. Chang, H.J. Hwang, B.H. Lee, A Facile Method for Improving Detectivity of Graphene/p-Type Silicon Heterojunction Photodetector, Laser & Photonics Reviews, 15 (2021) 2000557.
[29] Z. Sun, H. Chang, Graphene and graphene-like two-dimensional materials in photodetection: mechanisms and methodology, ACS nano, 8 (2014) 4133-4156.
[30] D. Castelló-Lurbe, H. Thienpont, N. Vermeulen, Predicting Graphene's Nonlinear-Optical Refractive Response for Propagating Pulses, Laser & Photonics Reviews, 14 (2020) 1900402.
[31] C. Ban, X. Wang, Z. Zhou, H. Mao, S. Cheng, Z. Zhang, Z. Liu, H. Li, J. Liu, W. Huang, A universal strategy for stretchable polymer nonvolatile memory via tailoring nanostructured surfaces, Scientific reports, 9 (2019) 1-7.
[32] H. Wang, H. Feng, J. Li, Graphene and graphene-like layered transition metal dichalcogenides in energy conversion and storage, Small, 10 (2014) 2165-2181.

Journal Pre-proof

[33] S. Yang, C. Jiang, S.-h. Wei, Gas sensing in 2D materials, Applied Physics Reviews, 4 (2017) 021304.
[34] C. Zhu, D. Du, Y. Lin, Graphene and graphene-like 2D materials for optical biosensing and bioimaging: A review, 2D Materials, 2 (2015) 032004.
[35] Y. Zhou, Controllable design, synthesis and characterization of nanostructured rare earth metal oxides, Physical Sciences Reviews, 5 (2020).
[36] B.M. Reddy, P. Bharali, P. Saikia, A Comprehensive Overview on Synthesis Techniques of Nanostructured Oxides, New Nanotechniques, 243-276.
[37] R. Kumar, E. Joanni, R.K. Singh, D.P. Singh, S.A. Moshkalev, Recent advances in the synthesis and modification of carbon-based 2D materials for application in energy conversion and storage, Progress in Energy and Combustion Science, 67 (2018) 115-157.
[38] S. Shamaila, A.K.L. Sajjad, S.A. Farooqi, N. Jabeen, S. Majeed, I. Farooq, Advancements in nanoparticle fabrication by hazard free eco-friendly green routes, Applied Materials Today, 5 (2016) 150-199.
[39] N. Hidayah, W.-W. Liu, C.-W. Lai, N. Noriman, C.-S. Khe, U. Hashim, H.C. Lee, Comparison on graphite, graphene oxide and reduced graphene oxide: Synthesis and characterization, in: AIP Conference Proceedings, AIP Publishing, 2017, pp. 150002.
[40] Z. Song, Z. Wei, B. Wang, Z. Luo, S. Xu, W. Zhang, H. Yu, M. Li, Z. Huang, J. Zang, Sensitive room-temperature H2S gas sensors employing SnO2 quantum wire/reduced graphene oxide nanocomposites, Chemistry of Materials, 28 (2016) 1205-1212.
[41] D. Zhang, H. Chang, R. Liu, Humidity-sensing properties of one-step hydrothermally synthesized tin dioxide-decorated graphene nanocomposite on polyimide substrate, Journal of Electronic Materials, 45 (2016) 4275-4281.
[42] H. Le Trung, N.K. Tran, H.C. Manh, H.N. Duc, H.T.T. Quynh, T.H. Pham, Hydrothermal synthesis, structure, and photocatalytic properties of SnO2/rGO nanocomposites with different GO concentrations, Materials Research Express, 5 (2018) 095506.
[43] P. Tyagi, A. Sharma, M. Tomar, V. Gupta, A comparative study of RGO-SnO2 and MWCNT-SnO2 nanocomposites based SO2 gas sensors, Sensors and Actuators B: Chemical, 248 (2017) 980-986.
[44] D. Nath, S. Mandal, D. Deb, J. Rakshit, P. Dey, J. Roy, Light tuning DC and AC electrical properties of ZnO-rGO based hybrid nanocomposite film, Journal of Applied Physics, 123 (2018) 095115.
[45] N. Jain, S. Sharma, N.K. Puri, Investigation of charge transport mechanism in hydrothermally synthesized reduced graphene oxide (rGO) incorporated zinc oxide (ZnO) nanocomposite films, Journal of Materials Science: Materials in Electronics, (2021) 1-17.
[46] J. Ding, H. Chen, D. Feng, H. Fu, Investigation on photoluminescence emission of (reduced) graphene oxide paper, in: IOP Conf. Ser.: Mater. Sci. Eng, 2018, pp. 012097.
[47] T. Iqbal, M. Haqnawaz, M. Sultan, M.B. Tahir, M.I. Khan, K.N. Riaz, M. Ijaz, M. Rafique, Novel graphene-based transparent electrodes for perovskite solar cells, International Journal of Energy Research, 42 (2018) 4866-4874.
[48] X.-S. Hu, Y. Shen, Y.-T. Zhang, H.-F. Zhang, L.-H. Xu, Y.-J. Xing, Synthesis of flower-like CuS/reduced graphene oxide (RGO) composites with significantly enhanced photocatalytic performance, Journal of Alloys and Compounds, 695 (2017) 1778-1785.
[49] G. Eda, Y.Y. Lin, C. Mattevi, H. Yamaguchi, H.A. Chen, I.S. Chen, C.W. Chen, M. Chhowalla, Blue photoluminescence from chemically derived graphene oxide, Advanced materials, 22 (2010) 505-509.
[50] A. Singh, B. Ahmed, A. Singh, A.K. Ojha, Photodegradation of phenanthrene catalyzed by rGO sheets and disk like structures synthesized using sugar cane juice as a reducing agent, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 204 (2018) 603-610.

[51] F. Gu, S.F. Wang, C.F. Song, M.K. Lü, Y.X. Qi, G.J. Zhou, D. Xu, D.R. Yuan, Synthesis and luminescence properties of SnO2 nanoparticles, Chemical Physics Letters, 372 (2003) 451-454.
[52] Y. Chen, Y. Niu, T. Tian, J. Zhang, Y. Wang, Y. Li, L.-C. Qin, Microbial reduction of graphene oxide by Azotobacter chroococcum, Chemical Physics Letters, 677 (2017) 143-147.
[53] N. Garino, A. Sacco, M. Castellino, J.A. Muñoz-Tabares, A. Chiodoni, V. Agostino, V. Margaria, M. Gerosa, G. Massaglia, M. Quaglio, Microwave-assisted synthesis of reduced graphene oxide/SnO2 nanocomposite for oxygen reduction reaction in microbial fuel cells, ACS Applied Materials & Interfaces, 8 (2016) 4633-4643.
[54] B. Babu, I.N. Reddy, K. Yoo, D. Kim, J. Shim, Bandgap tuning and XPS study of SnO2 quantum dots, Materials Letters, 221 (2018) 211-215.
[55] B. Singh, B. Mehta, Govind, X. Feng, K. Müllen, Electronic interaction and bipolar resistive switching in copper oxide-multilayer graphene hybrid interface: Graphene as an oxygen ion storage and blocking layer, Applied Physics Letters, 99 (2011) 222109.
[56] H. Jadhav, S. Suryawanshi, M. More, S. Sinha, Pulsed laser deposition of tin oxide thin films for field emission studies, Applied Surface Science, 419 (2017) 764-769.
[57] S. Li, W. Xie, S. Wang, X. Jiang, S. Peng, D. He, Facile synthesis of rGO/SnO 2 composite anodes for lithium ion batteries, Journal of Materials Chemistry A, 2 (2014) 17139-17145.
[58] B. Singh, B. Mehta, M. Singh, G. Gupta, L. Dössel, X. Feng, K. Müllen, Structural and electronic interaction at CuO-hexa-peri-hexabenzocoronene hybrid interface, Applied Physics Letters, 98 (2011) 072101.
[59] D. Yang, R. Yang, K. Wang, C. Wu, X. Zhu, J. Feng, X. Ren, G. Fang, S. Priya, S.F. Liu, High efficiency planar-type perovskite solar cells with negligible hysteresis using EDTA-complexed SnO2, Nature communications, 9 (2018) 1-11.
[60] Z. Wang, T. Han, T. Fei, S. Liu, T. Zhang, Investigation of microstructure effect on NO2 sensors based on SnO2 nanoparticles/reduced graphene oxide hybrids, ACS applied materials & interfaces, 10 (2018) 41773-41783.
[61] U.B. Isyaku, M.H.B.M. Khir, I.M. Nawi, M. Zakariya, F. Zahoor, ZnO based resistive random access memory device: a prospective multifunctional next-generation memory, IEEE Access, (2021).
[62] G. Ahmed, M. Hanif, K. Mahmood, R. Yao, H. Ning, D. Jiao, M. Wu, J. Khan, Z. Liu, Lattice defects of ZnO and hybrids with GO: Characterization, EPR and optoelectronic properties, AIP Advances, 8 (2018) 025218.
[63] L.-G. Wang, Z.-Y. Cao, X. Qian, L. Zhu, D.-P. Cui, A.-D. Li, D. Wu, Atomic layer deposited oxide-based nanocomposite structures with embedded CoPt x nanocrystals for resistive random access memory applications, ACS Applied Materials & Interfaces, 9 (2017) 6634-6643.
[64] R. Singh, R. Kumar, A. Kumar, D. Kumar, M. Kumar, Enhanced resistive switching in graphene oxide based composite thin film for nonvolatile memory applications, Materials Research Express, 6 (2019) 105621.
[65] R. Vartak, A. Rag, S. De, S. Bhat, Study of Ag/RGO/ITO sandwich structure for resistive switching behavior deposited on plastic substrate, Applied Nanoscience, 8 (2018) 1343-1351.

Journal Pre-proof

**Figure caption:**

**Fig. 1.** Schematic diagram illustrating the various steps used in the hydrothermal synthesis of RGO-$SnO_2$ nanocomposite powder.

**Fig. 2.** Schematic diagram showing various steps used in resistive memory device fabrication (a) spin-coating (b) thermal deposition of top Al electrode through shadow mask (c) final MIM configuration.

**Fig. 3.** XRD diffractogram of RGO, $SnO_2$, and RGO-$SnO_2$ nanocomposite powder.

**Fig. 4.** Raman spectra of pure $SnO_2$, pure RGO, and RGO-$SnO_2$ nanocomposite.

**Fig. 5.** UV-Vis spectra of a) pure RGO nanosheets b) pure $SnO_2$ nanoparticle, and c) RGO-$SnO_2$ nanocomposite synthesized at 180°C for 8 hours, respectively.

**Fig. 6.** SEM micrographs of pure RGO nanosheets (a), pure $SnO_2$ nanoparticle (b), and RGO-$SnO_2$ (c) nanocomposite.

**Fig. 7.** Photoluminescence spectra of the pure $SnO_2$ nanoparticle, pure RGO nanosheets, and RGO-$SnO_2$ nanocomposite.

**Fig. 8.** XPS survey spectra of pure RGO, $SnO_2$ and RGO-$SnO_2$ nanocomposite powder indicated major peaks of C1s, O1s, and Sn3d.

**Fig. 9.** High-resolution core-level XPS spectra for pure RGO, pure $SnO_2$, and RGO-$SnO_2$ nanocomposite: (a) C1s peak; (b) O1s peak; (c) Sn 3d spectra. The larger binding energy shifts in the XPS spectra for RGO-SnO2 in comparison to pure RGO and $SnO_2$ are due to interaction at the RGO-$SnO_2$ interface.

**Fig. 10.** Schematic illustrating the Energy band diagram leading to the possible reason for enhanced memory properties for RGO-$SnO_2$ composites.

**Fig. 11.** I-V curve showing resistive switching characteristics of (a) Al/$SnO_2$/ITO and (b) Al/RGO-$SnO_2$/ITO devices. (c) I-V characteristics of Al/RGO-$SnO_2$/ITO show the linear ohmic behavior with a slope of about 1.2 in the LRS in a double-logarithmic plot. Linearity is also observed in HRS at lower voltages with a slope of about 1.2. Deviation from the linearity is observed in HRS at higher voltages with a slope of about 1.92.

**Fig. 12.** The improved memory behavior of the Al/RGO-$SnO_2$/ITO device is shown in terms of (a) endurance performance for about 100 switching cycles and (b) retention of resistances

in high resistance state (HRS) and low resistance state (LRS) for about 1000 seconds, thereby indicating good performance.

**Fig. 13.** (a, b) Schematic illustration of the distribution of oxygen vacancies, (c, d) Filament formation due to oxygen vacancy migration (e, f) Filament rupture in Al/$SnO_2$/ITO and Al/RGO-$SnO_2$/ITO based memory device respectively.

**CRediT authorship contribution statement**

**Km. Komal:** Conceptualization, Methodology, Investigation, Data Curation, Writing - Original Draft

**Dr. Bharti Singh:** Conceptualization, Methodology, Writing - Review & Editing at different stages, Supervision

**Dr. Govind Gupta:** Characterization, Visualization

**Dr. Mukhtiyar Singh:** Revising the manuscript critically for important intellectual content, Supervision

Journal Pre-proof

**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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof

## Highlights:

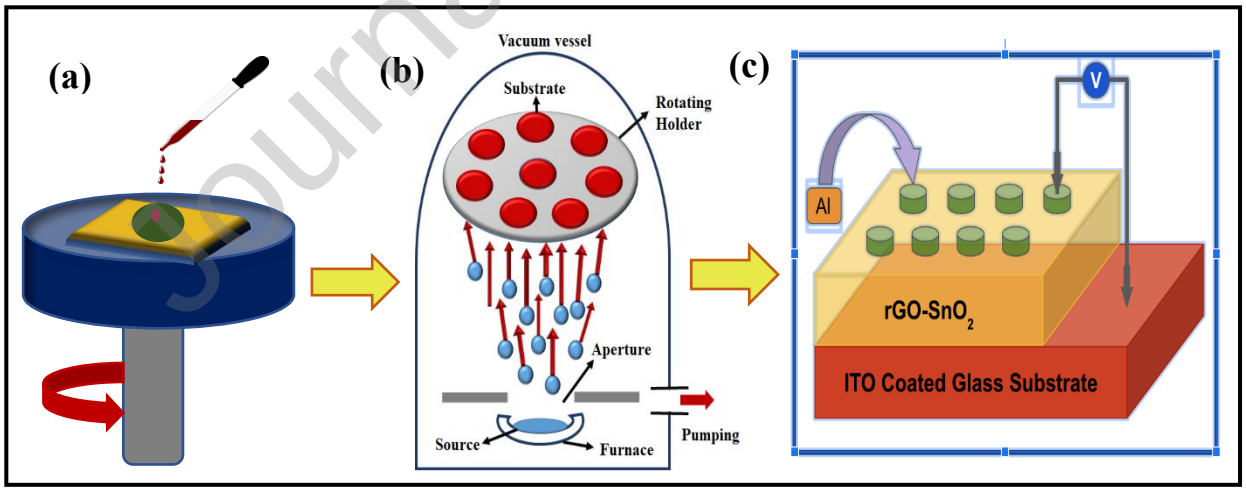
- RGO-$SnO_2$ nanocomposite was synthesized by hydrothermal technique.
- Structural and compositional analysis of $SnO_2$ and RGO-$SnO_2$ nanocomposite powder.
- Improved resistive switching of $SnO_2$-RGO composite in comparison to pure $SnO_2$.

Journal Pre-proof

**Figures:**

**Fig.1**

**Fig.2**




Journal Pre-proof

Journal Pre-proof

**Fig.3**

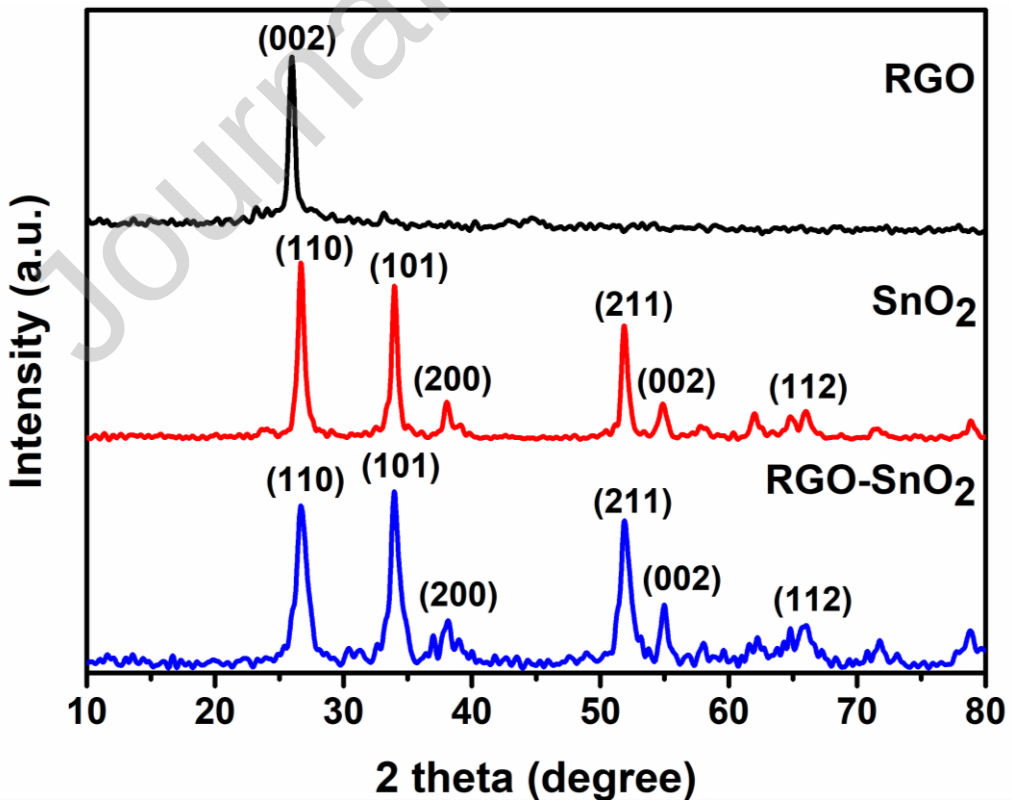
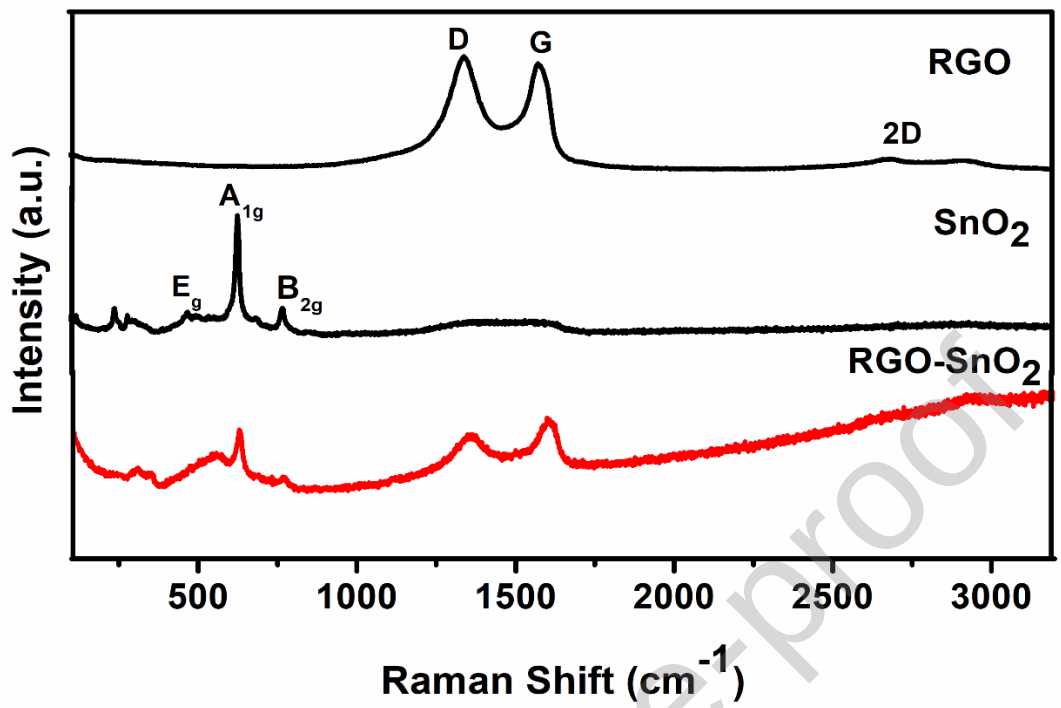
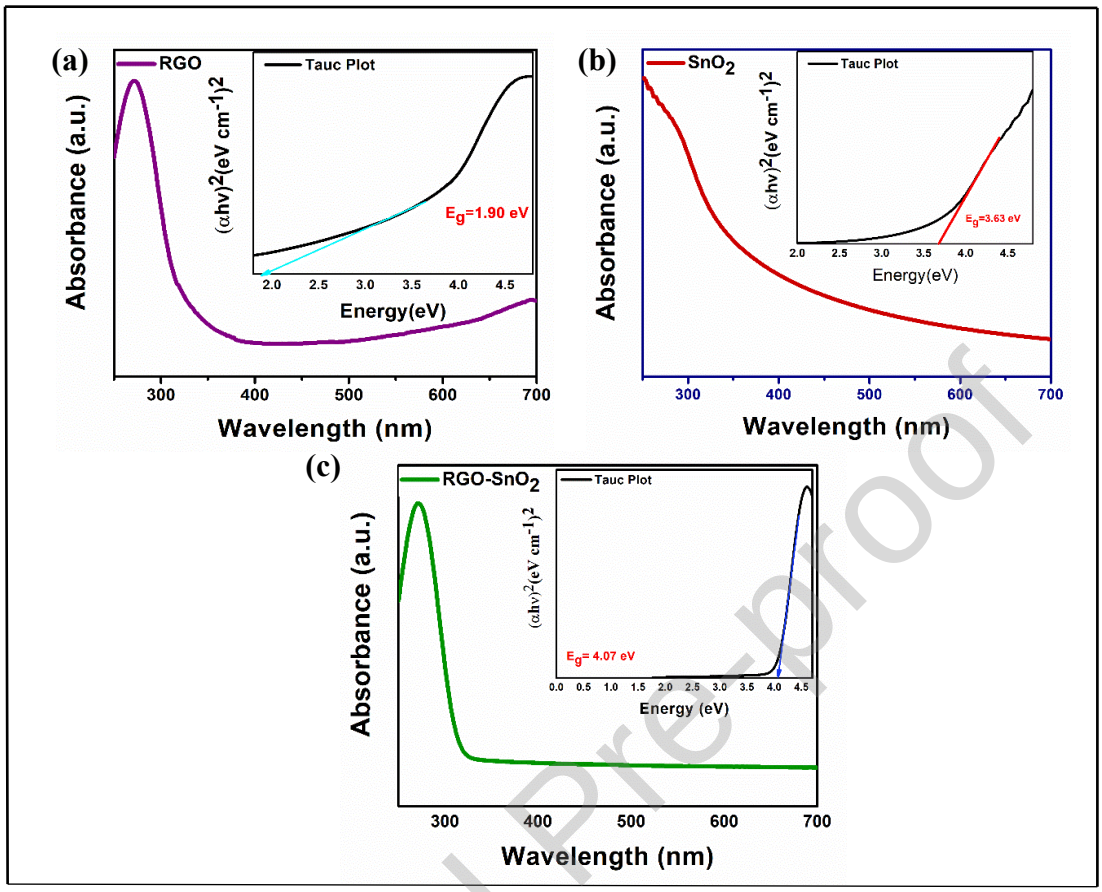


**Fig.4**




Journal Pre-proof

**Fig.5**




Journal Pre-proof

**Fig.6**

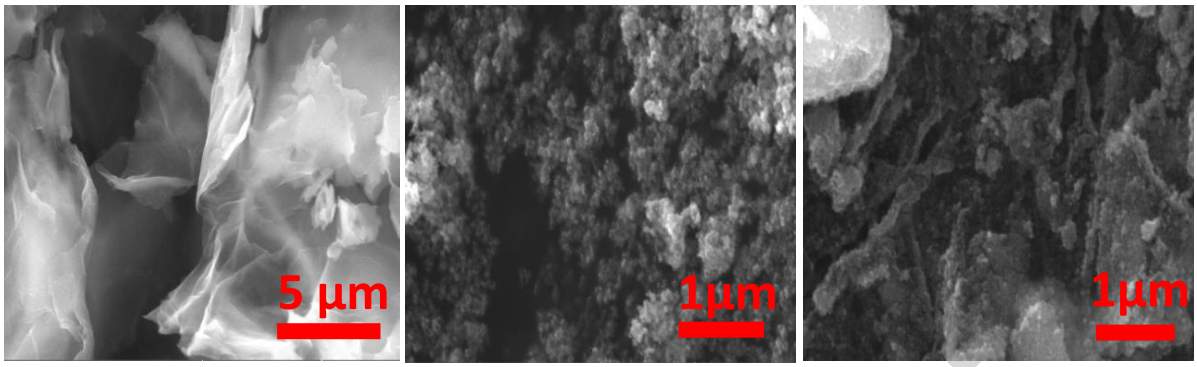
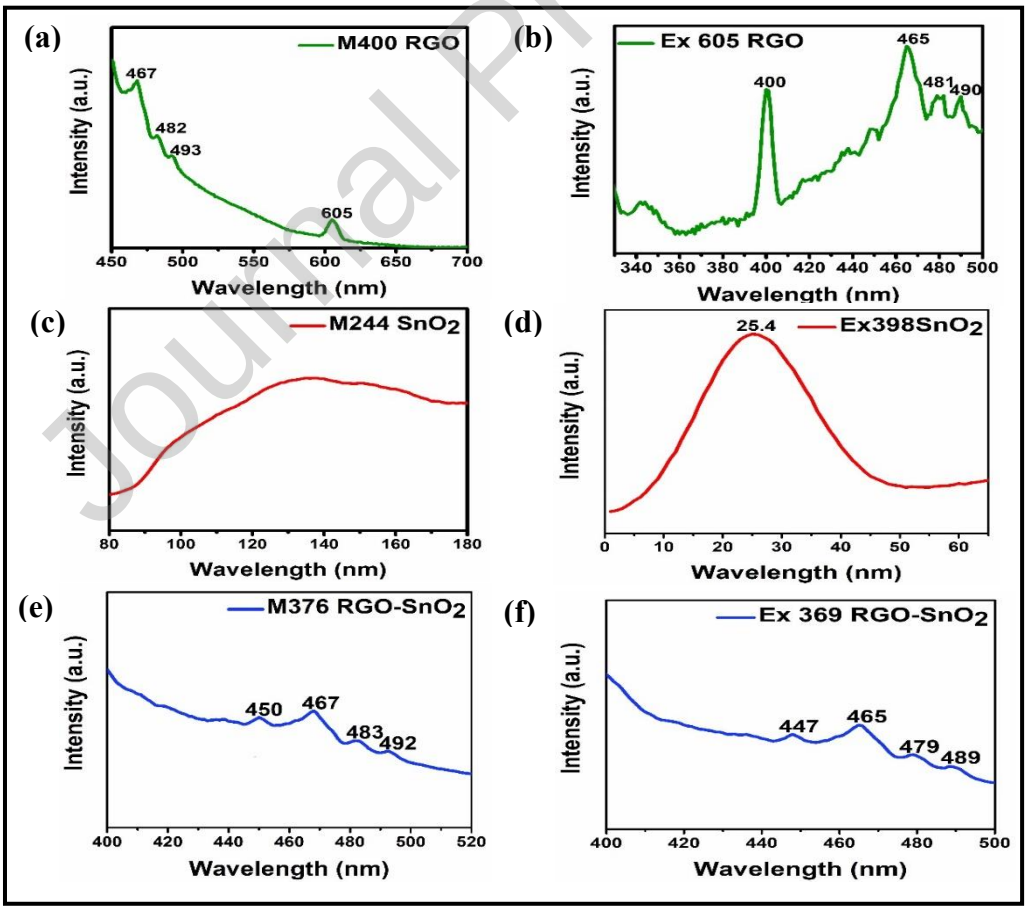



**Fig.7**

**Fig.8**




Journal Pre-proof

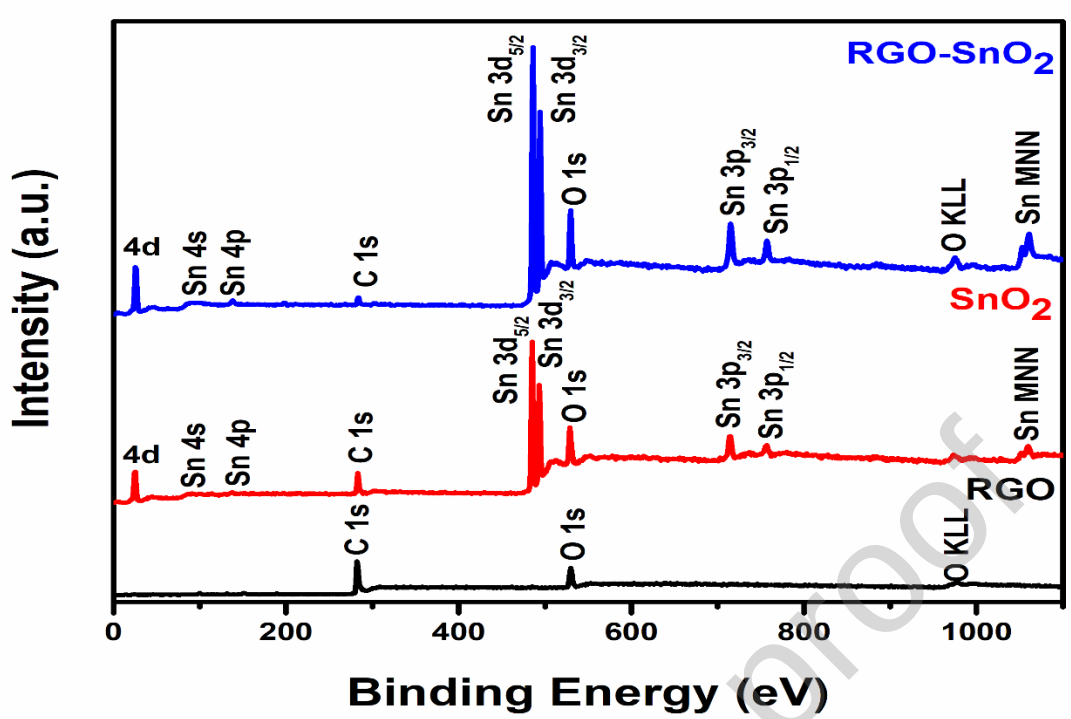

RGO-SnO2
SnO2
RGO
Sn 3d5/2
Sn 3d3/2
O 1s
Sn 3p3/2
Sn 3p1/2
O KLL
Sn MNN
4d
Sn 4s
Sn 4p
C 1s
Intensity (a.u.)
0
200
400
600
800
1000
Binding Energy (eV)

Fig.9

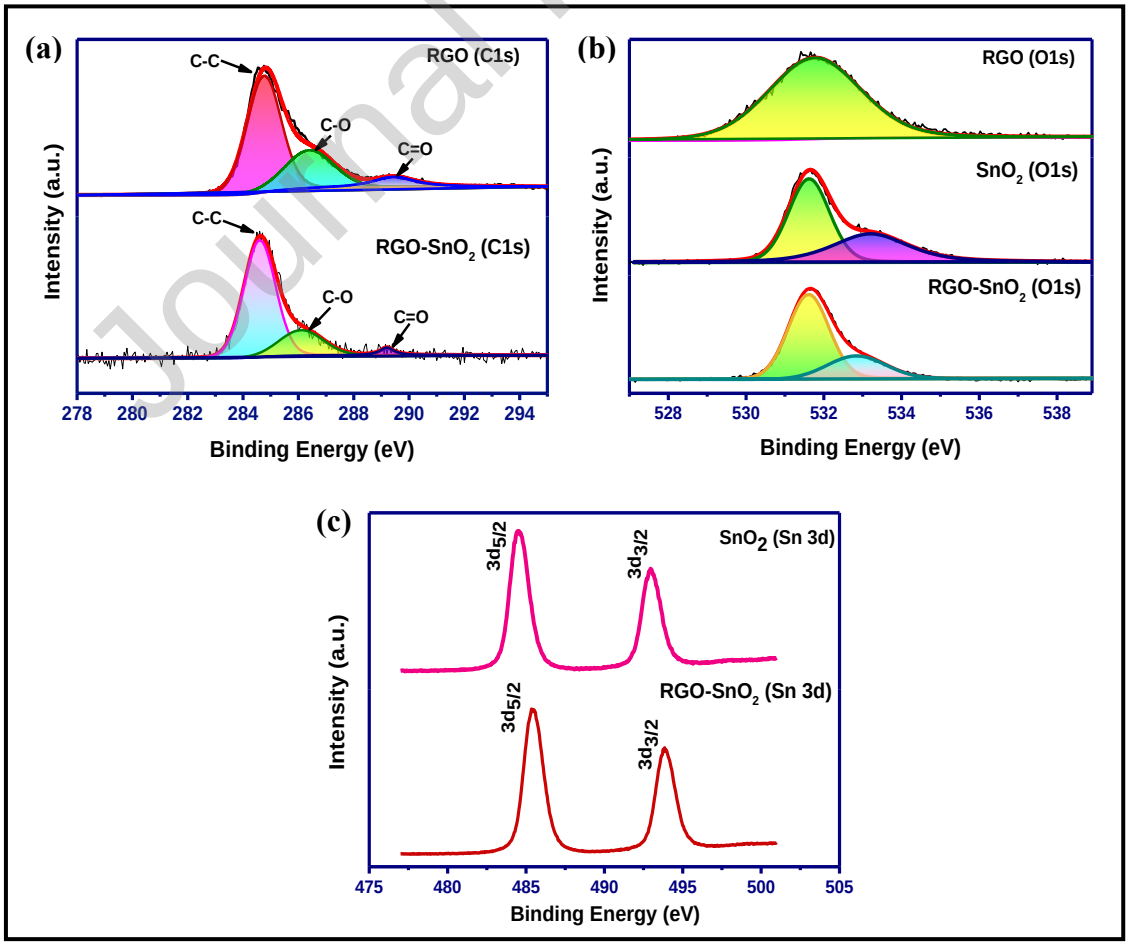
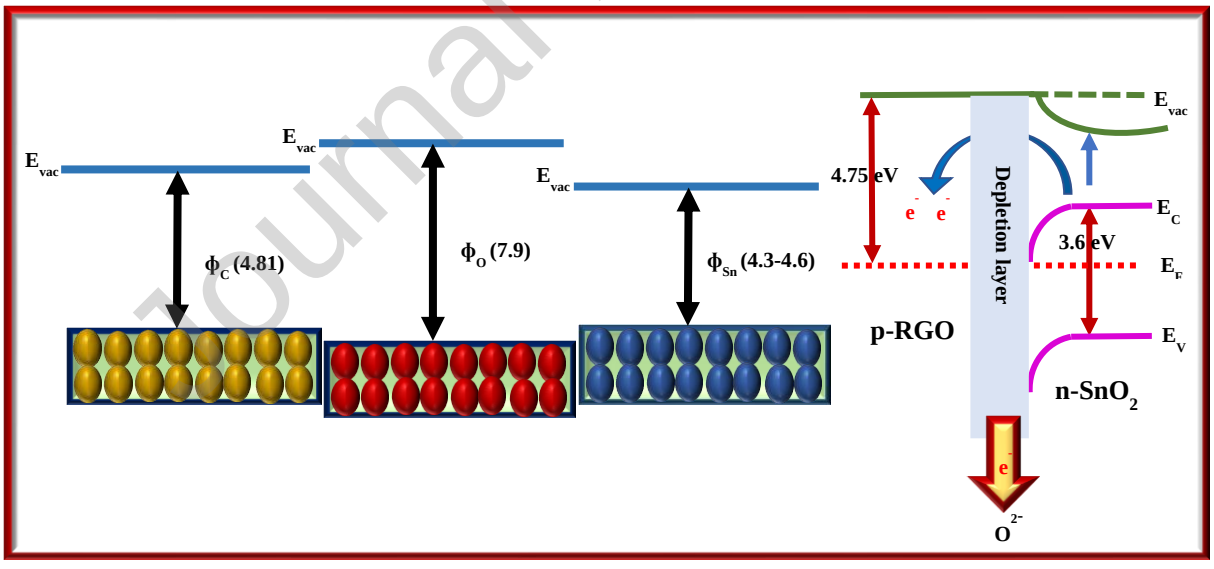
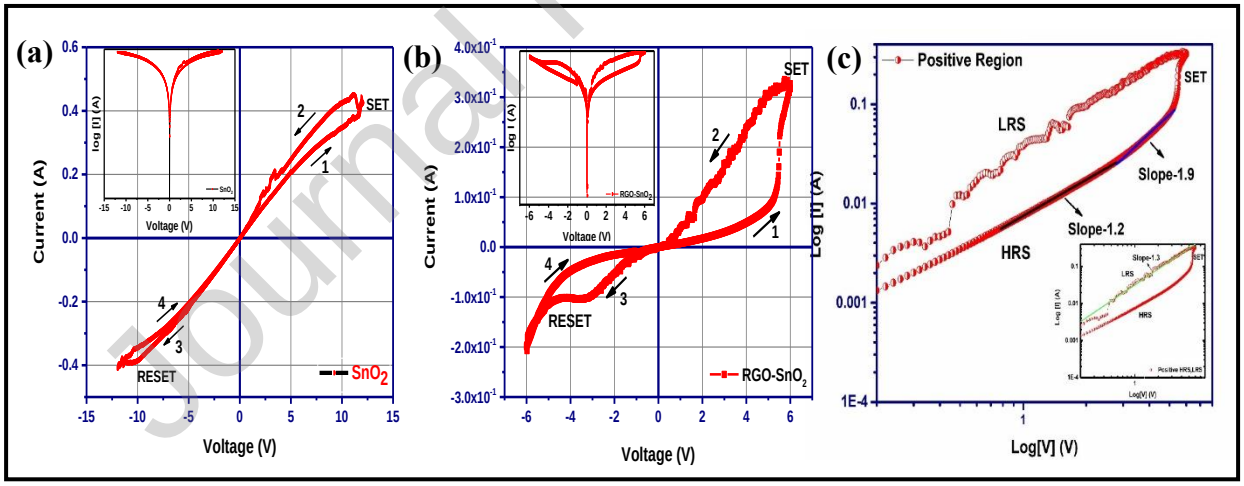

(a)
RGO (C1s)
C-C
C-O
C=O
RGO-SnO2 (C1s)
Intensity (a.u.)
278
280
282
284
286
288
290
292
294
Binding Energy (eV)
(b)
RGO (O1s)
SnO2 (O1s)
RGO-SnO2 (O1s)
528
530
532
534
536
538
(c)
SnO2 (Sn 3d)
RGO-SnO2 (Sn 3d)
3d5/2
3d3/2
475
480
485
490
495
500
505

**Fig. 10**



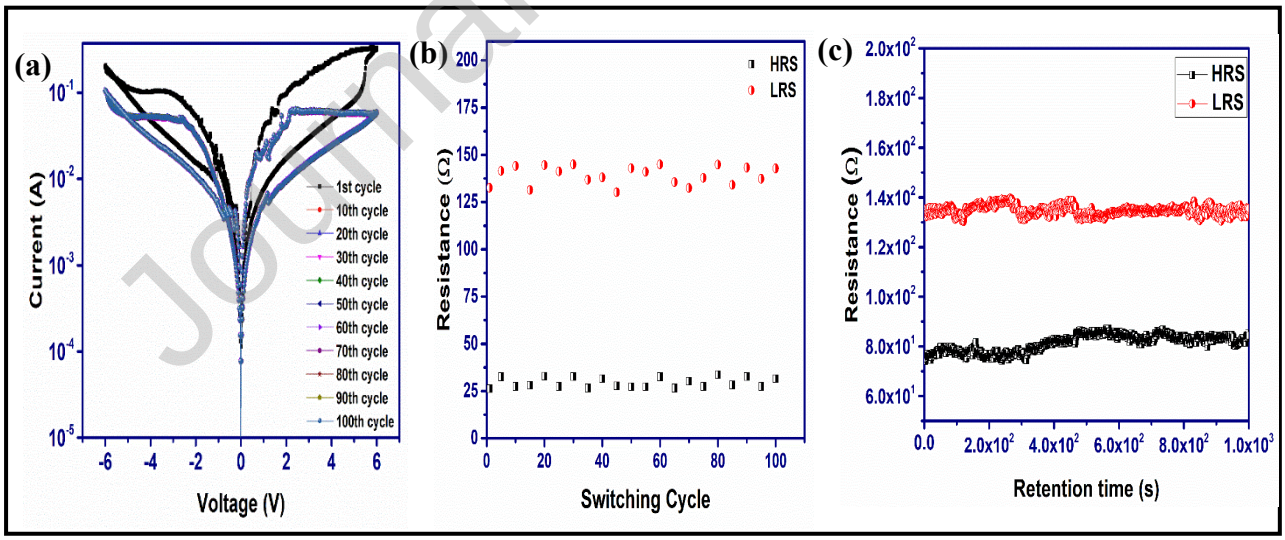

Journal Pre-proof

**Fig.11**




Journal Pre-proof

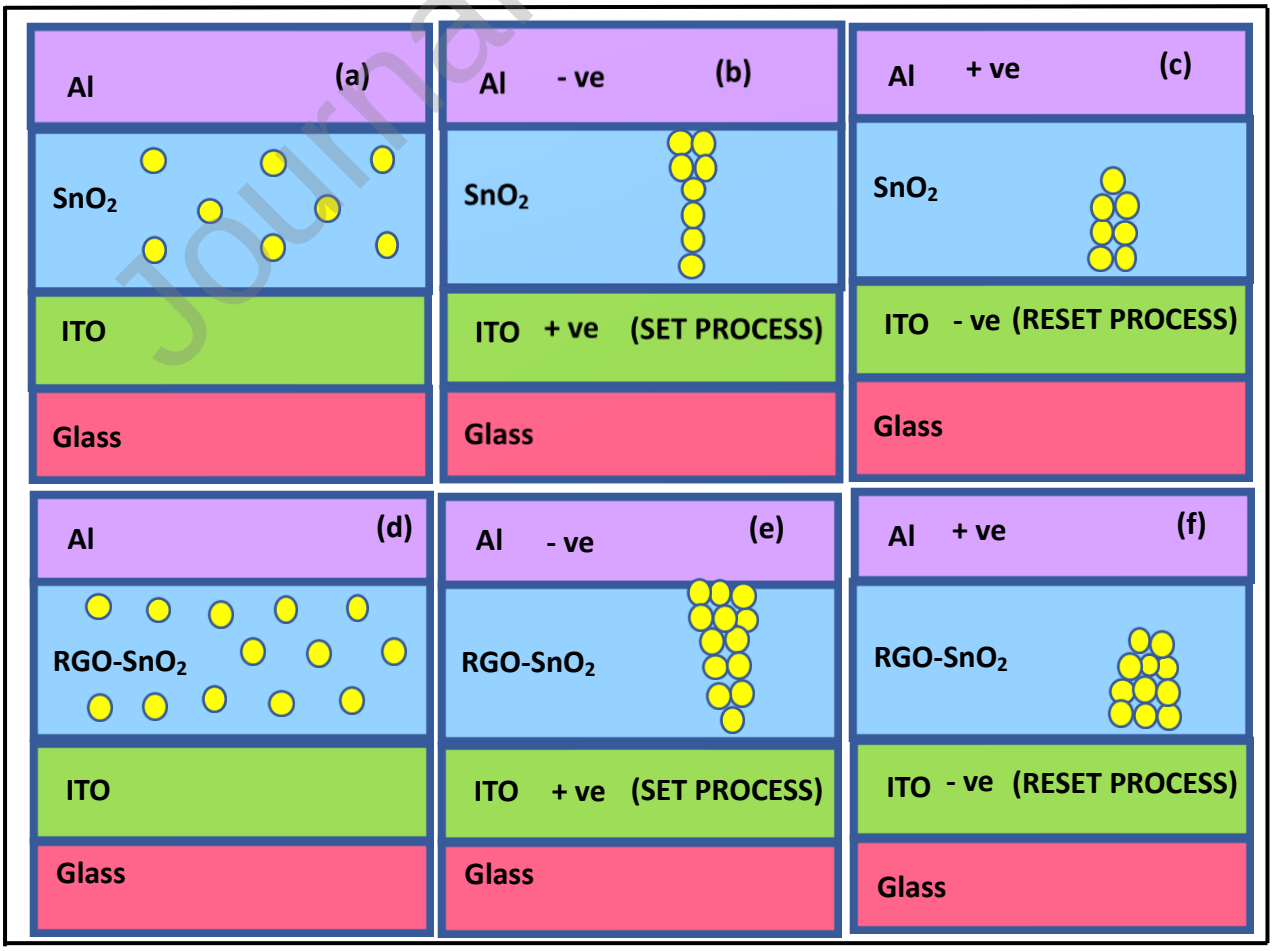
**Fig.12**




Journal Pre-proof

Journal Pre-proof

**Fig. 13**



Soft Computing
https://doi.org/10.1007/s00500-022-07269-4

APPLICATION OF SOFT COMPUTING

# Kernel picture fuzzy clustering with spatial neighborhood information for MRI image segmentation

**Dhirendra Kumar**[1] · **Inder Khatri**[1] · **Aaryan Gupta**[1] · **Rachana Gusain**[2]

Accepted: 13 May 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

Image segmentation task becomes complex in the presence of vague boundary structure and spatially distributed noise. In the literature, fuzzy set and its extension-based clustering methods are utilized to address vagueness for the image segmentation problem. In this work, picture fuzzy set theoretic clustering method has been proposed to enhance MRI image segmentation that is robust to the vague boundary structure, nonlinearity and spatially distributed noise. Most of the related work for handling noise in segmentation process is majorly based on smoothing the image, which results in the loss of fine structure. Moreover, the nonlinearity present in the data leads to inaccurate segmentation result. In this work, we have defined the optimization problem for clustering the pixel intensity values for image segmentation using the picture fuzzy set theoretic approach, which handle the vagueness. Also, we have included a spatial neighborhood information term in the optimization problem of the proposed method that avoids smoothing and preserves fine details in the segmentation process. Further, kernel distance measure is utilized to capture the nonlinear structure present in image in the proposed optimization problem. The experiments are carried out on a synthetic image dataset and two publicly available brain MRI datasets. The comparison with the state-of-the-art methods shows that the proposed picture fuzzy clustering method provides better segmentation performance in terms of average segmentation accuracy and Dice score.

**Keywords** Picture fuzzy sets · Picture fuzzy clustering · Kernel distance measures · Image segmentation · Magnetic resonance imaging · Spatial neighborhood information

## 1 Introduction

Clustering in general refers to a method of partitioning an unlabeled data set into subdivisions, commonly known as clusters, based on intrinsic properties of the data. The main idea is to explore and analyze the structural information associated with the unlabeled data. The technique has been widely applied in a variety of fields such as pattern recognition, content retrieval, image processing, plant disease classification, Hand gesture classification (Berry and Castellanos 2004; Gadekallu et al. 2021a, b; Jain et al. 1999; Rui et al. 1999). The conventional way of obtaining clusters is the hard *c*-means (HCM) clustering method, which results in c crisp partitions of the data set (Lloyd 1982). Assigning a data point to exactly one cluster ignores the uncertainty about the data point belonging to more than one cluster especially at the boundary and therefore tending to lose its interpretability for many real-world applications. Fuzzy *c*-means (FCM) (Bezdek 1981) overcomes this problem by assigning membership values for each data point to *c* number of clusters. In this way, each cluster can be represented by fuzzy sets.

We can use clustering to partition the pixel intensity values of an image into a number of similar regions referred as image segmentation. Image segmentation is a fundamental and important step in image analysis problems. Now a days, the medical imaging is popular for delineating the internal structure of human organs for finding out the abnormalities

✉ Dhirendra Kumar
dhirendrakumar@dtu.ac.in

Inder Khatri
inderkhatri999@gmail.com

Aaryan Gupta
aryan227227@gmail.com

Rachana Gusain
rachna.jnu2010@gmail.com

[1] Department of Applied Mathematics, Delhi Technological University, Delhi, India

[2] Department of Computer Science, Doon University, Dehra Dun, Uttarakhand, India

Published online: 09 July 2022

for early detection of diseases. Huge amount of medical images from various imaging technologies such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), mammogram, X-rays, and ultrasound, are produced for disease diagnosis. Human expertise is crucial in detecting and diagnosis of a disease by identifying some abnormalities in these medical images. This being a time-consuming process and computer-aided diagnostics can certainly help the expert radiologist in analyzing more images in less time.

MRI imaging is a popular modality for better understanding and analysis of a diseased organ and provides better resolution without bad effect on the tissue. MRI imaging is safe unlike X-rays and PET exposure, which harm the healthy tissues while scanning the body organs. During the image acquisition process using MRI machines, the main artifacts that affect the quality of the acquired MRI image are noise, intensity inhomogeneity and partial volume effect. Due to these artifacts, the segmentation performance for the MRI image may deteriorate. In the literature, many approaches have been proposed for MRI image segmentation such as level set (Alipour and Shanbehzadeh 2014; Li et al. 2011; Wang et al. 2010), graph cut (Chen et al. 2016), region growing (Sato et al. 2000), Atlas-guided methods (Rohlfing et al. 2004), artificial neural network (ANN) (Reddick et al. 1997), Markov random field model (Held et al. 1997) and clustering (Balafar et al. 2010; Benaichouche et al. 2013; Cocosco et al. 1997; Huang et al. 2015; Ji et al. 2014; Kannan et al. 2013; Krinidis and Chatzis 2010; Olabarriaga and Smeulders 2001; Pham et al. 2000; Qiu et al. 2013; Wang et al. 2013).

Among these segmentation techniques, clustering based on fuzzy set theory (Zadeh 1965) known as fuzzy $c$-means (FCM) (Bezdek James 1981) and its variants are found to be better in comparison with conventional hard clustering algorithm. The fuzzy clustering method is preferred as it models the related characteristics of MRI brain images, i.e., it allows each feature vector to belong to multiple clusters with membership degree whose value lies in the interval [0 1]. The performance of FCM method for image segmentation is not good in the presence of noise. To handle the noise, many variants of conventional FCM methods are proposed in the literature (Ahmed et al. 2002; Cai et al. 2007; Chen and Zhang 2004; Szilagyi et al. 2003).

Most of the methods discussed till looses fine image details and are dependent on selection of optimal parameter values. To address this issue, Krindis et al. (2010) proposed a fuzzy local information clustering method (FLICM) to tackle the problem of noise in image segmentation. Similar to FCM_S (Ahmed et al. 2002), it uses the neighboring pixels deviation from centroid's intensity, weighted by a fuzzy factor and spatial distance of neighbors. The method is parameter-free, but at the same time the FLICM requires to calculate the local information term in each iteration, which makes it time-consuming segmentation method. The performance of FLICM method for noisy image segmentation is satisfactory, but it has been reported in literature that the optimization problem for clustering is not minimizing the objective value rather converging the fuzzy partition matrix. Guo et al. proposed an adaptive fuzzy $C$-means based on local noise detecting (NDFCM), where the spatial parameters for each pixel were dependent on the noise level in a given immediate neighborhood (Guo et al. 2016). Despite being an noise adaptive algorithm, NDFCM has a high computational complexity due to its dependence on the three input parameters, which are required to be fine-tuned to get a decent performance.

Recently, Lei et al. proposed a fast and robust fuzzy $c$-means algorithm (FRFCM), which gave impressive results with considerably low time complexity (Lei et al. 2018). The method utilized morphological reconstruction operation as a pre-processing step, which made it robust to a variety of noises. To avoid the heavy computation in measuring the distance between the neighbor pixels and centroids to handle noisy pixels, the method used membership filtering as a post-processing step, which was equivalent to introducing neighbor information. Although the method was robust to a variety of noises, it was not able to perform well on high noise samples. When encountered with high level of noise in image, FRFCM is not able to preserve the sharp edges and shapes. Another research work (Zhang et al. 2018), termed as DSFCMN, modeled the deviation between the original pixel values and measured noisy pixels value as residual and used this value in optimization function. They considered the residual term to be sparse matrix and use the l1 norm measure in objective function as a constraint over residuals. However, DSFCMN did not show reliable performance when tested with higher noise samples. Further, Wang et al. proposed weighted residual fuzzy $C$-means (WRFCM), which used weighted l2-norm fidelity for making the residual estimation more reliable and showed better results compared to the previous works (Wang et al. 2020).

Further, to deal with nonlinear structure present in the image many works have been reported in the literature that utilize the kernel distance measure. The research work (Zhao et al. 2013) proposed a kernel generalized fuzzy $c$-means clustering with spatial information termed as KGFCM for image segmentation. Most of these methods are dependent on optimal selection of input parameters values for better segmentation result. For this, the optimal values of these parameters are usually obtained via a grid search method, which is a time-consuming process. Motivated by the FLICM method, Gong et al. (2012) proposed a variant of FLICM method by replacing the Euclidean distance with kernel metric and further introduced a trade-off weighted fuzzy factor to better use the neighbor information in an adaptive manner termed as KWFLICM. Although the performance of

KWFLICM method is better in comparison with the FLICM method, it inherits the problem of FLICM method.

The performance of the most of the method is depended on the selection of parameter involving for penalizing the noise in image segmentation problem. Selecting the correct parameter is a challenging task. Also, due to impreciseness in real data, the FCM-based clustering could not properly define the fuzzy partition matrix, i.e., the membership degree of the belongingness to a given cluster is imprecise. This imprecision arises an uncertainty in defining the membership value while fuzzy clustering process. To deal with this problem of uncertainty in defining of membership value, Atanassov proposed intuitionistic fuzzy set theory (Atanassov 1986), which can better represent the real data in comparison with fuzzy set. Incorporation of intuitionistic fuzzy set theory for clustering process provides better performance, which is more accurate, robust to noise and faster convergence in comparison with the conventional FCM (Iakovidis et al. 2008). The IFS theory exploits the non-membership degree along with membership degree value. In the literature, many research works have been suggested IFS theory based clustering as it exploits the IFS theory for dealing with the uncertainty in defining the membership values while clustering (Chaira 2011; Kumar et al. 2019a, b, c; Verma et al. 2016; Xu and Wu 2010; Zang et al. 2017). Recently, research work (Verma et al. 2016) proposed a parameter-free method to handle noise and outliers in the medical data termed as improved IFCM (IIFCM). IIFCM used both local spatial and grey-level information together for MRI segmentation. The method was found to be efficient on noisy data, but it is a considerably time-consuming method. Further research work (Kumar et al. 2019b) introduced a membership-dependent flexible neighborhood constraint to the IFCM method and proposed IFCM with spatial neighborhood information (IFCMSNI). The spatial neighborhood information term preserved the fine image details and is capable of dealing with high-intensity various noises. Their model gives better results on noisy MRI images compared to the counter ones.

It is observed that IFS representation of real-world information could not perform better when the data are complex. The IFS set theoretic approach for clustering is not efficient for the image containing uneven illumination and complex structure for image segmentation problem (Wu and Chen 2020). The research work (Cuong and Kreinovich 2013) suggested picture fuzzy set, an extension of IFS set, which increases the degree of representation of real-world data. Motivated by the fact that PFS has more generalized version of IFS theory, the research work (Son 2015) suggested a picture fuzzy c- means (PFCM) clustering method based on PFS theory. The research work (Thong 2016) suggested an another variant of fuzzy clustering of PFS (FC-PFS). In FC-PFS method, the degree of neutrality and refusal degree are treated as same, which is not reasonable. Motivated from this fact, the research work (Wu and Wu 2017) suggested a robust image segmentation algorithm to handle noise using spatial context to modified picture fuzzy clustering method on picture fuzzy sets (IPFCM). The research work (Wu and Chen 2020) modified the IPFCM method termed as AIPFCM method by introducing the adaptive weights in the optimization problem of the IPFCM method to enhance the segmentation performance. The AIPFCM method adaptive to the for improving the performance. query Please check the clarity of the sentence 'The AIPFCM method adaptive to the for improving the performance'.

Further, the research work (Wu and Chen 2020) proposed a robust adaptive entropy-weighted picture fuzzy clustering with spatial information (APFCMS), which handle the noise present in image segmentation process. In most of the method, the spatial neighborhood information term added could not efficiently handle the problem of noise due to involvement of image smoothing directly or indirectly in segmentation problem. The smoothing operation during segmentation process overlooked the fine image structures and edges in image. Also, non-Euclidean structures present in the image could not be easily handled by the existing methods.

Most of the related work for handling noise in segmentation process is majorly based on smoothing the image, which results in the loss of fine structure in image. Moreover, the nonlinearity present in the data leads to inaccurate segmentation result. The research problems discussed above motivate us to present an effort to solve the problems in related works with the help of proposed kernel picture fuzzy $c$-means with spatial neighborhood information (KPFCMSNI). The proposed KPFCMSNI method is a variant of PFCM clustering method, which incorporate the Gaussian kernel distance measure. It helps to cluster the data points in a robust way as the calculation of the distance is done in higher-dimensional space using kernel trick. Due to transformation of data points in higher-dimensional feature space also known as kernel space, the representational capability is increased. In this way, the inherent non-Euclidean structure in data is handled without increasing the computational complexity. Also, we have included a spatial neighborhood information term in the optimization problem of the proposed method. This term uses the neighborhood membership values to dampen the effect of noise that avoid smoothing and preserves fine details in the image segmentation process. Further, to check the efficacy of proposed KPFCMSNI method, experiments are performed on various real and synthetic datasets. The performance of the proposed method is compared quantitatively in terms of average segmentation accuracy (ASA) and Dice score (DS) with well-known existing methods such as FCM, FC_PFS, RPFCM, FCM_S, FLICM, IPFCM, AIPFCM, APFCMS, IFCMSNI, KWFLICM, NDFCM, KGFCM, IIFCM, FRFCM, DSFCMN and WRFCM. Further, statistical test is performed

to show the efficacy of the proposed KPFCMSNI method in comparison with state-of-the-art methods for image segmentation process.

The rest part of this paper is organized as follows: Section 2 describes the preliminaries. Section 3 describes clustering methods in picture fuzzy set framework. The proposed method is described in Sect. 4. The experimental results on various synthetic and real data are presented in Sect. 5. Finally, Sect. 6 includes the conclusion.

## 2 Preliminaries and notations

The description of notations used in this work and related definition is described here in this section.

**Definition 1** Fuzzy set: A fuzzy set is a set in which each member element will have the fractional membership via a membership function $\mu_A : X \rightarrow [0, 1]$, which gives the degree of belongingness (Zadeh 1965). If $A$ is a fuzzy set defined over a set $X$, it can be represented as:

$$A = \{(x, \mu_A(x)) : x \in X\} \tag{1}$$

**Definition 2** Picture fuzzy set (PFS): A picture fuzzy set $B$ is an extension of fuzzy set over $X$ which is represented as (Cuong and Kreinovich 2013):

$$B = \{(x, \mu_B(x), \eta_B(x), \gamma_B(x)) : x \in X \text{ and } 0 \leq \mu_B(x) + \eta_B(x) + \gamma_B(x) \leq 1\} \tag{2}$$

where $\mu_B : X \rightarrow [0, 1]$, $\eta_B : X \rightarrow [0, 1]$, $\gamma_B : X \rightarrow [0, 1]$ are positive membership value, neutral and negative membership functions of an element x in the set $B$.

The refusal degree $\xi_B(x)$ of an element x in the set $B$ is represented as $\xi_B(x) = 1 - (\mu_B(x) + \eta_B(x) + \gamma_B(x))$. In case $\xi_B(x) = 0$, the PFS $B$ reduces to IFS $B$ and when $\eta_B(x) = 0; \xi_B(x) = 0$, then the PFS $B$ is reduced to FS $B$ for all $x$ in $B$.

### 2.1 Kernel methods

Many kernel-based machine learning techniques are proved to be efficient for several classifier and feature extraction techniques such as support vector machines (SVMs) (Cristianini and Shawe-Taylor 2000; Vapnik 2013), kernel principal component analysis (KPCA) (Schölkopf et al. 1998) and kernel Fisher discriminant (KFD) (Roth and Steinhage 2000). In general, feature vectors associated with real data are not linearly separable and to learn nonlinear boundary for classification is very difficult. According to Cover's theorem (Cover 1965), the transformation of data points from original low-dimensional space to high-dimensional space can be beneficial as such data points are linear separable in higher-dimensional space. But, at the same time the computational complexity for learning the classifier for high-dimensional features space is high as it involves a nonlinear mapping $\phi$ from the original space to higher-dimensional space, such that a function $K(x_i, x_j)$ that satisfies Mercer's condition (J Mercer 1909) can be given as:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \tag{3}$$

where $x_i$ represent data in original space, $\phi(x_i)$ denotes the data in higher-dimensional space. From (3), it is clear that kernel function facilitates the ease of calculating require metric without involvement of data in higher-dimension space and save the time. Use of kernels has a wide application in many areas as they make it possible to map data onto a high-dimensional feature space so that the representation capability of linear machines can be enhanced. The linear kernel function, polynomial kernel function of degree $p$, sigmoid kernel function and the Gaussian radial basis function (RBF) kernel function are few important kernel functions, which are given as:

$$\begin{aligned} K(x_i, x_j) &= x_j^T x_i \\ K(x_i, x_j) &= (1 + x_j^T x_i)^p \\ K(x_i, x_j) &= \tanh(\alpha(x_j^T x_i) + \beta) \\ K(x_i, x_j) &= \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right) \end{aligned}$$

respectively. Let elements $\mathbf{x}_i$ and $\mathbf{x}_j$ mapped to higher-dimensional space through a nonlinear transform $\phi$, then Euclidean distance in kernel space can be derived as follows:

$$\|\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)\|^2 = \|\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)\|^T \|\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)\|$$

$$\|\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)\|^2 = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_i) + \phi(\mathbf{x}_j)^T \phi(\mathbf{x}_j) - 2\phi(\mathbf{x}_i)\phi(\mathbf{x}_j),$$

and according to (3), the above equation can be written as:

$$\|\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)\|^2 = K(\mathbf{x}_i, \mathbf{x}_i) + K(\mathbf{x}_j, \mathbf{x}_j) - 2K(\mathbf{x}_i, \mathbf{x}_j) \tag{4}$$

In case of Gaussian RBF kernel, this can be written as:

$$\|\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)\|^2 = 2(1 - K(\mathbf{x}_i, \mathbf{x}_j)) \tag{5}$$

Similarly, the Euclidean distance for other kernel space can be calculated after replacing the corresponding kernel function in (4).

## 3 Related works

It is observed that IFS representation of real-world information could not perform better when the data are complex. The IFS set theoretic approach for clustering is not efficient for the image containing uneven illumination and complex structure for image segmentation problem (Wu and Chen 2020). The research work (Cuong and Kreinovich 2013) suggested picture fuzzy set, an extension of IFS set, which increases the degree of representation of real-world data. In the following subsections, clustering methods based on picture fuzzy set theory are discussed in detail with corresponding optimization problem.

### 3.1 Picture fuzzy clustering

The research work (Son 2015) proposed the picture fuzzy clustering, which modified the conventional fuzzy c-mean to incorporate the advantage of the PFS theory. This increases the interpretability of the clustering result and increases the representational capabilities. Let $X$ represent the set of $N$ data points to be clustered in $c$ groups; then, the optimization problem of picture fuzzy clustering can be given as (Son 2015):

$$\min J_\lambda(\mathbf{U}, \mathbf{V} : \mathbf{X}) = \sum_{i=1}^{c}\sum_{j=1}^{N}\left(\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}\right)^m \|\mathbf{x}_j - \mathbf{v}_i\|^2 \tag{6}$$

with the following constraints

$$\begin{aligned} 0 \le \mu_{ij}, \eta_{ij}, \xi_{ij} \le 1 \\ 0 \le \mu_{ij} + \eta_{ij} + \xi_{ij} \le 1 \; 1 \le i \le c \; 1 \le j \le N \\ \sum_{i=1}^{c} \frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}} = 1, \; 1 \le j \le N \\ \sum_{i=1}^{c} (\eta_{ij} + \frac{\xi_{ij}}{c}) = 1, \; 1 \le j \le N \end{aligned}$$

where $\mathbf{x}_j$, $\mathbf{v}_i$, and $m$ represent the $j$th datapoint, $i$th cluster centroid and degree of fuzziness, respectively. The membership degree, neutrality degree and refusal degree corresponding to $j$th data point to $i$th cluster are denoted by $\mu_{ij}, \eta_{ij}$ and $\xi_{ij}$, respectively. These notations have a similar meaning throughout the manuscript. The optimization problem (6) can be solved using the Lagrange method of undetermined multipliers, and the solution can be given as:

$$\mu_{ij} = (1-\eta_{ij}-\xi_{ij})\frac{\|\mathbf{x}_j - \mathbf{v}_i\|^{\frac{-2}{m-1}}}{\sum_{k=1}^{c}\|\mathbf{x}_j - \mathbf{v}_k\|^{\frac{-2}{m-1}}} \tag{7}$$

$$\eta_{ij} = 1 - \xi_{ij} + \frac{\frac{c-1}{c\sum_{k=1}^{c}\xi_{kj}}}{\sum_{k=1}^{c}\frac{\mu_{ij}}{\mu_{kj}}\left(\frac{\|\mathbf{x}_j-\mathbf{v}_i\|}{\|\mathbf{x}_j-\mathbf{v}_k\|}\right)^{\frac{2}{m+1}}} \tag{8}$$

$$\mathbf{v}_i = \frac{\sum_{j=1}^{N}\left(\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}\right)^m \mathbf{x}_j}{\sum_{j=1}^{N}\left(\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}\right)^m} \tag{9}$$

$$\xi_{ij} = 1 - (\mu_{ij} + \eta_{ij}) - (1 - (\mu_{ij} + \eta_{ij})^\alpha)^{\frac{1}{\alpha}} \tag{10}$$

Using the above solution, an alternating optimization algorithm can be designed and segmentation results can be obtained.

### 3.2 Fuzzy clustering on picture fuzzy sets (FC_PFS)

The research work (Thong 2016) proposed the variant of picture fuzzy clustering termed as FC_PFS. The FC_PFS method guarantees the working on the PFS sets since at least one of two uncertain factors, namely the neutral and refusal degrees, always exists in the model. Let $X$ represent the set of $N$ data points to be clustered in $c$ groups; then, the optimization problem of the FC_PFS method can be given as (Thong 2016):

$$\begin{aligned} \min J_\lambda(\mathbf{U}, \mathbf{V} : \mathbf{X}) \; &= \sum_{i=1}^{c}\sum_{j=1}^{N}(\mu_{ij}(2-\xi_{ij}))^m \|\mathbf{x}_j - \mathbf{v}_i\|^2 \\ &+ \sum_{i=1}^{c}\sum_{j=1}^{N}\eta_{ij}(\log\eta_{ij} + \xi_{ij}) \end{aligned} \tag{11}$$

with the following constraints

$$\begin{aligned} 0 \le \mu_{ij}, \eta_{ij}, \xi_{ij} \le 1 \\ 0 \le \mu_{ij} + \eta_{ij} + \xi_{ij} \le 1 \; 1 \le i \le c \; 1 \le j \le N \\ \sum_{i=1}^{c} \mu_{ij}(2-\xi_{ij}) = 1, \; 1 \le j \le N \\ \sum_{i=1}^{c} (\eta_{ij} + \frac{\xi_{ij}}{c}) = 1, \; 1 \le j \le N \end{aligned}$$

The optimization problem (11) can be solved using Lagrange method of undetermined multipliers, and the solution can be given as:

$$\mu_{ij} = \frac{1}{(2-\xi_{ij})}\frac{\|\mathbf{x}_j - \mathbf{v}_i\|^{\frac{-2}{m-1}}}{\sum_{k=1}^{c}\|\mathbf{x}_j - \mathbf{v}_k\|^{\frac{-2}{m-1}}} \tag{12}$$

$$\eta_{ij} = \frac{\exp(-\xi_{ij})}{\sum_{k=1}^{c}\exp(-\xi_{kj})}\left(1 - \frac{\sum_{k=1}^{c}\xi_{kj}}{c}\right) \tag{13}$$

$$\mathbf{v}_i = \frac{\sum_{j=1}^{N}(\mu_{ij}(2-\xi_{ij}))^m \mathbf{x}_j}{\sum_{j=1}^{N}(\mu_{ij}(2-\xi_{ij}))^m} \tag{14}$$

$$\xi_{ij} = 1 - (\mu_{ij} + \eta_{ij}) - (1 - (\mu_{ij} + \eta_{ij})^{\alpha})^{\frac{1}{\alpha}} \tag{15}$$

Using the above solution, an alternating optimization algorithm can be designed and segmentation results can be obtained.

### 3.3 Robust picture fuzzy *c*-means clustering (RPFCM)

The FC_PFS method is time-consuming so in the research work (Wu and Wu 2017) an improved version with symmetric regularizing term is proposed termed as robust picture fuzzy c means (RPFCM) clustering. Let $X$ represent the set of $N$ data points to be clustered in $c$ groups; then, the optimization problem of picture fuzzy clustering can be given as (Wu and Wu 2017):

$$\begin{aligned}\min J_{\lambda}(\mathbf{U}, \mathbf{V} : \mathbf{X}) &= \sum_{i=1}^{c}\sum_{j=1}^{N}\left(\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}\right)^{m}\|\mathbf{x}_j - \mathbf{v}_i\|^2 \\ &\quad + \sum_{i=1}^{c}\sum_{j=1}^{N}(\xi_{ij}\eta_{ij}^2 + \xi_{ij}^2\eta_{ij})\end{aligned} \tag{16}$$

with the following constraints

$$\begin{aligned}&0 \le \mu_{ij}, \eta_{ij}, \xi_{ij} \le 1 \\ &0 \le \mu_{ij} + \eta_{ij} + \xi_{ij} \le 1 \; 1 \le i \le c \; 1 \le j \le N \\ &\textit{subject to} \sum_{i=1}^{c}\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}} = 1, \; 1 \le j \le N \\ &\sum_{i=1}^{c}(\eta_{ij} + \frac{\xi_{ij}}{c}) = 1, \; 1 \le j \le N\end{aligned}$$

The optimization problem (16) can be solved using Lagrange method of undetermined multipliers, and the solution can be given as:

$$\mu_{ij} = (1-\eta_{ij}-\xi_{ij})\frac{\|\mathbf{x}_j - \mathbf{v}_i\|^{\frac{-2}{m-1}}}{\sum_{k=1}^{c}\|\mathbf{x}_j - \mathbf{v}_k\|^{\frac{-2}{m-1}}} \tag{17}$$

$$\eta_{ij} = \frac{\xi_{ij}(2\eta_{ij} + \xi_{ij})}{c(\eta_{ij} + 2\xi_{ij})} \tag{18}$$

$$\mathbf{v}_i = \frac{\sum_{j=1}^{N}\left(\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}\right)^{m}\mathbf{x}_j}{\sum_{j=1}^{N}\left(\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}\right)^{m}} \tag{19}$$

$$\xi_{ij} = 1 - (\mu_{ij} + \eta_{ij}) - (1 - (\mu_{ij} + \eta_{ij})^{\alpha})^{\frac{1}{\alpha}} \tag{20}$$

Using the above solution, an alternating optimization algorithm can be designed and segmentation results can be obtained.

### 3.4 Improved picture fuzzy *c*-means clustering (IPFCM)

The RPFCM method does not perform well in the presence of noise for image segmentation process, so in the research work (Wu and Wu 2017) an improved version with spatial regularizing term is proposed termed as improved picture fuzzy c means (IPFCM) clustering. Let $X$ represent the set of $N$ data points to be clustered in $c$ groups and $\bar{X}$ represents the mean of the data point around the neighborhood; then, the optimization problem of picture fuzzy clustering can be given as (Wu and Wu 2017):

$$\begin{aligned}&\min J_{\lambda}(\mathbf{U}, \mathbf{V} : \mathbf{X}) \\ &= \sum_{i=1}^{c}\sum_{j=1}^{N}\left(\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}\right)^{m}\big(\|\mathbf{x}_j \\ &\quad -\mathbf{v}_i\|^2 + \beta\|\bar{\mathbf{x}}_j - \mathbf{v}_i\|^2\big) + \sum_{i=1}^{c}\sum_{j=1}^{N}\left(\xi_{ij}\eta_{ij}^2 + \xi_{ij}^2\eta_{ij}\right)\end{aligned} \tag{21}$$

with the following constraints

$$\begin{aligned}&0 \le \mu_{ij}, \eta_{ij}, \xi_{ij} \le 1 \\ &0 \le \mu_{ij} + \eta_{ij} + \xi_{ij} \le 1 \; 1 \le i \le c \; 1 \le j \le N \\ &\textit{subject to} \sum_{i=1}^{c}\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}} = 1, \; 1 \le j \le N \\ &\sum_{i=1}^{c}(\eta_{ij} + \frac{\xi_{ij}}{c}) = 1, \; 1 \le j \le N\end{aligned}$$

where $\beta$ is spatial regularization parameter. The optimization problem (21) can be solved using the Lagrange method of undetermined multipliers, and the solution can be given as:

$$\begin{aligned}&\mu_{ij} \\ &= (1-\eta_{ij}-\xi_{ij})\frac{(\|\mathbf{x}_j - \mathbf{v}_i\|^2 + \gamma\|\bar{\mathbf{x}}_j - \mathbf{v}_i\|^2)^{\frac{-1}{m-1}}}{\sum_{k=1}^{c}(\|\mathbf{x}_j - \mathbf{v}_k\|^2 + \gamma\|\bar{\mathbf{x}}_j - \mathbf{v}_k\|^2)^{\frac{-1}{m-1}}}\end{aligned} \tag{22}$$

$$\eta_{ij} = \frac{\xi_{ij}(2\eta_{ij} + \xi_{ij})}{c(\eta_{ij} + 2\xi_{ij})} \tag{23}$$

$$\mathbf{v}_i = \frac{\sum_{j=1}^{N}\left(\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}\right)^{m}(\mathbf{x}_j + \gamma\bar{\mathbf{x}}_j)}{\sum_{j=1}^{N}\left(\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}\right)^{m}(1+\gamma)} \tag{24}$$

$$\xi_{ij} = 1 - (\mu_{ij} + \eta_{ij}) - (1 - (\mu_{ij} + \eta_{ij})^{\alpha})^{\frac{1}{\alpha}} \tag{25}$$

Using the above solution, an alternating optimization algorithm can be designed, and segmentation results can be obtained.

## 4 Proposed kernel picture fuzzy c means with spatial neighborhood information (KPFCMSNI) method

The proposed kernel picture fuzzy c means with spatial neighborhood information clustering method is the extended version of picture fuzzy c means for segmentation of images corrupted with noise. The optimization problem (26) of the proposed KPFCMSNI method utilizes the kernel distance measure along with a robust spatial neighborhood information term for handling the nonlinear structure and noise present in image. The optimization problem (26) consists of mainly three terms along with the underlined constraints on the variable involved for producing the valid partition matrix. The optimization problem of the proposed method will help to cluster the data points in a robust way as the distance is measured in higher-dimensional space using kernel trick. Due to transformation of data points in higher-dimensional feature space, the representational capability is increased and hence revealed the inherent non-Euclidean structures present in data without increasing the computational complexity using kernel trick. Let $X$ represent the set of $N$ data points to be clustered in $c$ groups; the optimization problem of the proposed KPFCMSNI method can be given as:

$$\begin{aligned}\min J_{\lambda}^{\phi}(\mathbf{U},\mathbf{V}:\mathbf{X}) \;&= \sum_{i=1}^{c}\sum_{j=1}^{N}\left(\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}\right)^{m}\|\phi(\mathbf{x}_j)\\ &\quad -\phi(\mathbf{v}_i)\|^2\\ &\quad +\beta\sum_{i=1}^{c}\sum_{j=1}^{N}\left(\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}\right)^{m} SNI_{ij}\\ &\quad +\sum_{i=1}^{c}\sum_{j=1}^{N}(\xi_{ij}\eta_{ij}^2+\xi_{ij}^2\eta_{ij}) \qquad (26)\end{aligned}$$

Subject to the constraints

$$\begin{aligned}&0\le\mu_{ij},\eta_{ij},\xi_{ij}\le 1\\ &0\le\mu_{ij}+\eta_{ij}+\xi_{ij}\le 1\ 1\le i\le c\ 1\le j\le N\\ &\textit{subject to}\sum_{i=1}^{c}\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}=1,\ 1\le j\le N\\ &\sum_{i=1}^{c}(\eta_{ij}+\frac{\xi_{ij}}{c})=1,\ 1\le j\le N\end{aligned}$$

where

$$\mathrm{SNI}_{ij}=\sum_{k\in N_j}\frac{\mu_{ik}\|\phi(\mathbf{x}_j)-\phi(\mathbf{v}_i)\|^2}{\sum_{k\in N_j}\mu_{ik}}$$

$\phi$ is mapping used to transform samples from low-dimensional to higher-dimensional space and $\beta$ is spatial regularization parameter. For Gaussian radial basis function (RBF) kernel function, (26) using (5) can be rewritten as:

$$\begin{aligned}\min J_{\lambda}^{\phi}(\mathbf{U},\mathbf{V}:\mathbf{X}) \;&= \sum_{i=1}^{c}\sum_{j=1}^{N}\left(\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}\right)^{m}\\ &\quad (1-K(\mathbf{x}_j,\mathbf{v}_i))\\ &\quad +\beta\sum_{i=1}^{c}\sum_{j=1}^{N}\left(\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}\right)^{m}\mathrm{SNI}_{ij}\\ &\quad +\sum_{i=1}^{c}\sum_{j=1}^{N}(\xi_{ij}\eta_{ij}^2+\xi_{ij}^2\eta_{ij}) \qquad (27)\end{aligned}$$

Subject to the constraints

$$\begin{aligned}&0\le\mu_{ij},\eta_{ij},\xi_{ij}\le 1\\ &0\le\mu_{ij}+\eta_{ij}+\xi_{ij}\le 1\ 1\le i\le c\ 1\le j\le N\\ &\textit{subject to}\sum_{i=1}^{c}\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}=1,\ 1\le j\le N\\ &\sum_{i=1}^{c}(\eta_{ij}+\frac{\xi_{ij}}{c})=1,\ 1\le j\le N\end{aligned}$$

Where

$$\mathrm{SNI}_{ij}=\sum_{k\in N_j}\frac{\mu_{ik}(1-K(\mathbf{x}_k,\mathbf{v}_i))}{\sum_{k\in N_j}\mu_{ik}}$$

$K(\mathbf{x}_j,\mathbf{v}_i)=\exp(\frac{-\|\mathbf{x}_j-\mathbf{v}_i\|^2}{\sigma^2})$, cluster centers $\mathbf{v}_i\ \ \forall i=1,2,\ldots,c$ and $\mathbf{x}_j\ \forall j=1,2,\ldots,N$. After solving the optimization problem (27) using Lagrange method of undetermined multiplier, the membership value $\mu_{ij}$, cluster center $\mathbf{v}_i$, neutrality degree $\eta_{ij}$ and refusal degree $\xi_{ij}$ can, respectively, be given as:

$$\mu_{ij}=(1-\eta_{ij}-\xi_{ij})\frac{\left((1-K(\mathbf{x}_j,\mathbf{v}_i))+\beta\mathrm{SNI}_{ij}\right)^{-\frac{1}{m-1}}}{\sum_{l=1}^{c}\left((1-K(\mathbf{x}_j,\mathbf{v}_l))+\beta\mathrm{SNI}_{lj}\right)^{-\frac{1}{m-1}}} \qquad (28)$$

$$\mathbf{v}_i=\frac{\sum_{j=1}^{N}\left(\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}\right)^{m}\left(K(\mathbf{x}_j,\mathbf{v}_i)\mathbf{x}_j+\beta\sum_{k\in N_j}\frac{\mu_{ik}K(\mathbf{x}_k,\mathbf{v}_i)\mathbf{x}_k}{\sum_{k\in N_j}\mu_{ik}}\right)}{\sum_{j=1}^{N}\left(\frac{\mu_{ij}}{1-\eta_{ij}-\xi_{ij}}\right)^{m}\left(K(\mathbf{x}_j,\mathbf{v}_i)+\beta\sum_{k\in N_j}\frac{\mu_{ik}K(\mathbf{x}_k,\mathbf{v}_i)}{\sum_{k\in N_j}\mu_{ik}}\right)} \qquad (29)$$

$$\eta_{ij}=\frac{\xi_{ij}(2\eta_{ij}+\xi_{ij})}{c(\eta_{ij}+2\xi_{ij})} \qquad (30)$$

$$\xi_{ij}=1-(\mu_{ij}+\eta_{ij})-(1-(\mu_{ij}+\eta_{ij})^{\alpha})^{\frac{1}{\alpha}} \qquad (31)$$

The key features of the proposed KPFCMSNI method are as follows:

– The clustering in PFS framework effectively deals with vagueness as the representation capability of cluster is increased in PFS framework.
– It reveals the inherent non-Euclidean structures by incorporating the kernel space transformation and that too without increasing the computational complexity.
– The proposed approach uses the neighborhood membership values to dampen the effect of noise that avoid smoothing and preserves fine details in the image segmentation process.
– The proposed method is robust to noise and nonlinear structure present in image.

Notations related to proposed methodologies are described in Table 1. Outline of the iterative procedure for finding the solution of the proposed method is given below.

**Algorithm 1** Proposed KPFCMSNI algorithm.

**Require:** Set values for the degree of fuzziness $m > 1$, negation parameter $\alpha$, spatial regularization parameter $\beta$, the number of clusters $c$ and the error $\epsilon$.

1: Initialize the cluster centroid $v_i^{(k)}$, $1 \le i \le c$ randomly
2: Calculate the membership $U^{(0)} = \{\mu_{ij}^{(0)}\}_{c\times N}$ using (28)
3: $k \leftarrow 1$
4: **repeat**
5: Update the centers $v_i^{(k)}$, $1 \le i \le c$ using (29)
6: Compute the membership matrix $U^{(k)} = \{\mu_{ij}^{(k)}\}_{c\times N}$, $1 \le i \le c$ and $1 \le j \le N$ using (28)
7: Update the $\eta_{ij}$, $1 \le i \le c$ and $1 \le j \le N$ using (30)
8: Update the $\xi_{ij}$, $1 \le i \le c$ and $1 \le j \le N$ using (31)
9: $k \leftarrow k + 1$
10: **until** $||U^{(k+1)} - U^{(k)}|| + ||\eta^{(k+1)} - \eta^{(k)}|| + ||\xi^{(k+1)} - \xi^{(k)}|| < \epsilon$
11: **return** the centers of clusters $v_i$ and the membership degrees $\mu_{ij}$.

The flowchart of the proposed KPFCMSNI method for image segmentation process is given in Fig. 1.

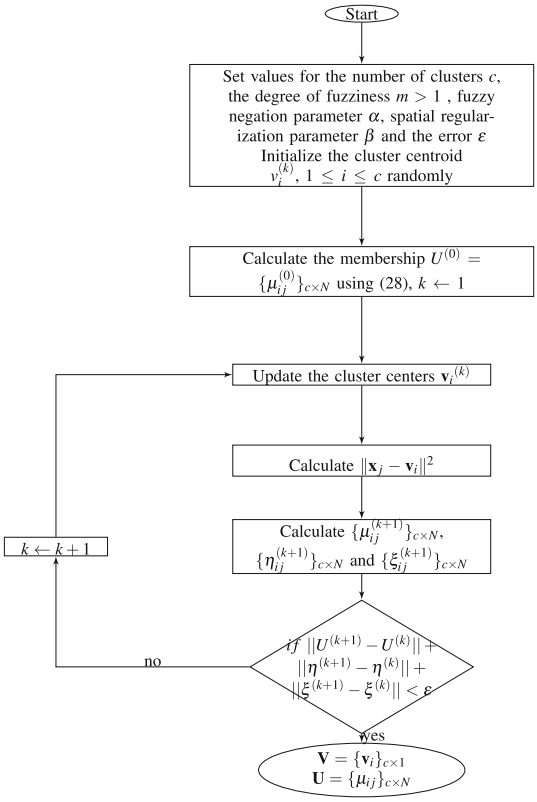



**Fig. 1** Flow chart of the proposed KPFCMSNI method

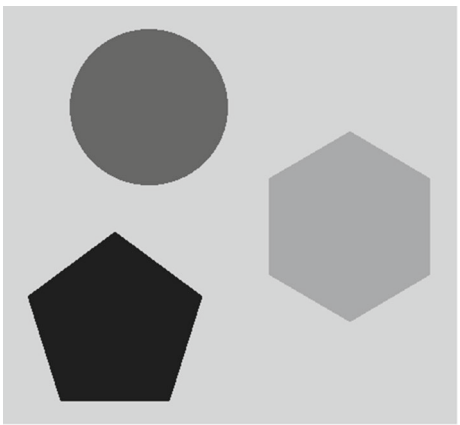


**Fig. 2** Synthetic gray image

## 5 Experimental setup and results

To check the effectiveness of the proposed KPFCMSNI method over related works FCM, FC_PFS, RPFCM, FCM_S, FLICM, IPFCM, AIPFCM, APFCMS, IFCMSNI, KWFLICM, NDFCM, KGFCM, IIFCM, FRFCM, DSFCMN and WRFCM, experiments have been performed on one synthetic dataset and two publicly available datasets. The performance measures used for comparison are average segmentation accuracy ($ASA$) and Dice score ($DS$) (Chuang et al. 2006; Gokul and Gopal 2017; Qiu et al. 2013; Vovk et al. 2007), which are, respectively, given as:

$$ASA = \sum_{i=1}^{c} \frac{|X_i \cap Y_i|}{\sum_{j=1}^{c} |X_j|} \tag{32}$$

$$\mathrm{DS} = \frac{2|X_i \cap Y_i|}{|X_i| + |Y_i|} \tag{33}$$

where $c$ is the number of cluster, $N$ is the total number of datapoints, $\mu_{ij}$ is the membership of $j$th datapoint in $i$th cluster, $m$ is the fuzzifier factor, $X_i$ denotes the pixels belonging to the manual segmented image (ground truth), $Y_i$ denotes the pixels belonging to the experimental segmented image and $|X_i|$ denotes the cardinality of $X_i$. The optimal values of parameters involved for each of the methods are obtained using grid search.

**Table 1** Notations related to proposed methodology

| Notations | Description |
|---|---|
| $\phi(x_i)$ | Data in higher-dimensional space |
| $\mu_A(x)$ | Positive membership function where $A$ is a fuzzy/PFS set defined over a set $X$ |
| $\eta_A(x)$ | Neutral membership function where $A$ is a fuzzy/PFS set defined over a set $X$ |
| $\gamma_A(x)$ | Negative membership function where $A$ is a fuzzy/PFS set defined over a set $X$ |
| $\xi_A(x)$ | The refusal degree of x where $A$ is a fuzzy/PFS set defined over a set $X$ |
| $K(x_i, x_j)$ | Kernel function to measure distance between $x_i$ and $x_j$ in higher-dimensional space |
| $\mathbf{x}_j$ | $j$th datapoint |
| $\mathbf{v}_i$ | $i$th cluster centroid |
| $m$ | Degree of fuzziness |
| $\mu_{ij}$ | Membership degree corresponding to $j$th data point to $i$th cluster |
| $\eta_{ij}$ | Neutrality degree corresponding to $j$th data point to $i$th cluster |
| $\xi_{ij}$ | Refusal degree corresponding to $j$th data point to $i$th cluster |
| $N$ | The set of data points |
| $c$ | The set of groups to be clustered in |
| $\bar{x}$ | The mean of the data point around the neighborhood of $x$ |
| $\beta$ | Spatial regularization parameter |

### 5.1 Datasets

To validate the performance of the proposed approach, a synthetic gray image is generated. The synthetic gray image has four regions with intensities 4, 76, 148 and 201 (Fig. 2). For experimental purpose, this gray image is corrupted with Gaussian noise, Rician noise and salt and pepper noise. Other than this synthetic gray image dataset, we have utilized two publicly available MRI datasets.

BrainWeb simulated MRI brain volumes, a publicly available dataset[1] from the McConnell Brain Imaging Center of the Montreal Neurological Institute, McGill University (Cocosco et al. 1997) consist of several simulated $T1$-weighted brain volume data with different intensity inhomogeneity (0%, 20% and 40%) and noise (1%, 3%, 5%, 7% and 9%) of resolution $1 \times 1 \times 1$ mm with $181 \times 217 \times 181$ dimension.

Another publicly available real MRI brain images have been acquired from the Internet Brain Segmentation Repository (IBSR)[2] along with the given ground truth. For all the MRI images, scull striping is done using the brain extraction tool.[3]

[1] BrainWeb [online], available: http://www.brainweb.bic.mni.mcgill.ca/brainweb.

[2] IBSR [online], available: http://www.cma.mgh.harvard.edu/ibsr/.

[3] Brain Extraction Tool (BET) [online], available: http://www.fmrib.ox.ac.uk/fsl/.

### 5.2 Results on synthetic dataset

We demonstrate the robustness of the proposed method in handling various types of noise with different levels. We conducted the segmentation experiments on synthetic image data with added Gaussian white noise (standard deviations of 10, 15 and 20), Rician noise (standard deviations of 5, 10 and 15) and salt and pepper noise (noisy pixel percentage of 5, 10 and 15). The results measuring average segmentation accuracy and Dice score for these experiments are shown in Tables 2, 3, 4, 5, 6 and 7. The results show that the proposed method KPFCMSNI performs better than other comparative methods in suppressing different types of noise. The IFCMSNI method also gives good segmentation results for the image corrupted by Rician noise and salt and pepper noise. Figure 3 shows the qualitative result obtained on the synthetic gray image corrupted with Gaussian white noise with standard deviation 10. It can be noted from Fig. 3 that the segmentation result obtained for the proposed KPFCMSNI method is better in comparison with the state-of-the-art methods. The proposed approach suppresses the noise without loosing fine structural details of image in the segmentation process.

### 5.3 Results on simulated MRI dataset

The results for experiments conducted on simulated MRI BrainWeb dataset for various measures are recorded in Tables 8, 9, 10 and 11. From the experimental results, it is evident that KPFCMSNI performs better than other methods in most of the cases. Average segmentation accuracy and Dice scores of FCM, FC_PFS and RPFCM are nearly the same for each image in the dataset and are also lower than those obtained

**Table 2** Average segmentation accuracy for synthetic image dataset corrupted by Gaussian noise

| ImageName | FCM | FC_PFS | RPFCM | FLICM | FCM_S | IPFCM | AIPFCM | APFCMS | IFCMSNI | KWFLICM | NDFCM | KGFCM | IIFCM | FRFCM | DSFCMN | WRFCM | KPFCMSNI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GN_10 | 0.7909 | 0.8054 | 0.7909 | 0.9933 | 0.9954 | 0.9836 | 0.9812 | 0.9812 | 0.9994 | 0.9997 | 0.9892 | 0.9546 | 0.9966 | 0.9883 | 0.9836 | 0.9913 | 0.9994 |
| GN_15 | 0.6365 | 0.6412 | 0.6365 | 0.9707 | 0.9780 | 0.9735 | 0.9804 | 0.9805 | 0.9619 | 0.9990 | 0.9869 | 0.9807 | 0.9645 | 0.9866 | 0.9576 | 0.9836 | 0.9967 |
| GN_20 | 0.5672 | 0.5742 | 0.5587 | 0.9371 | 0.6607 | 0.9592 | 0.9758 | 0.9767 | 0.8404 | 0.9972 | 0.9790 | 0.9750 | 0.9111 | 0.9813 | 0.9445 | 0.9789 | 0.9909 |

**Table 3** Dice score for synthetic image dataset corrupted by Gaussian noise

| Images\Methods | | FCM | FC_PFS | RPFCM | FCM_S | FLICM | IPFCM | AIPFCM | APFCMS | IFCMSNI | KWFLICM | NDFCM | KGFCM | IIFCM | FRFCM | DSFCMN | WRFCM | KPFCMSNI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GN_10 | Region 1 | 0.9751 | 0.9765 | 0.9751 | 0.9999 | 0.9841 | 0.9902 | 0.9880 | 0.9880 | 1.0000 | 0.9998 | 0.9878 | 0.9882 | 0.9994 | 0.9989 | 0.9781 | 0.9883 | 1.0000 |
| | Region 2 | 0.9022 | 0.9143 | 0.9022 | 0.9919 | 0.9720 | 0.9872 | 0.9788 | 0.9787 | 0.9999 | 0.9999 | 0.9762 | 0.9801 | 0.9987 | 0.9984 | 0.9673 | 0.9823 | 0.9996 |
| | Region 3 | 0.5092 | 0.5328 | 0.5092 | 0.9812 | 0.9951 | 0.9447 | 0.9381 | 0.9380 | 0.9973 | 0.9989 | 0.9691 | 0.9403 | 0.9868 | 0.955 | 0.9573 | 0.9751 | 0.9975 |
| | Region 4 | 0.8363 | 0.8482 | 0.8363 | 0.9977 | 0.9979 | 0.9896 | 0.9890 | 0.9890 | 0.9995 | 0.9997 | 0.9952 | 0.9892 | 0.9977 | 0.9915 | 0.9921 | 0.9961 | 0.9996 |
| GN_15 | Region 1 | 0.9014 | 0.9072 | 0.9014 | 0.9873 | 0.9906 | 0.9889 | 0.9878 | 0.9878 | 0.9995 | 0.9994 | 0.9875 | 0.9884 | 0.9874 | 0.9977 | 0.9401 | 0.9902 | 0.9998 |
| | Region 2 | 0.7110 | 0.7330 | 0.7110 | 0.9806 | 0.9477 | 0.9847 | 0.9793 | 0.9793 | 0.9981 | 0.9997 | 0.9768 | 0.9815 | 0.9734 | 0.9958 | 0.9166 | 0.9873 | 0.9975 |
| | Region 3 | 0.3228 | 0.3333 | 0.3228 | 0.9235 | 0.9107 | 0.9100 | 0.9351 | 0.9353 | 0.8650 | 0.9966 | 0.9601 | 0.9335 | 0.8755 | 0.9499 | 0.8849 | 0.9445 | 0.9869 |
| | Region 4 | 0.7105 | 0.7105 | 0.7105 | 0.9864 | 0.9318 | 0.9822 | 0.9883 | 0.9883 | 0.9709 | 0.9988 | 0.9934 | 0.9882 | 0.9772 | 0.9905 | 0.9812 | 0.9895 | 0.9979 |
| GN_20 | Region 1 | 0.8510 | 0.8602 | 0.8510 | 0.9762 | 0.9176 | 0.9869 | 0.9871 | 0.9872 | 0.9834 | 0.9989 | 0.9856 | 0.9871 | 0.9618 | 0.9962 | 0.9376 | 0.9889 | 0.9985 |
| | Region 2 | 0.5532 | 0.5897 | 0.5532 | 0.9221 | 0.8549 | 0.9799 | 0.9782 | 0.9781 | 0.9417 | 0.9991 | 0.9748 | 0.9791 | 0.9063 | 0.9933 | 0.9005 | 0.9839 | 0.9944 |
| | Region 3 | 0.2554 | 0.2690 | 0.2522 | 0.4019 | 0.7885 | 0.8644 | 0.9193 | 0.9222 | 0.5919 | 0.9901 | 0.9317 | 0.9158 | 0.7274 | 0.9314 | 0.8508 | 0.9286 | 0.9639 |
| | Region 4 | 0.6583 | 0.6583 | 0.6467 | 0.6812 | 0.9743 | 0.9717 | 0.9849 | 0.9856 | 0.8759 | 0.9981 | 0.9877 | 0.9841 | 0.9452 | 0.9868 | 0.9709 | 0.9864 | 0.9940 |

**Table 4** Average segmentation accuracy for synthetic image dataset corrupted by Rician noise

| Images\Methods | FCM | FC_PFS | RPFCM | FLICM | FCM_S | IPFCM | AIPFCM | APFCMS | IFCMSNI | KWFLICM | NDFCM | KGFCM | IIFCM | FRFCM | DSFCMN | WRFCM | KPFCMSNI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RN_05 | 1.0000 | 1.0000 | 1.0000 | 0.9890 | 0.9974 | 0.9980 | 0.9817 | 1.0000 | 1.0000 | 1.0000 | 0.9899 | 1.0000 | 1.0000 | 0.9900 | 0.9900 | 0.9917 | 1.0000 |
| RN_10 | 0.9960 | 0.9960 | 0.9960 | 0.9955 | 0.9968 | 0.9961 | 0.9816 | 0.9816 | 1.0000 | 1.0000 | 0.9897 | 0.9993 | 1.0000 | 0.9895 | 0.9813 | 0.9916 | 1.0000 |
| RN_15 | 0.9053 | 0.9241 | 0.8953 | 0.9929 | 0.9956 | 0.9874 | 0.9814 | 0.9814 | 0.9999 | 0.9998 | 0.9896 | 0.9924 | 0.9993 | 0.9889 | 0.9781 | 0.9915 | 0.9982 |

**Table 5** Dice score for synthetic image dataset corrupted by Rician noise

| Images\Methods | | FCM | FC_PFS | RPFCM | FCM_S | FLICM | IPFCM | AIPFCM | APFCMS | IFCMSNI | KWFLICM | NDFCM | KGFCM | IIFCM | FRFCM | DSFCMN | WRFCM | KPFCMSNI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RN_05 | Region 1 | 1.0000 | 1.0000 | 1.0000 | 0.9879 | 0.9882 | 0.9999 | 0.9880 | 1.0000 | 1.0000 | 1.0000 | 0.9881 | 1.0000 | 1.0000 | 1.0000 | 0.9871 | 0.9967 | 1.0000 |
| | Region 2 | 1.0000 | 1.0000 | 1.0000 | 0.9875 | 0.9881 | 0.9999 | 0.9787 | 1.0000 | 1.0000 | 1.0000 | 0.9762 | 1.0000 | 1.0000 | 1.0000 | 0.9824 | 0.9880 | 1.0000 |
| | Region 3 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9570 | 0.9920 | 0.9397 | 1.0000 | 1.0000 | 1.0000 | 0.9721 | 1.0000 | 1.0000 | 0.9604 | 0.9713 | 0.9771 | 1.0000 |
| | Region 4 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9954 | 0.9985 | 0.9894 | 1.0000 | 1.0000 | 1.0000 | 0.9956 | 1.0000 | 1.0000 | 0.9924 | 0.9951 | 0.9811 | 1.0000 |
| RN_10 | Region 1 | 0.9996 | 0.9996 | 0.9996 | 0.9926 | 0.9906 | 0.9993 | 0.9880 | 0.9880 | 1.0000 | 1.0000 | 0.9879 | 0.9999 | 1.0000 | 0.9997 | 0.9808 | 0.9965 | 1.0000 |
| | Region 2 | 0.9994 | 0.9995 | 0.9994 | 0.9878 | 0.9865 | 0.9992 | 0.9789 | 0.9789 | 1.0000 | 1.0000 | 0.9764 | 0.9999 | 1.0000 | 0.9999 | 0.9758 | 0.9880 | 1.0000 |
| | Region 3 | 0.9842 | 0.9843 | 0.9842 | 0.9935 | 0.9837 | 0.9846 | 0.9392 | 0.9393 | 0.9999 | 0.9998 | 0.9711 | 0.9971 | 0.9999 | 0.9588 | 0.9480 | 0.9768 | 0.9998 |
| | Region 4 | 0.9971 | 0.9971 | 0.9971 | 0.9994 | 0.9998 | 0.9971 | 0.9892 | 0.9893 | 0.9999 | 0.9999 | 0.9955 | 0.9994 | 0.9999 | 0.9921 | 0.9887 | 0.9818 | 0.9999 |
| RN_15 | Region 1 | 0.9836 | 0.9846 | 0.9836 | 0.9957 | 0.9861 | 0.9945 | 0.9879 | 0.9879 | 1.0000 | 0.9999 | 0.9878 | 0.9970 | 0.9998 | 0.9988 | 0.9722 | 0.9964 | 0.9997 |
| | Region 2 | 0.9615 | 0.9698 | 0.9584 | 0.9955 | 0.9723 | 0.9937 | 0.9793 | 0.9793 | 1.0000 | 0.9999 | 0.9773 | 0.9961 | 0.9997 | 0.9989 | 0.9620 | 0.9881 | 0.9996 |
| | Region 3 | 0.7194 | 0.7650 | 0.6972 | 0.9858 | 0.9947 | 0.9551 | 0.9385 | 0.9384 | 0.9996 | 0.9994 | 0.9699 | 0.9722 | 0.9972 | 0.9569 | 0.9403 | 0.9764 | 0.9929 |
| | Region 4 | 0.9305 | 0.9450 | 0.9224 | 0.9973 | 0.9969 | 0.9914 | 0.9891 | 0.9891 | 0.9999 | 0.9998 | 0.9954 | 0.9948 | 0.9994 | 0.9918 | 0.9889 | 0.9821 | 0.9987 |

**Table 6** Average segmentation accuracy for synthetic image dataset corrupted by salt and pepper noise

| Images\Methods | FCM | FC_PFS | RPFCM | FLICM | FCM_S | IPFCM | AIPFCM | APFCMS | IFCMSNI | KWFLICM | NDFCM | KGFCM | IIFCM | FRFCM | DSFCMN | WRFCM | KPFCMSNI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SP_05 | 0.9696 | 0.9696 | 0.9696 | 0.8383 | 0.9671 | 0.9679 | 0.9642 | 0.9696 | 0.9918 | 0.9937 | 0.8846 | 0.8374 | 0.9907 | 0.9885 | 0.9585 | 0.9592 | 0.9927 |
| SP_10 | 0.9392 | 0.9392 | 0.9392 | 0.7176 | 0.9322 | 0.9363 | 0.9196 | 0.9392 | 0.9828 | 0.9827 | 0.8112 | 0.7912 | 0.9698 | 0.9804 | 0.9082 | 0.9002 | 0.9884 |
| SP_15 | 0.9087 | 0.9087 | 0.9087 | 0.6061 | 0.8963 | 0.9041 | 0.8594 | 0.9087 | 0.9605 | 0.9633 | 0.7606 | 0.6941 | 0.9406 | 0.9662 | 0.8957 | 0.7011 | 0.9665 |

**Table 7** Dice score for synthetic image dataset corrupted by salt and pepper noise

| Images\Methods | | FCM | FC_PFS | RPFCM | FCM_S | FLICM | IPFCM | AIPFCM | APFCMS | IFCMSNI | KWFLICM | NDFCM | KGFCM | IIFCM | FRFCM | DSFCMN | WRFCM | KPFCMSNI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SP_05 | Region 1 | 0.8950 | 0.8950 | 0.8950 | 0.8951 | 0.9090 | 0.9711 | 0.9786 | 0.8951 | 0.9934 | 0.9809 | 0.9808 | 0.6816 | 0.9632 | 0.9929 | 0.9704 | 0.97068 | 0.9971 |
| | Region 2 | 0.9726 | 0.9726 | 0.9726 | 0.9725 | 0.8739 | 0.8893 | 0.9539 | 0.9726 | 0.9700 | 0.9926 | 0.9565 | 0.0004 | 0.9899 | 0.9941 | 0.9566 | 0.9605 | 0.9815 |
| | Region 3 | 0.9751 | 0.9751 | 0.9751 | 0.9651 | 0.6334 | 0.9686 | 0.8916 | 0.9751 | 0.9848 | 0.9928 | 0.6852 | 0.8455 | 0.9955 | 0.9606 | 0.8724 | 0.87506 | 0.9739 |
| | Region 4 | 0.9813 | 0.9813 | 0.9813 | 0.9794 | 0.8844 | 0.9801 | 0.9786 | 0.9813 | 0.9962 | 0.9961 | 0.9128 | 0.9709 | 0.9946 | 0.9922 | 0.9752 | 0.97491 | 0.9971 |
| SP_10 | Region 1 | 0.8061 | 0.8061 | 0.8061 | 0.9149 | 0.9644 | 0.9391 | 0.9484 | 0.8061 | 0.9880 | 0.9461 | 0.9556 | 0.6831 | 0.8875 | 0.9760 | 0.8274 | 0.93834 | 0.9902 |
| | Region 2 | 0.9473 | 0.9473 | 0.9473 | 0.7688 | 0.8518 | 0.7984 | 0.9021 | 0.9473 | 0.9543 | 0.9797 | 0.9021 | 0.0013 | 0.9721 | 0.9763 | 0.7596 | 0.88547 | 0.9781 |
| | Region 3 | 0.9473 | 0.9473 | 0.9473 | 0.9471 | 0.4615 | 0.9361 | 0.7751 | 0.9473 | 0.9591 | 0.9844 | 0.5658 | 0.7750 | 0.9824 | 0.9476 | 0.7638 | 0.72284 | 0.9628 |
| | Region 4 | 0.9622 | 0.9622 | 0.9622 | 0.9622 | 0.7709 | 0.9601 | 0.9516 | 0.9622 | 0.9908 | 0.9891 | 0.8553 | 0.9410 | 0.9821 | 0.9879 | 0.9679 | 0.94005 | 0.9942 |
| SP_15 | Region 1 | 0.7291 | 0.7291 | 0.7291 | 0.8716 | 0.9044 | 0.8936 | 0.9136 | 0.7291 | 0.9812 | 0.8878 | 0.9209 | 0.6833 | 0.8000 | 0.9490 | 0.8908 | 0.73405 | 0.9852 |
| | Region 2 | 0.9182 | 0.9182 | 0.9182 | 0.6781 | 0.7820 | 0.7236 | 0.8439 | 0.9182 | 0.9018 | 0.9575 | 0.8667 | 0.0434 | 0.9437 | 0.9449 | 0.8415 | 0.21681 | 0.9339 |
| | Region 3 | 0.9208 | 0.9208 | 0.9208 | 0.9079 | 0.3801 | 0.9028 | 0.6509 | 0.9208 | 0.9070 | 0.9682 | 0.5026 | 0.2773 | 0.9646 | 0.9262 | 0.6172 | 0.26724 | 0.8847 |
| | Region 4 | 0.9431 | 0.9431 | 0.9431 | 0.9425 | 0.6603 | 0.9393 | 0.9105 | 0.9431 | 0.9771 | 0.9768 | 0.8121 | 0.9003 | 0.9639 | 0.9791 | 0.9507 | 0.88575 | 0.9824 |

for the other methods; in most of the cases, we have considered in this work. Their accuracies reduce significantly for an image with increased level of Rician noise. Also, their Dice scores corresponding to GM are comparatively lower and give better segmentation results on CSF and WM. The Dice scores obtained for FCM_S and FLICM on CSF are particularly low in comparison with all other methods. We further observe that the methods IPFCM and AIPFCM give very similar values for both the performance metrics, except for the case of CSF, where IPFCM results are better. It is also to be noted that the method APFCMS does not give good results on WM and GM images, especially the ones with high intensity of noise. The results obtained using IFCMSNI are also quite close to the results of the proposed KPFCMSNI method.

Figures 4, 5 and 6 show the variation in performance, in terms of ASA, of all the methods considered in this work for 0%, 20% and 40% intensity non-uniformity, respectively. For a fixed INU level, we consider six different Rician noise levels 0%, 1%, 3%, 5%, 7% and 9% and observe that KPFCMSNI is very effective in handling the noise. The ASA of all the methods decrease with the increase in the intensity non-uniformity. But the performance of the proposed KPFCMSNI method does not degrade much and performs better than all other state-of-the-art methods even in noisy images with high level of intensity non-uniformity.

### 5.4 Results on real MRI dataset

The experimental results on real MRI dataset shown in Tables 12, 13 and 14 point out that the proposed method KPFCMSNI gives best results in terms of ASA in almost all the cases. Also, the DS of KPFCMSNI corresponding to both GM and WM is very close to the optimal values. In case of GM, we notice that the Dice scores of FLICM are quite high for some of the image cases. But we can also observe that for all these cases, for instance 110_03, 1_24, etc., the results obtained using KPFCMSNI are better among all the methods other than FLICM. Average segmentation accuracy and Dice scores of FCM, FC_PFS and RPFCM are nearly the same for each image in the dataset and are also lower than those obtained for the other methods; in most of the cases, we have considered in this work. Their Dice scores for GM are comparatively lower. FCM_S results are poor for some of the images like 205_03, 1_24, etc. corresponding to GM. The methods IPFCM and AIPFCM give similar values for both the performance metrics. For real MRI images, APFCMS gives better results than both IPFCM and AIPFCM, where the results are significantly lower than other methods. Unlike in the case of simulated brain images dataset, we observe that the performance of IFCMSNI is not good on GM.

Following are some key observations based on the experimental results:

1. The methods FCM, FC_PFS and RPFCM, which do not consider the spatial information, are not effective in suppressing the noise. This is evident from their poor ASA and DS scores in the experimental results.
2. Taking local spatial information into account, as done in FLICM, FCM_S and IPFCM, helps in denoising to some extent but at the cost of loss in finer details resulting in edge blurring.

### 5.5 Statistical test

Friedman test, a two-way nonparametric statistical test, is conducted to find out the significant difference among the proposed and other segmentation methods for both the publicly available datasets. The null hypothesis ($H_0$) of this test is that there is no significant difference in the performance of the proposed and other segmentation methods, whereas the alternative hypothesis ($H_1$) defines as the performance of the proposed and other methods is different. For a given performance measure $M$, the $H_0$ and $H_1$ can be defined as:

$$\begin{aligned} H_0 : \mu_{\text{FCM}} &= \mu_{\text{FC-PFS}} = \mu_{\text{RPFCM}} = \mu_{\text{FLICM}} \\ &= \mu_{\text{FCM_S}} = \mu_{\text{IPFCM}} \\ &= \mu_{\text{AIPFCM}} = \mu_{\text{APFCMS}} = \mu_{\text{IFCMSNI}} \\ &= \mu_{\text{KWFLICM}} = \mu_{\text{NDFCM}} \\ &= \mu_{\text{KGFCM}} \\ &= \mu_{\text{IIFCM}} = \mu_{\text{FRFCM}} = \mu_{\text{DSFCMN}} \\ &= \mu_{\text{WRFCM}} = \mu_{\text{KPFCMSNI}} \end{aligned} \tag{34}$$

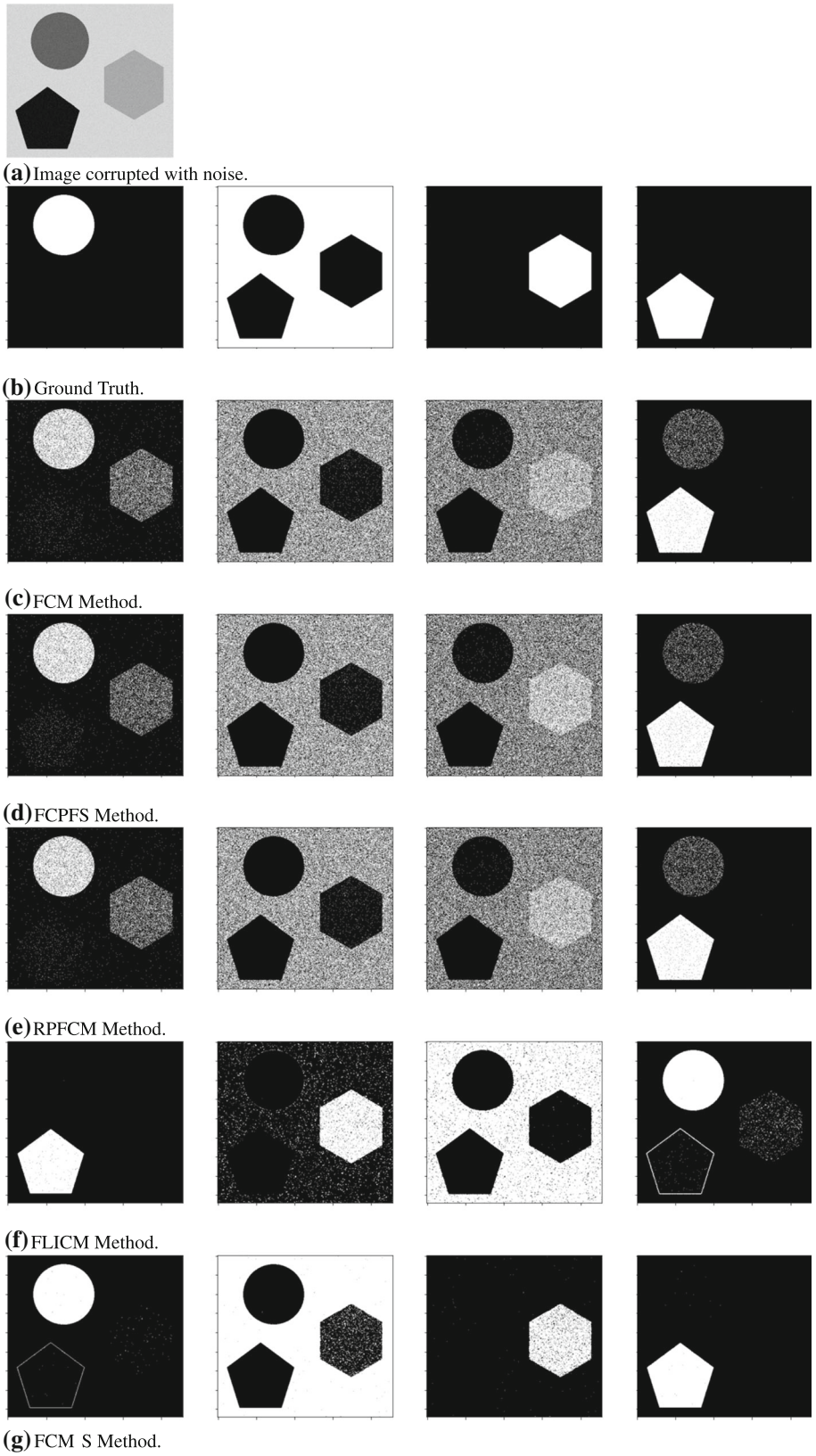
$$\begin{aligned} H_1 : \mu_{\text{FCM}} &\neq \mu_{\text{FC-PFS}} \neq \mu_{\text{RPFCM}} \\ &\neq \mu_{\text{FLICM}} \neq \mu_{\text{FCM_S}} \neq \mu_{\text{IPFCM}} \\ &\neq \mu_{\text{AIPFCM}} \neq \mu_{\text{APFCMS}} \neq \mu_{\text{IFCMSNI}} \\ &\neq \mu_{\text{KWFLICM}} \neq \mu_{\text{NDFCM}} \neq \mu_{\text{KGFCM}} \\ &\neq \mu_{\text{IIFCM}} \neq \mu_{\text{FRFCM}} \neq \mu_{\text{DSFCMN}} \\ &\neq \mu_{\text{WRFCM}} \neq \mu_{\text{KPFCMSNI}}. \end{aligned} \tag{35}$$

The rank of different segmentation methods, according to the different performance measures, is obtained for comparing the methods separately. In the Friedman test, the average rank $R_j$ of $j$th methods for a given $N$ number of images is obtained with respect to a given performance measure as:

$$R_j = \frac{1}{N} \sum_{i=1}^{N} r_i^j \tag{36}$$

where $r_i^j \in \{1, 2, \ldots, k\} (1 \leq i \leq N, 1 \leq j \leq k)$ is rank value for $i$th image and $j$th method. Table 15 shows the average Friedman ranking of different segmentation methods corresponding to ASA for 18 BrainWeb brain images, 10 real brain images and 9 synthetic images used for experiment

**Fig. 3** Synthetic gray image corrupted with noise (Cont..)



**(a)** Image corrupted with noise.

**(b)** Ground Truth.

**(c)** FCM Method.

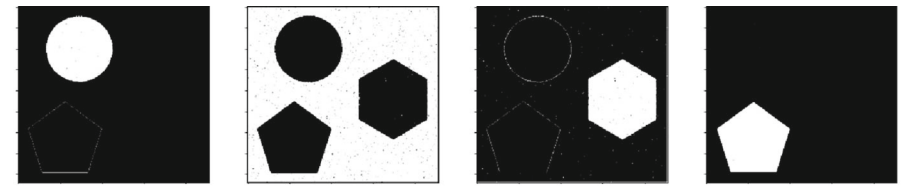
**(d)** FCPFS Method.

**(e)** RPFCM Method.

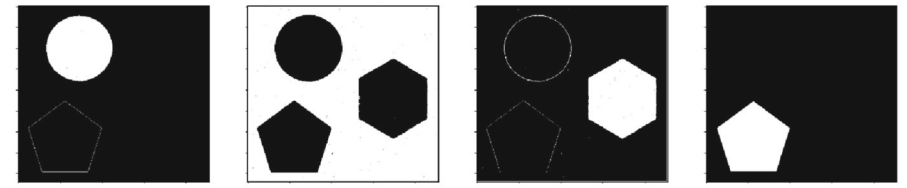
**(f)** FLICM Method.

**(g)** FCM_S Method.

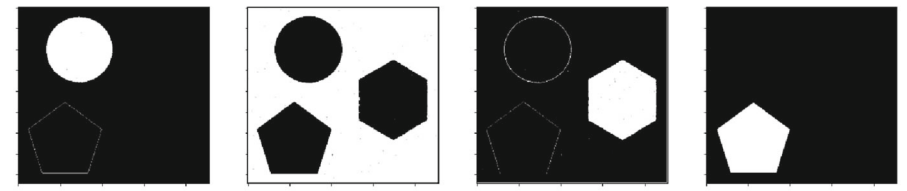
**Fig. 3** continued



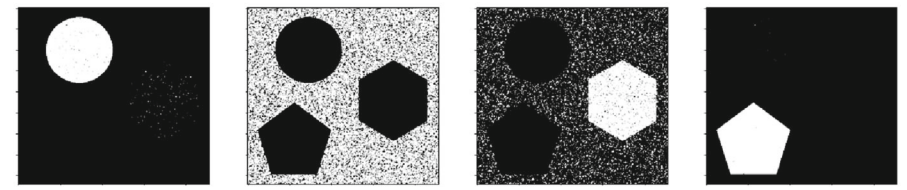
**(h)** IPFCM Method.



**(i)** AIPFCM Method.

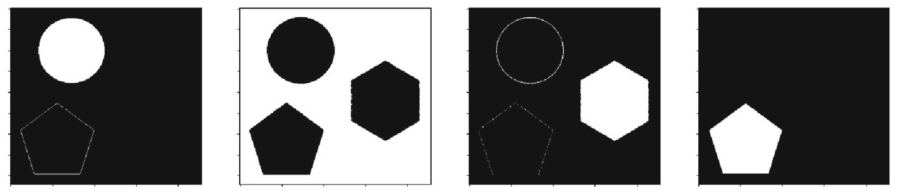


**(j)** APFCMS Method.

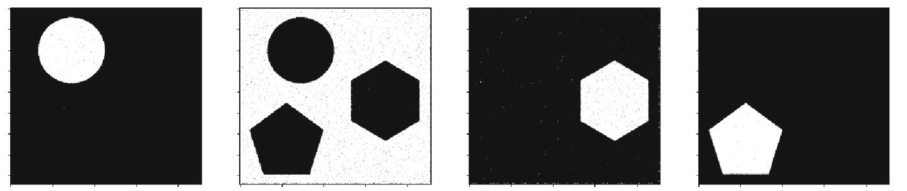


**(k)** IFCMSNI Method.

**(l)** KWFLICM Method.

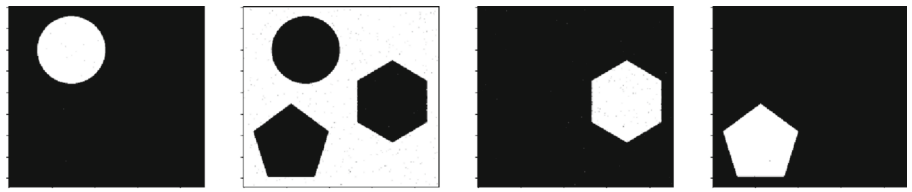


**(m)** NDFCM Method.

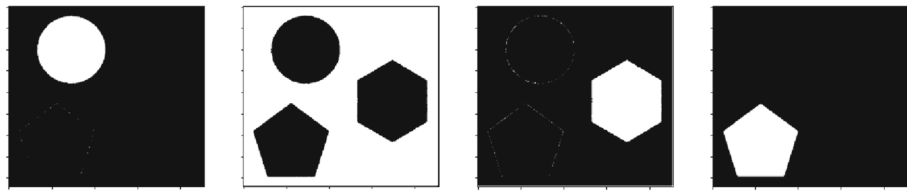


**(n)** KGFCM Method.

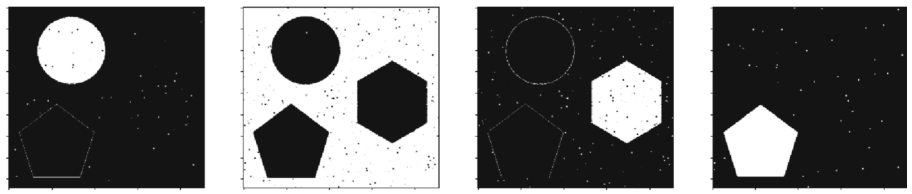
**Fig. 3** continued



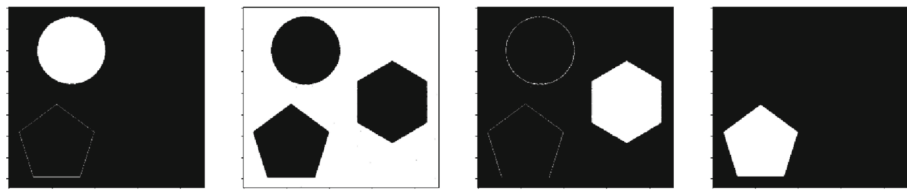
**(o)** IIFCM Method.



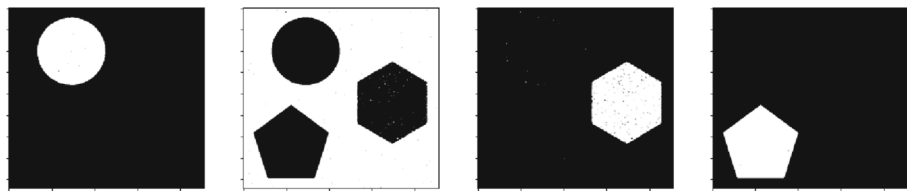
**(p)** FRFCM Method.



**(q)** DSFCMN Method.



**(r)** WRFCM Method.



**(s)** KPFCMSNI Method.

(Derrac et al. 2011; Friedman 1937). Lowest numerical value of rank for a segmentation method shows its better performance compared to other methods for a given performance measure. On the basis of Friedman ranking, the proposed method KPFCMSNI performs better in terms of $ASA$. The statistical hypothesis test proposed by Iman and Davenportis is used. The statistic $F_{\mathrm{ID}}$ is defined by Iman and Davenport (Iman and Davenport 1980) as:

$$F_{\mathrm{ID}} = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2} \tag{37}$$

which is distributed according to F-distribution with $k-1$ and $(k-1)(N-1)$ degrees of freedom, where $\chi_F^2$ is the Friedman's statistic defined as $\frac{12N}{k(k+1)}\left[\sum_j R_j^2 - \frac{k(k+1)^2}{4}\right]$. In our experiments, $k = 17$ and $N = 37$. The $p$-values obtained by Iman and Davenport statistic are 1.19E-36 corresponding to the performance measures $ASA$, which advocate the rejection of null hypothesis $H_0$ as there is significant difference among different segmentation methods at the significance level of 0.05.

However, these $p$-values so obtained are not suitable for comparison with the control method, i.e., the one that emerges with the lowest rank. So adjusted $p$-values (Derrac et al. 2011) are computed, which take into account the error accumulated and provide the correct correlation. This is done with respect to a control method, which is the proposed method KPFCMSNI (lowest rank for $ASA$). For this, a set of post hoc procedures are defined and adjusted $p$-values are computed. The most widely used post hoc method

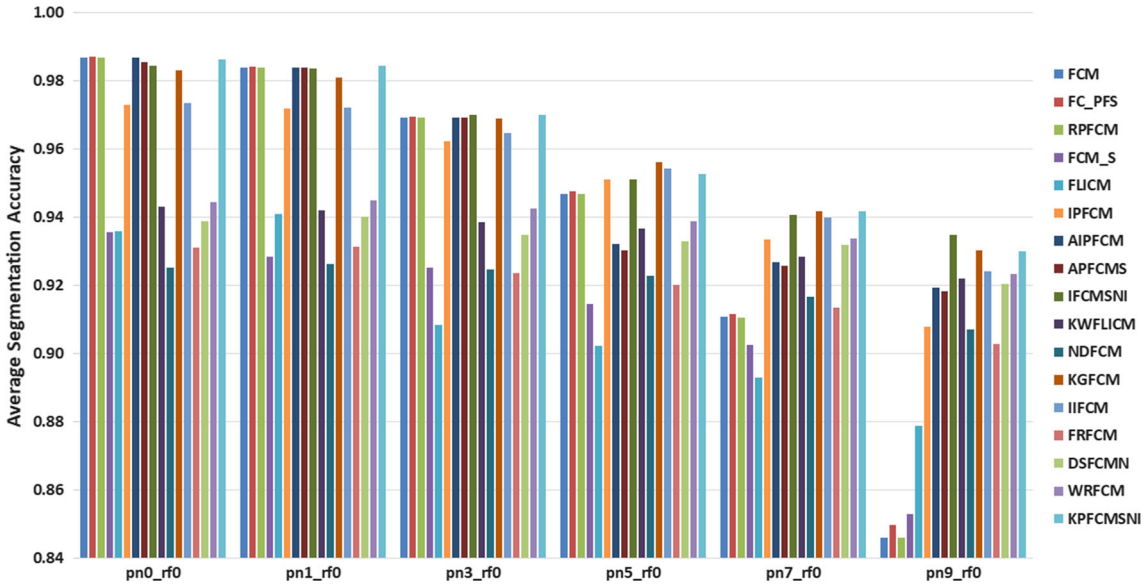



**Fig. 4** Variation in performance in terms of ASA on images with 0% INU

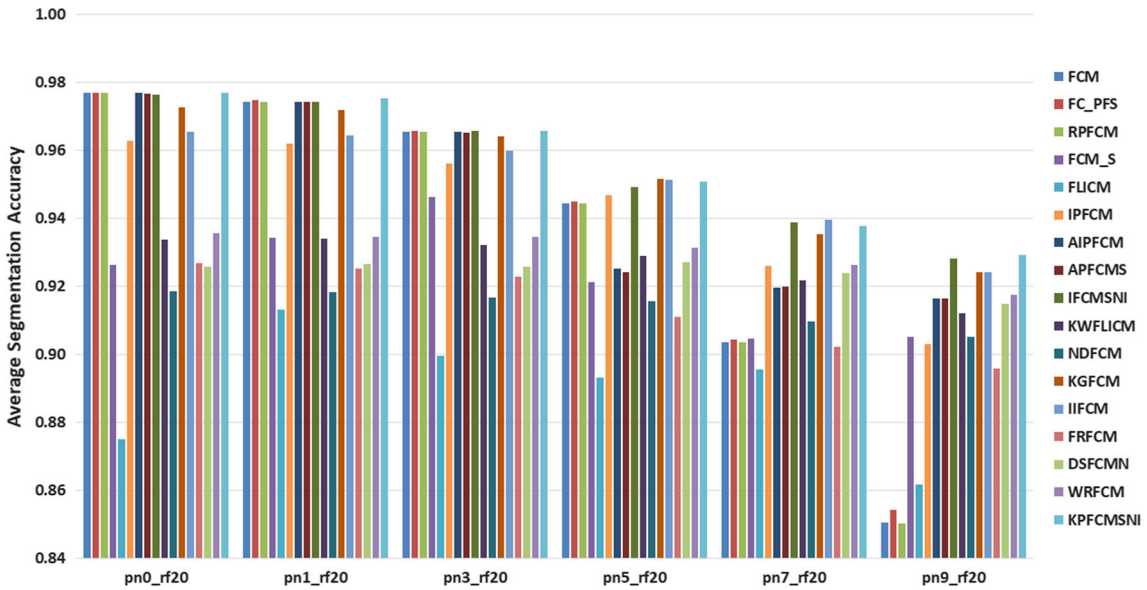



**Fig. 5** Variation in performance in terms of ASA on images with 20% INU

(Derrac et al. 2011) to obtain adjusted $p$-values is Holm procedure. Table 16 shows the various values of adjusted $p$-values obtained. From this table, it indicates that the performance of proposed KPFCMSNI method in terms of ASA performance measures has no significant difference. Also, it can be noted that there is statistical difference in terms of $ASA$ between proposed method and IFCMSNI method.

## 6 Conclusion

In this research work, a picture fuzzy set theory-based fuzzy clustering method is discussed to enhance the MRI image segmentation task in the presence of vague boundary structures, noise and nonlinearity. Few works on picture fuzzy clustering have been reported for handling the noise in segmentation process to handle the vagueness and noise. The optimization problem of the proposed clustering method includes a spatial regularization term that uses the neighbor-

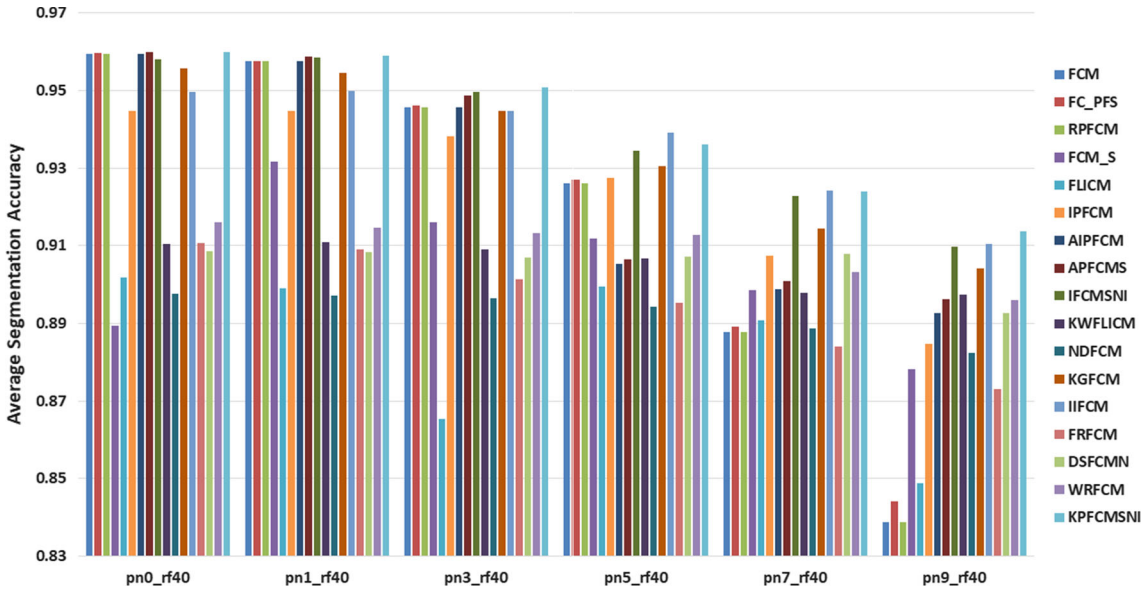



**Fig. 6** Variation in performance in terms of ASA on images with 40% INU

**Table 8** Average segmentation accuracy for BrainWeb dataset

| Methods\Images | pn7_rf0 | pn9_rf0 | pn7_rf20 | pn9_rf20 | pn7_rf40 | pn9_rf40 |
|---|---|---|---|---|---|---|
| FCM | 0.9106 | 0.8459 | 0.9036 | 0.8506 | 0.8877 | 0.8388 |
| FC_PFS | 0.9115 | 0.8496 | 0.9045 | 0.8543 | 0.8892 | 0.8441 |
| RPFCM | 0.9105 | 0.8459 | 0.9036 | 0.8504 | 0.8877 | 0.8388 |
| FCM_S | 0.9025 | 0.8529 | 0.9046 | 0.9052 | 0.8984 | 0.8783 |
| FLICM | 0.8930 | 0.8788 | 0.8957 | 0.8619 | 0.8908 | 0.8488 |
| IPFCM | 0.9333 | 0.9078 | 0.9262 | 0.9030 | 0.9073 | 0.8848 |
| AIPFCM | 0.9268 | 0.9193 | 0.9197 | 0.9164 | 0.8988 | 0.8927 |
| APFCMS | 0.9255 | 0.9180 | 0.9199 | 0.9165 | 0.9007 | 0.8962 |
| IFCMSNI | 0.9406 | 0.9347 | 0.9389 | 0.9282 | 0.9229 | 0.9096 |
| KWFLICM | 0.9283 | 0.9220 | 0.9217 | 0.9123 | 0.8979 | 0.8973 |
| NDFCM | 0.9165 | 0.9070 | 0.9099 | 0.9053 | 0.8887 | 0.8824 |
| KGFCM | 0.9417 | 0.9302 | 0.9355 | 0.9242 | 0.9144 | 0.9041 |
| IIFCM | 0.9397 | 0.9239 | 0.9397 | 0.9241 | 0.9243 | 0.9105 |
| FRFCM | 0.9134 | 0.9027 | 0.9024 | 0.8960 | 0.8841 | 0.8730 |
| DSFCMN | 0.9318 | 0.9202 | 0.9240 | 0.9148 | 0.9079 | 0.8926 |
| WRFCM | 0.9335 | 0.9232 | 0.9264 | 0.9175 | 0.9032 | 0.8959 |
| KPFCMSNI | 0.9416 | 0.9299 | 0.9379 | 0.9292 | 0.9241 | 0.9138 |

**Table 9** Dice score for WM BrainWeb dataset

| Methods\Images | pn7_rf0 | pn7_rf20 | pn7_rf40 | pn9_rf0 | pn9_rf20 | pn9_rf40 |
|---|---|---|---|---|---|---|
| FCM | 0.9348 | 0.9286 | 0.9144 | 0.8843 | 0.8878 | 0.8794 |
| FC_PFS | 0.9351 | 0.9288 | 0.9144 | 0.8851 | 0.8890 | 0.8803 |
| RPFCM | 0.9348 | 0.9286 | 0.9142 | 0.8843 | 0.8878 | 0.8794 |
| FCM_S | 0.9239 | 0.9518 | 0.9450 | 0.9517 | 0.9318 | 0.9260 |
| FLICM | 0.9589 | 0.9540 | 0.9296 | 0.9510 | 0.9527 | 0.9193 |
| IPFCM | 0.9520 | 0.9458 | 0.9283 | 0.9314 | 0.9277 | 0.9111 |
| AIPFCM | 0.9528 | 0.9478 | 0.9286 | 0.9477 | 0.9462 | 0.9245 |
| APFCMS | 0.9377 | 0.9328 | 0.9216 | 0.8996 | 0.9016 | 0.8919 |
| IFCMSNI | 0.9683 | 0.9688 | 0.9624 | 0.9496 | 0.9419 | 0.9313 |
| KWFLICM | 0.9595 | 0.9531 | 0.9288 | 0.9534 | 0.9474 | 0.9297 |
| NDFCM | 0.9475 | 0.9419 | 0.9234 | 0.9413 | 0.9395 | 0.9189 |
| KGFCM | 0.9591 | 0.9546 | 0.9432 | 0.9525 | 0.9474 | 0.9269 |
| IIFCM | 0.9603 | 0.9574 | 0.9435 | 0.9478 | 0.9465 | 0.9338 |
| FRFCM | 0.9518 | 0.9424 | 0.9242 | 0.9435 | 0.9366 | 0.916 |
| DSFCMN | 0.9557 | 0.9496 | 0.9318 | 0.9489 | 0.9463 | 0.9218 |
| WRFCM | 0.9575 | 0.9522 | 0.9311 | 0.9511 | 0.9472 | 0.9273 |
| KPFCMSNI | 0.9691 | 0.9682 | 0.9630 | 0.9526 | 0.9454 | 0.9377 |

**Table 10** Dice score for GM BrainWeb dataset

| Methods\Images | pn7_rf0 | pn7_rf20 | pn7_rf40 | pn9_rf0 | pn9_rf20 | pn9_rf40 |
|---|---|---|---|---|---|---|
| FCM | 0.8817 | 0.8718 | 0.8534 | 0.8018 | 0.8101 | 0.7923 |
| FC_PFS | 0.8825 | 0.8726 | 0.8551 | 0.8063 | 0.8141 | 0.7989 |
| RPFCM | 0.8816 | 0.8716 | 0.8531 | 0.8016 | 0.8101 | 0.7923 |
| FCM_S | 0.8396 | 0.8512 | 0.8721 | 0.7898 | 0.8915 | 0.8487 |
| FLICM | 0.9137 | 0.9061 | 0.8569 | 0.8778 | 0.9021 | 0.8113 |
| IPFCM | 0.9095 | 0.9006 | 0.8769 | 0.8759 | 0.8714 | 0.8484 |
| AIPFCM | 0.9033 | 0.8955 | 0.8687 | 0.8930 | 0.8916 | 0.8611 |
| APFCMS | 0.8838 | 0.8765 | 0.8626 | 0.8183 | 0.8276 | 0.8125 |
| IFCMSNI | 0.9417 | 0.9424 | 0.9307 | 0.9103 | 0.8973 | 0.8748 |
| KWFLICM | 0.8791 | 0.8774 | 0.8715 | 0.8725 | 0.8645 | 0.8704 |
| NDFCM | 0.8502 | 0.8447 | 0.8304 | 0.8364 | 0.8362 | 0.8254 |
| KGFCM | 0.9212 | 0.9134 | 0.8861 | 0.9067 | 0.8989 | 0.8731 |
| IIFCM | 0.9169 | 0.9171 | 0.8983 | 0.894 | 0.8956 | 0.8777 |
| FRFCM | 0.8857 | 0.8718 | 0.8496 | 0.8702 | 0.8643 | 0.8367 |
| DSFCMN | 0.9128 | 0.8999 | 0.8795 | 0.8961 | 0.8908 | 0.857 |
| WRFCM | 0.9127 | 0.9043 | 0.8744 | 0.8989 | 0.8934 | 0.865 |
| KPFCMSNI | 0.9445 | 0.9434 | 0.9325 | 0.9146 | 0.8977 | 0.8847 |

**Table 11** Dice score for CSF BrainWeb dataset

| Methods\Images | pn7_rf0 | pn7_rf20 | pn7_rf40 | pn9_rf0 | pn9_rf20 | pn9_rf40 |
|---|---|---|---|---|---|---|
| FCM | 0.9145 | 0.9118 | 0.9007 | 0.8577 | 0.8607 | 0.8490 |
| FC_PFS | 0.9167 | 0.9138 | 0.9037 | 0.8654 | 0.8672 | 0.8613 |
| RPFCM | 0.9144 | 0.9116 | 0.9003 | 0.8576 | 0.8605 | 0.8490 |
| FCM_S | 0.8980 | 0.8013 | 0.8479 | 0.7072 | 0.8514 | 0.8664 |
| FLICM | 0.8286 | 0.8265 | 0.8736 | 0.8202 | 0.8128 | 0.8029 |
| IPFCM | 0.9333 | 0.9293 | 0.9189 | 0.9157 | 0.9099 | 0.9001 |
| AIPFCM | 0.8677 | 0.8597 | 0.8524 | 0.8588 | 0.8549 | 0.8454 |
| APFCMS | 0.9186 | 0.9190 | 0.9098 | 0.8816 | 0.8874 | 0.8829 |
| IFCMSNI | 0.9640 | 0.9638 | 0.9533 | 0.9459 | 0.9219 | 0.9236 |
| KWFLICM | 0.9188 | 0.9103 | 0.8819 | 0.9114 | 0.9003 | 0.8797 |
| NDFCM | 0.8895 | 0.8817 | 0.8551 | 0.8776 | 0.8773 | 0.8478 |
| KGFCM | 0.9264 | 0.9321 | 0.9253 | 0.8903 | 0.8861 | 0.8864 |
| IIFCM | 0.9280 | 0.9378 | 0.9289 | 0.9179 | 0.9213 | 0.9168 |
| FRFCM | 0.8312 | 0.8254 | 0.8204 | 0.8267 | 0.8242 | 0.811 |
| DSFCMN | 0.8785 | 0.8726 | 0.862 | 0.8648 | 0.856 | 0.3712 |
| WRFCM | 0.8795 | 0.8719 | 0.8609 | 0.8669 | 0.8612 | 0.8515 |
| KPFCMSNI | 0.9678 | 0.9671 | 0.9569 | 0.8979 | 0.8823 | 0.8714 |

**Table 12** Average segmentation accuracy for IBSR dataset

| Methods\Cases | 110_03 | 111_02 | 112_02 | 11_03 | 12_03 | 13_03 | 16_03 | 202_03 | 205_03 | 1_24 |
|---|---|---|---|---|---|---|---|---|---|---|
| FCM | 0.5906 | 0.6133 | 0.5707 | 0.6266 | 0.5766 | 0.5577 | 0.4726 | 0.6044 | 0.5855 | 0.6055 |
| FC_PFS | 0.5969 | 0.6215 | 0.5731 | 0.6315 | 0.5794 | 0.5596 | 0.4783 | 0.6098 | 0.5906 | 0.6116 |
| RPFCM | 0.5909 | 0.6135 | 0.5708 | 0.6269 | 0.5772 | 0.5588 | 0.4707 | 0.6057 | 0.5856 | 0.6058 |
| FCM_S | 0.6898 | 0.6890 | 0.6267 | 0.6346 | 0.6888 | 0.6401 | 0.5823 | 0.6522 | 0.6534 | 0.7048 |
| FLICM | 0.7269 | 0.6423 | 0.5984 | 0.6490 | 0.6632 | 0.3366 | 0.5334 | 0.6670 | 0.6565 | 0.7112 |
| IPFCM | 0.6329 | 0.6454 | 0.6051 | 0.6656 | 0.6125 | 0.5938 | 0.5294 | 0.6306 | 0.6083 | 0.6357 |
| AIPFCM | 0.6017 | 0.6331 | 0.5842 | 0.6406 | 0.5958 | 0.5703 | 0.5743 | 0.6222 | 0.6012 | 0.6491 |
| APFCMS | 0.6582 | 0.6763 | 0.6773 | 0.6824 | 0.6368 | 0.5962 | 0.6877 | 0.6748 | 0.6997 | 0.6839 |
| IFCMSNI | 0.7098 | 0.7102 | 0.6864 | 0.7217 | 0.7156 | 0.7005 | 0.6106 | 0.7068 | 0.6943 | 0.6907 |
| KWFLICM | 0.5148 | 0.6790 | 0.5662 | 0.7125 | 0.7239 | 0.6611 | 0.5910 | 0.5624 | 0.5448 | 0.6893 |
| NDFCM | 0.7521 | 0.7196 | 0.7042 | 0.7389 | 0.7290 | 0.7761 | 0.6998 | 0.7088 | 0.7129 | 0.6945 |
| KGFCM | 0.7175 | 0.6100 | 0.6769 | 0.6181 | 0.5769 | 0.6493 | 0.6793 | 0.6076 | 0.6941 | 0.6390 |
| IIFCM | 0.7055 | 0.7129 | 0.6872 | 0.7448 | 0.7031 | 0.7348 | 0.6748 | 0.7077 | 0.7249 | 0.6734 |
| FRFCM | 0.6695 | 0.6762 | 0.6742 | 0.7156 | 0.7089 | 0.6521 | 0.6858 | 0.6811 | 0.6924 | 0.6813 |
| DSFCMN | 0.7397 | 0.7062 | 0.6888 | 0.6195 | 0.7119 | 0.6971 | 0.6633 | 0.7062 | 0.7479 | 0.6799 |
| WRFCM | 0.6919 | 0.6723 | 0.7193 | 0.6935 | 0.7106 | 0.6770 | 0.6735 | 0.6678 | 0.6902 | 0.6714 |
| KPFCMSNI | 0.7287 | 0.7254 | 0.7266 | 0.7240 | 0.7510 | 0.7897 | 0.7003 | 0.7145 | 0.7111 | 0.6957 |

**Table 13** Dice score for WM IBSR dataset

| Methods\Cases | 110_03 | 111_02 | 112_02 | 11_03 | 12_03 | 13_03 | 16_03 | 202_03 | 205_03 | 1_24 |
|---|---|---|---|---|---|---|---|---|---|---|
| FCM | 0.7511 | 0.7625 | 0.7334 | 0.7785 | 0.7423 | 0.7352 | 0.5818 | 0.7738 | 0.7390 | 0.7716 |
| FC_PFS | 0.7503 | 0.7626 | 0.7315 | 0.7786 | 0.7388 | 0.7311 | 0.5804 | 0.7735 | 0.7379 | 0.7718 |
| RPFCM | 0.7511 | 0.7625 | 0.7335 | 0.7784 | 0.7426 | 0.7355 | 0.5814 | 0.7744 | 0.7389 | 0.7718 |
| FCM_S | 0.7879 | 0.7337 | 0.7817 | 0.7188 | 0.7438 | 0.8244 | 0.7681 | 0.7762 | 0.7819 | 0.7999 |
| FLICM | 0.7736 | 0.7568 | 0.7850 | 0.7505 | 0.8052 | 0.5664 | 0.7800 | 0.7935 | 0.7751 | 0.7961 |
| IPFCM | 0.7839 | 0.7880 | 0.7634 | 0.8069 | 0.7789 | 0.7716 | 0.6685 | 0.7958 | 0.7640 | 0.7917 |
| AIPFCM | 0.7617 | 0.7794 | 0.7470 | 0.7904 | 0.7625 | 0.7474 | 0.7248 | 0.7884 | 0.7566 | 0.7964 |
| APFCMS | 0.7763 | 0.7895 | 0.7724 | 0.7888 | 0.7826 | 0.7590 | 0.7498 | 0.8024 | 0.7836 | 0.7948 |
| IFCMSNI | 0.8077 | 0.8009 | 0.7826 | 0.7923 | 0.8310 | 0.8217 | 0.7513 | 0.8115 | 0.7888 | 0.7954 |
| KWFLICM | 0.7247 | 0.8008 | 0.7503 | 0.7936 | 0.8343 | 0.8085 | 0.7680 | 0.8012 | 0.7316 | 0.7958 |
| NDFCM | 0.8011 | 0.7955 | 0.7570 | 0.7837 | 0.7732 | 0.8269 | 0.7689 | 0.8035 | 0.7824 | 0.7897 |
| KGFCM | 0.8061 | 0.7950 | 0.7845 | 0.8257 | 0.7856 | 0.8239 | 0.7838 | 0.8041 | 0.7855 | 0.7970 |
| IIFCM | 0.7974 | 0.7980 | 0.7798 | 0.7952 | 0.8103 | 0.8085 | 0.7338 | 0.8068 | 0.7821 | 0.7910 |
| FRFCM | 0.7896 | 0.7847 | 0.7655 | 0.8056 | 0.8166 | 0.7820 | 0.7702 | 0.8044 | 0.7783 | 0.7811 |
| DSFCMN | 0.7690 | 0.7279 | 0.7596 | 0.7276 | 0.7592 | 0.7859 | 0.7364 | 0.7385 | 0.7721 | 0.7119 |
| WRFCM | 0.8031 | 0.7935 | 0.7791 | 0.7854 | 0.8316 | 0.8277 | 0.7704 | 0.8075 | 0.7840 | 0.7915 |
| KPFCMSNI | 0.8020 | 0.7931 | 0.7690 | 0.7868 | 0.8098 | 0.8159 | 0.7540 | 0.8035 | 0.7766 | 0.7883 |

**Table 14** Dice score for GM IBSR dataset

| Methods\Cases | 110_03 | 111_02 | 112_02 | 11_03 | 12_03 | 13_03 | 16_03 | 202_03 | 205_03 | 1_24 |
|---|---|---|---|---|---|---|---|---|---|---|
| FCM | 0.6143 | 0.6380 | 0.5763 | 0.6611 | 0.5893 | 0.5720 | 0.5204 | 0.6024 | 0.5854 | 0.6289 |
| FC_PFS | 0.6230 | 0.6488 | 0.5816 | 0.6679 | 0.5956 | 0.5774 | 0.5297 | 0.6101 | 0.5938 | 0.6368 |
| RPFCM | 0.6146 | 0.6381 | 0.5763 | 0.6615 | 0.5897 | 0.5732 | 0.5189 | 0.6039 | 0.5854 | 0.6292 |
| FCM_S | 0.7218 | 0.7455 | 0.8027 | 0.6280 | 0.7361 | 0.6284 | 0.5306 | 0.5985 | 0.4099 | 0.4278 |
| FLICM | 0.8417 | 0.4025 | 0.5568 | 0.6423 | 0.5463 | 0.3427 | 0.4650 | 0.4894 | 0.5877 | 0.7680 |
| IPFCM | 0.6540 | 0.6694 | 0.6062 | 0.6969 | 0.6233 | 0.6059 | 0.5515 | 0.6239 | 0.6043 | 0.6575 |
| AIPFCM | 0.6240 | 0.6572 | 0.5872 | 0.6734 | 0.6071 | 0.5835 | 0.5839 | 0.6176 | 0.5985 | 0.6718 |
| APFCMS | 0.6799 | 0.6998 | 0.6783 | 0.7125 | 0.6473 | 0.6111 | 0.6889 | 0.6683 | 0.6838 | 0.7070 |
| IFCMSNI | 0.7216 | 0.7292 | 0.6837 | 0.7490 | 0.7201 | 0.7104 | 0.6020 | 0.6988 | 0.6675 | 0.7124 |
| KWFLICM | 0.5385 | 0.7062 | 0.5700 | 0.7399 | 0.7338 | 0.6755 | 0.5965 | 0.5440 | 0.5514 | 0.7182 |
| NDFCM | 0.7694 | 0.7442 | 0.6828 | 0.7705 | 0.7050 | 0.7954 | 0.7050 | 0.7044 | 0.7000 | 0.7199 |
| KGFCM | 0.7304 | 0.6264 | 0.6729 | 0.6415 | 0.5787 | 0.6459 | 0.6817 | 0.5870 | 0.6742 | 0.6570 |
| IIFCM | 0.7284 | 0.7390 | 0.6875 | 0.7768 | 0.7218 | 0.7593 | 0.6887 | 0.7060 | 0.7102 | 0.7009 |
| FRFCM | 0.6909 | 0.7079 | 0.6852 | 0.7511 | 0.7302 | 0.6783 | 0.6953 | 0.6796 | 0.6852 | 0.7134 |
| DSFCMN | 0.7479 | 0.6935 | 0.6771 | 0.6135 | 0.6844 | 0.6727 | 0.6367 | 0.6849 | 0.7288 | 0.6791 |
| WRFCM | 0.7006 | 0.6883 | 0.7177 | 0.7194 | 0.7114 | 0.6821 | 0.6589 | 0.6499 | 0.6656 | 0.6911 |
| KPFCMSNI | 0.7471 | 0.7533 | 0.7316 | 0.7566 | 0.7832 | 0.8007 | 0.6883 | 0.7275 | 0.7058 | 0.7307 |

hood membership values to dampen the effect of noise. In this way, the proposed clustering method avoids smoothing and preserves fine details in the image segmentation process. To capture the nonlinear structure present in image kernel distance measure is used in the proposed optimization problem. The experiments are carried out on a synthetic image dataset and two publicly available brain MRI datasets. The comparison with the state-of-the-art method shows that the proposed picture fuzzy clustering method provides better segmentation performance in terms of average segmentation accuracy and Dice score. Statistical test has been conducted, which advocates the superior performance of the proposed method over other related methods. Selection of spatial regularization parameter is crucial for getting accurate segmentation result.

**Table 15** Average Friedman ranking of algorithms

| Algorithms | Ranking | Algorithms | Ranking |
|---|---|---|---|
| KPFCMSNI | 2.08 | AIPFCM | 9.92 |
| IFCMSNI | 4.39 | NDFCM | 10.24 |
| IIFCM | 5.45 | DSFCMN | 10.27 |
| APFCMS | 7.76 | FRFCM | 10.80 |
| KGFCM | 8.16 | FCM_S | 11.05 |
| KWFLICM | 8.61 | RPFCM | 11.31 |
| WRFCM | 8.86 | FCM | 11.39 |
| IPFCM | 9.74 | FLICM | 13.19 |
| FC_PFS | 9.77 | | |

**Table 16** Adjusted $p$-value (Friedman)

| Algorithms | Unadjusted $p$ | $p_Holm$ | Algorithms | Unadjusted $p$ | $p_Holm$ |
|---|---|---|---|---|---|
| FLICM | 3.04E-21 | 4.86E-20 | FC_PFS | 5.78E-11 | 4.62E-10 |
| FCM | 2.18E-15 | 3.27E-14 | IPFCM | 6.74E-11 | 4.72E-10 |
| RPFCM | 3.80E-15 | 5.31E-14 | WRFCM | 7.55E-09 | 4.53E-08 |
| FCM_S | 2.13E-14 | 2.76E-13 | KWFLICM | 2.71E-08 | 1.35E-07 |
| FRFCM | 1.14E-13 | 1.36E-12 | KGFCM | 2.22E-07 | 8.89E-07 |
| DSFCMN | 3.05E-12 | 3.36E-11 | APFCMS | 1.34E-06 | 4.01E-06 |
| NDFCM | 3.60E-12 | 3.60E-11 | IIFCM | 0.004 | 0.008 |
| AIPFCM | 2.46E-11 | 2.21E-10 | IFCMSNI | 0.049 | 0.049 |

It will be interesting future direction to develop method, which adaptively select the spatial regularization parameter. Also, other extension of fuzzy sets for developing clustering method for image segmentation can be investigated in future.

**Author Contributions** There is equal contributions in this research from all the authors of this article.

**Funding** We have received no funding for this research.

**Data availability** Image data utilized for analysis of the proposed approach is publicly available.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

## References

Ahmed MN, Yamany SM, Mohamed N, Farag AA, Moriarty T (2002) A modified fuzzy $c$-means algorithm for bias field estimation and segmentation of MRI data. IEEE Trans Med Imag 21(3):193–199

Alipour S, Shanbehzadeh J (2014) Fast automatic medical image segmentation based on spatial kernel fuzzy $c$-means on level set method. Mach Vis Appl 25(6):1469–1488

Atanassov KT (1986) Intuitionistic fuzzy sets. Fuzzy Sets Syst 20(1):87–96

Balafar MA, Ramli AR, Saripan MI, Mashohor S (2010) Review of brain MRI image segmentation methods. Artif Intell Rev 33(3):261–274

Benaichouche A, Oulhadj H, Siarry P (2013) Improved spatial fuzzy $c$-means clustering for image segmentation using PSO initialization, mahalanobis distance and post-segmentation correction. Digit Sign Process 23(5):1390–1400

Berry MW, Castellanos M (2004) Survey of text mining. Comput Rev 45(9):548

Bezdek JC (1981) Objective function clustering. In: Pattern recognition with fuzzy objective function algorithms. Springer, pp 43–93

Bezdek JC (1981) Pattern recognition with fuzzy function algorithms

Cai W, Chen S, Zhang D (2007) Fast and robust fuzzy $c$-means clustering algorithms incorporating local information for image segmentation. Pattern Recognit 40(3):825–838

Chaira T (2011) A novel intuitionistic fuzzy $c$ means clustering algorithm and its application to medical images. Appl Soft Comput 11(2):1711–1717

Chen S, Zhang D (2004) Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. Syst Man Cybern Part B: Cybern IEEE Trans 34(4):1907–1916

Chen X, Nguyen BP, Chui CK, Ong SH (2016) Automated brain tumor segmentation using kernel dictionary learning and superpixel-level features. In: Systems, man, and cybernetics (SMC), 2016 IEEE international conference on, pp 002547–002552. IEEE

Chuang KS, Tzeng HL, Chen S, Wu J, Chen TJ (2006) Fuzzy $c$-means clustering with spatial information for image segmentation. Comput Med Imag Gr 30(1):9–15

Cocosco CA, Kollokian V, Kwan RKS, Pike GB, Evans AC (1997) Brainweb: online interface to a 3d MRI simulated brain database. In: NeuroImage. Citeseer

Cover TM (1965) Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. IEEEE Trans Electron Comput EC–14(3):326–334. https://doi.org/10.1109/PGEC.1965.264137

Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press

Cuong BC, Kreinovich V (2013) Picture fuzzy sets-a new concept for computational intelligence problems. In: 2013 Third world congress on information and communication technologies (WICT 2013), pp 1–6. IEEE (2013)

Derrac J, García S, Molina D, Herrera F (2011) A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. Swarm Evol Comput 1(1):3–18

Friedman M (1937) The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J Am Stat Assoc 32(200):675–701

Gadekallu TR, Alazab M, Kaluri R, Maddikunta PKR, Bhattacharya S, Lakshmanna K et al (2021a) Hand gesture classification using a novel CNN-crow search algorithm. Complex Intell Syst 7(4):1855–1868

Gadekallu TR, Rajput DS, Reddy M, Lakshmanna K, Bhattacharya S, Singh S, Jolfaei A, Alazab M (2021b) A novel PCA-whale optimization-based deep neural network model for classification of tomato plant diseases using GPU. J Real-Time Image Process 18(4):1383–1396

Gokul Y, Gopal M (2017) An authoritative method using fuzzy logic to evaluate maintainability index and utilizability of software. Adv Model Anal B 60(3):566–580

Gong M, Liang Y, Shi J, Ma W, Ma J (2012) Fuzzy *c*-means clustering with local information and kernel metric for image segmentation. IEEE Trans Image Process 22(2):573–584

Guo FF, Wang XX, Shen J (2016) Adaptive fuzzy *c*-means algorithm based on local noise detecting for image segmentation. IET Image Process 10(4):272–279

Held K, Kops ER, Krause BJ, Wells WM, Kikinis R, Muller-Gartner HW (1997) Markov random field segmentation of brain MR images. IEEE Trans Med Imag 16(6):878–886

Huang CW, Lin KP, Wu MC, Hung KC, Liu GS, Jen CH (2015) Intuitionistic fuzzy *c*-means clustering algorithm with neighborhood attraction in segmenting medical image. Soft Comput 19(2):459–470

Iakovidis D, Pelekis N, Kotsifakos E, Kopanakis I (2008) Intuitionistic fuzzy clustering with applications in computer vision. In: Advanced concepts for intelligent vision systems. Springer, pp 764–774

Iman RL, Davenport JM (1980) Approximations of the critical region of the Fbietkan statistic. Commun Stat-Theor Methods 9(6):571–595

J Mercer B (1909) Xvi. functions of positive and negative type, and their connection the theory of integral equations. Phil Trans R Soc Lond A 209(441–458):415–446

Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. ACM Comput Surv (CSUR) 31(3):264–323

Ji ZX, Sun QS, Xia DS (2014) A framework with modified fast FCM for brain MR images segmentation (retraction of vol 44, pg 999, 2011). Patt Recogn 47(12):3979–3979

Kannan S, Devi R, Ramathilagam S, Takezawa K (2013) Effective FCM noise clustering algorithms in medical images. Comput Biol Med 43(2):73–83

Krinidis S, Chatzis V (2010) A robust fuzzy local information *c*-means clustering algorithm. IEEE Trans Image Process 19(5):1328–1337

Kumar D, Agrawal R, Verma H (2019a) Kernel intuitionistic fuzzy entropy clustering for MRI image segmentation. Soft Comput: 1–24

Kumar D, Agrawal RK, Kirar JS (2019b) Intuitionistic fuzzy clustering method with spatial information for mri image segmentation. In: 2019 IEEE international conference on fuzzy systems (FUZZ-IEEE), pp 1–7. IEEE

Kumar D, Verma H, Mehra A, Agrawal R (2019c) A modified intuitionistic fuzzy *c*-means clustering approach to segment human brain MRI image. Multimed Tools Appl 78(10):12663–12687

Lei T, Jia X, Zhang Y, He L, Meng H, Nandi AK (2018) Significantly fast and robust fuzzy *c*-means clustering algorithm based on morphological reconstruction and membership filtering. IEEE Trans Fuzzy Syst 26(5):3027–3041

Li C, Huang R, Ding Z, Gatenby JC, Metaxas DN, Gore JC (2011) A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI. IEEE Trans Image Process 20(7):2007–2016

Lloyd S (1982) Least squares quantization in PCM. IEEE Trans Inf Theor 28(2):129–137

Olabarriaga SD, Smeulders AW (2001) Interaction in the segmentation of medical images: a survey. Med Image Anal 5(2):127–142

Pham DL, Xu C, Prince JL (2000) Current methods in medical image segmentation. Ann Rev Biomed Eng 2(1):315–337

Qiu C, Xiao J, Yu L, Han L, Iqbal MN (2013) A modified interval type-2 fuzzy c-means algorithm with application in MR image segmentation. Patt Recogn Lett 34(12):1329–1338

Reddick WE, Glass JO, Cook EN, Elkin TD, Deaton RJ (1997) Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks. IEEE Trans Med Imag 16(6):911–918

Rohlfing T, Brandt R, Menzel R, Maurer CR (2004) Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. NeuroImage 21(4):1428–1442

Roth V, Steinhage V (2000) Nonlinear discriminant analysis using kernel functions. In: Advances in neural information processing systems, pp 568–574

Rui Y, Huang TS, Chang SF (1999) Image retrieval: current techniques, promising directions, and open issues. J Vis Commun Image Represent 10(1):39–62

Sato M, Lakare S, Wan M, Kaufman A, Nakajima M (2000) A gradient magnitude based region growing algorithm for accurate segmentation. In: Image processing, 2000. Proceedings. 2000 international conference on, vol. 3, pp 448–451. IEEE (2000)

Schölkopf B, Smola A, Müller KR (1998) Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput 10(5):1299–1319

Son LH (2015) Dpfcm: A novel distributed picture fuzzy clustering method on picture fuzzy sets. Expert Syst Appl: Int J 42(1):51–66

Szilagyi L, Benyo Z, Szilágyi SM, Adam H (2003) Mr brain image segmentation using an enhanced fuzzy c-means algorithm. In: Engineering in medicine and biology society, 2003. Proceedings of the 25th annual international conference of the IEEE, vol. 1, pp 724–726. IEEE

Thong PH et al (2016) Picture fuzzy clustering: a new computational intelligence method. Soft Comput 20(9):3549–3562

Vapnik V (2013) The nature of statistical learning theory. Springer science & Business Media

Verma H, Agrawal R, Sharan A (2016) An improved intuitionistic fuzzy *c*-means clustering algorithm incorporating local information for brain image segmentation. Appl Soft Comput 46:543–557

Vovk U, Pernus F, Likar B (2007) A review of methods for correction of intensity inhomogeneity in MRI. IEEE Trans Med Imag 26(3):405–421

Wang C, Pedrycz W, Li Z, Zhou M (2020) Residual-driven fuzzy *c*-means clustering for image segmentation. IEEE/CAA J Autom Sin 8(4):876–889

Wang L, Chen Y, Pan X, Hong X, Xia D (2010) Level set segmentation of brain magnetic resonance images based on local gaussian distribution fitting energy. J Neurosci Methods 188(2):316–325

Wang Z, Song Q, Soh YC, Sim K (2013) An adaptive spatial information-theoretic fuzzy clustering algorithm for image segmentation. Comput Vis Image Underst 117(10):1412–1420

Wu C, Chen Y (2020) Adaptive entropy weighted picture fuzzy clustering algorithm with spatial information for image segmentation. Appl Soft Comput 86:105888

Wu C, Wu Q (2017) A robust image segmentation algorithm based on modified picture fuzzy clustering method on picture fuzzy sets. J Xi'an Univ Posts Telecommun 22(5):37–43

Xu Z, Wu J (2010) Intuitionistic fuzzy *c*-means clustering algorithms. J Syst Eng Electron 21(4):580–590

Zadeh LA (1965) Fuzzy sets. Inf Control 8(3):338–353

Zang W, Zhang W, Zhang W, Liu X (2017) A kernel-based intuitionistic fuzzy *c*-means clustering using a DNA genetic algorithm for magnetic resonance image segmentation. Entropy 19(11):578

Zhang Y, Bai X, Fan R, Wang Z (2018) Deviation-sparse fuzzy *c*-means with neighbor information constraint. IEEE Trans Fuzzy Syst 27(1):185–199

Zhao F, Jiao L, Liu H (2013) Kernel generalized fuzzy *c*-means clustering with spatial information for image segmentation. Digit Sign Process 23(1):184–199

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

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



# Memory Recurrent Elman Neural Network-Based Identification of Time-Delayed Nonlinear Dynamical System

Rajesh Kumar

***Abstract*—In this article, an attempt has been made to propose an improved version of the classical Elman neural network (ENN) and its application is presented to identify the unknown dynamics of time-delayed nonlinear plants. The model proposed, known as memory recurrent ENN (MRENN), consists of an additional number of weighted self-feedback loops, an extra output context layer, and weighted connections of input signals (through adjustable weights) to the output-layer neuron. The model is proposed to be given only two signals as its inputs: plant unit-delayed value and the present value of the externally applied signal irrespective of the actual order of the plant (which most of the time may not be known). To guarantee stability, the parameters of the MRENN model are updated using the equations that are obtained by applying Lyapunov-stability criteria. Furthermore, the recursive learning rate scheme is constructed for speeding up the learning process. From the simulation results, MRENN model appears to have produced comparably superior outcomes when the performance of the suggested model is compared with that of other well-known models.**

***Index Terms*—Adaptive learning rate, Elman neural network (ENN), Lyapunov stability analysis, nonlinear system identification, recurrent neural network (RNN).**

## I. Introduction

FOR A stable controller design, there is a requirement to determine the dynamics of the system/plant. This is quite a challenging task as most plants are described by nonlinear equations. Many of the times accurate information regarding the dynamics is not available [1]. The problem of system identification has been dealt using various types of intelligent tools, such as artificial neural network (ANN) models [2], [3], fuzzy-logic-based systems [3], [4], etc. ANN models mimic the behavior of the human brain and are found to be very effective in solving system identification and control problems [5], [35]. They have remarkable features, such as adaptive learning, parallel nonlinear processing, function approximation, and generalization [6], [7]. To use any ANN model for performing the nonlinear system identification task one is only required to observe the input–output data from the system instead of figuring out the actual dynamics linking these data values and is responsible for controlling its behavior [8]. With the help of input–output data and the learning algorithm, the ANN model parameters undergoes adjustment and an empirical model is established that is quite simple as compared to the actual differential equations (which are unknown) describing the given system. Feedforward neural network (FFNN) and recurrent neural network (RNN) models are the two categories to which any given ANN model belongs. The former does not include any kind of feedback loops (in its structure) but only has feedforward connections whereas the latter includes both feedforward as well as feedback loops/connections. In the case of the FFNN model, there is only one direction in which the signal flows (which starts from the input layer and then through the hidden layer(s) reach out the output layer) [10], [11]. FFNN models, which include multilayered perceptron (MLP) [12], RBFN [13], [37], and functional link neural network (FLNN) [14] have been extensively used in solving various engineering problems, such as pattern recognition [15], classification [16], etc.

In most identification problems, the output of the dynamic plant can be computed by using the past values of the plant and the current and previous values of the external input. Recognizing such temporal patterns requires the ANN model to have some kind of memory in its structure [17]. But most of the ANN models are of a static type and, hence, are suitable only for memoryless transformations (since their output depends upon the current values of the inputs). To apply them for identifying the dynamic plants, a sufficient number of past values [using the time-delayed lines (TDLs)] of both the plant and external input $r(k)$ are to be applied as inputs signals [18], [19], [36]. Then, the ANN model can be expected to capture the memoryless transformations. This, however, requires the availability of information regarding the exact order of the given plant (or on the upper bound on the order). But unfortunately, most of the time such information is not available and even if it is available, the approximation of the input–output data relationship using the ANN model is not found to produce versatile dynamic models. On the other hand, the proposed memory recurrent Elman neural network (MRENN) model consists of several tunable weighted loops that provide the unit-delayed outputs of various neurons (to which they are connected) and, hence, contribute to establishing the required memory feature into the structure. The MRENN model can recognize the temporal patterns existing between

Manuscript received 24 March 2022; accepted 22 June 2022. This article was recommended by Associate Editor X. Wang.

The author is with the Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India (e-mail: rajeshmahindru23@gmail.com; rajeshkumar@dtu.ac.in).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSMC.2022.3186610.

Digital Object Identifier 10.1109/TSMC.2022.3186610

2168-2216 © 2022 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 July 22,2022 at 10:20:10 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.



the input–output data (obtained from the plant). Furthermore, it does not require any information regarding the exact order (of the given plant) as it is provided with only two input signals (viz., unit-delayed output of the plant and the present value of the input). This eliminates the necessity of knowing the order of the plant and, hence, it can be applied to identify any complex nonlinear dynamics.

### A. Related Work

A number of successful Elman neural network (ENN) and other RNN model-based control and identification schemes for nonlinear dynamical systems have been implemented by various researchers [24]–[26], [34], etc. In particular, Ge *et al.* [27] have proposed an ANN model which involves the time delay and a Lyapunov-stability-based learning mechanism. The proposed model has shown good performance in identifying and controlling the various nonlinear systems. In [28], a backpropagation-based RNN is implemented for designing the stable model-predictive control scheme. The proposed configuration is tested and compared with other methods and has performed better. In [29], a two-layered neural-network-based identification model (coupled with a feedback term) is presented. The learning algorithm is developed using the Lyapunov-stability principles. The proposed scheme is successfully applied to identify the unknown dynamics of two complex nonlinear systems (under the effect of the disturbances). Bonassi and Scattolini [30] have implemented an internal model control strategy based on two gated recurrent units (GRUs). One GRU is used to implement the identifier for approximating the unknown dynamic function of the plant and the other is used as a controller. The proposed framework is tested on the Quadruple Tank benchmark system.

Yao *et al.* [31] have created an identification model for nonlinear dynamical systems using an adaptive echo-state neural network. The Lyapunov-stability principles are used to create the parameter update equations. By using it to forecast the Mackey–Glass chaotic time series, the model's performance is evaluated. A unique self-organizing interval Type-2 fuzzy-neural network is created in [32] to address the modeling and prediction issue. The suggested approach involves automatically adjusting the model's parameters and structure. Additionally, it is evaluated against the current approaches and proven to perform better.

### B. Contributions of This Article

The contributions of the present research are as follows:

1) In addition to the weighted linear feed through of input signals (to the output layer) and an additional self-feedback adjustable weighted loops, the proposed MRENN model has an additional context layer compared to the standard ENN structure. The MRENN model's ability of handling any dynamic systems has been improved by these new enhancements.
2) In most of the neural network models, there is a requirement of knowing the order of the plant so that the desired input signals can be applied to the model. This information, however, is not required by the MRENN model since it is only provided with two inputs ($r(k)$ and $y_d(k-1)$) and is found to successfully identify the unknown dynamics of the given system.
3) Many researchers have used the BP method to develop the parameter update equation for the neural network model. Though the BP method is quite simple, it usually suffers from the local minima problem. In the proposed adaptive BP method, the Lyapunov stability criteria are used to develop the update equation. It involves the creation of the energy surface having only single global minimum point. This approach also ensures the overall stability of the system.
4) Most of the training methods available in the literature involve a constant value of the learning rate. If its value is set close to zero, then the training procedure will be quite slow and if its higher value is selected then the training procedure may become unstable. Thus, in this article, an attempt has been made to develop an adaptive learning rate scheme that updates the learning rate value (in each iteration).
5) The working of the MRENN is also compared with ENN, FFNN, and RBFN models. These models are applied to approximate the unknown mathematical equations connecting the input–output data concerned with the two considered benchmark systems. The simulation results show that, in comparison to the findings obtained with the other three models, the MRENN model has produced more accurate results.

## II. Problem Description

The following general difference equation can be used to describe the input–output data relationship of any dynamic plant:

$$y_d(k) = f\big[y_d(k), y_d(k-1), \ldots, y_d(k-u) \\ r(k), r(k-1), \ldots, r(k-v)\big] \tag{1}$$

where $y_d(k)$ and $r(k)$ are plant output current value and the externally applied bounded input signal value, respectively. Furthermore, $u$ and $v$ denote the time-lagged values of these signals. The unknown function $f$ is assumed to be nonlinear but differentiable. To compute the MRENN model's output, only two inputs, viz., $y_d(k-1)$ and $r(k)$ are used. Thus

$$y_{mr}(k) = \hat{f}\big[y_d(k-1), r(k)\big] \tag{2}$$

where $y_{mr}(k)$ represents the current output of the MRENN model and $\hat{f}$ represents the dynamics of the MRENN model. The above identification model is of series–parallel type and the objective of training is to achieve $\hat{f} \approx f$ which means $\|y_d(k) - y_{mr}(k)\| \leq \xi$ for some desired $\xi \to 0$ with a suitably defined norm (denoted by $\|.\|$) on the output space.

## III. Mathematical Formulation of Memory Recurrent Elman Neural Network

Fig. 1 shows the proposed MRENN model. A total of five layers are depicted in the image, including input to the hidden layer, input to the output layer, two context layers, and weighted direct linkages between the signals of the input-layer and output-layer neurons. The externally applied signals

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2022 at 10:20:10 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.



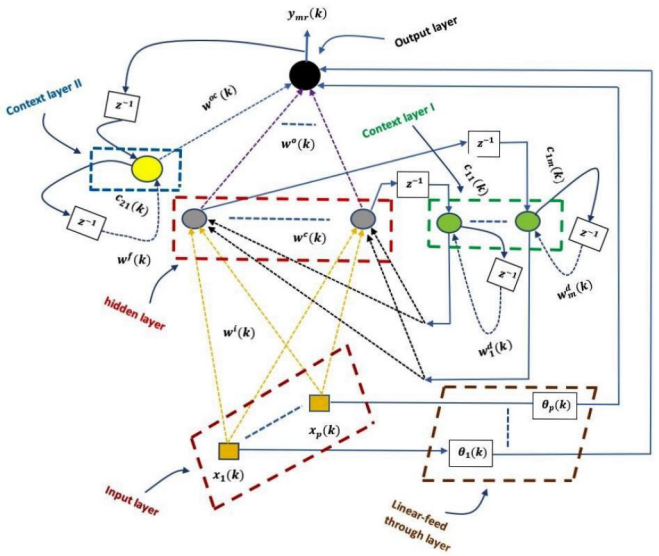



Fig. 1. Model framework of the proposed MRENN model.

[denoted by $x_1(k), x_2(k), \ldots, x_p(k)$] are the part of the input layer that establishes the connections to the hidden-layer neurons through tunable weights (denoted by $w^i(k)$ vector). The resulting outputs given by each of the hidden-layer neurons are having weighted connections (denoted by $w^o(k)$ vector) with the output-layer neuron. Additionally, we have a corresponding context node in the first context layer for each of the hidden-layer neurons. The output of each context node is connected to each of the hidden-layer neurons using context-layer weights, denoted by the $w^c(k)$ vector, and each context node stores the unit delayed output value of the corresponding hidden-layer neuron (using the TDL $z^{-1}$). The second context layer consists of the context node that receives a unit delayed value of the MRENN model output as its input and is having a weighted connection [denoted by $w^{oc}(k)$] with the MRENN output-layer neuron. Furthermore, each of the context nodes receives a unit delayed output value of its own via adjustable weights.

Finally, the fifth layer consists of adjustable parameters [denoted by $\theta_j(k)$] through which any $j$th input-layer signal is connected to the output-layer neuron. The structure will reduce to the traditional ENN model if the second context layer, linear feed-through layer, and self-feedback loops on each of the context nodes of the context layers are eliminated. Any $j$th hidden-layer neuron's induced field value is calculated as follows:

$$v_j(k) = \left\{\sum_{u=1}^{p} x_u(k) w_{uj}^i(k)\right\} + \left\{\sum_{q=1}^{m} c_{1q}(k) w_q^c(k)\right\} - w_j^h(k) \tag{3}$$

where $c_{1q}(k)$ denotes the output value of any $q$th context node of first context layer and $w_j^h(k)$ denotes the bias weight connected to the $j$th (having input $= -1$) hidden-layer neuron. The output of $c_{1q}(k)$ can be obtained as follows:

$$c_{1q}(k) = s_q(k-1) + c_{1q}(k-1) w_q^d(k) \tag{4}$$

and the output of any $j$th hidden-layer neuron is given by

$$s_j(k) = \tanh(v_j(k)) \tag{5}$$

where tanh(.) represents the tangent hyperbolic activation function (having a range of $[-1, 1]$). The output-layer neuron uses a linear activation function to prevent normalization of the output data. Finally, the MRENN model's output is calculated as follows:

$$y_{mr}(k) = \left\{\sum_{i=1}^{m} s_i(k) w_i^o(k)\right\} + \left\{\sum_{i=1}^{p} x_i(k)\theta_i(k)\right\} + c_{21}(k) w^{oc}(k) - w_o^h(k) \tag{6}$$

where the output expression for the second context-layer node can be written as follows:

$$c_{21}(k) = y_{mr}(k-1) + w^f(k) c_{21}(k-1). \tag{7}$$

## IV. Lyapunov Stability-Based Development of Adaptive BP Algorithm

The parameters of the proposed MRENN model are adjusted by developing an update equation which is based on the Lyapunov stability criteria. First, any positive-definite Lyapunov function $v_L(x)$ is selected which satisfy the following condition:

$$v_L(x) > 0 \quad \text{for all } x \text{ except } x = 0. \tag{8}$$

Now, if the following condition is met then the learning procedure is said to be asymptotic stable. That is

$$\frac{dv_L(x)}{dt} < 0 \quad \text{for all } x. \tag{9}$$

According to the gradient-descent principle, the general parameter update equation of any $j$th parameter of the MRENN model can be written as

$$\phi_j(k) = \phi_j(k-1) - \eta_i(k) \Delta\phi_j(k) \tag{10}$$

where $\phi_j(k)$ and $\phi_j(k-1)$ denotes the updated and previous instant values, respectively, and $\eta_i(k)$ represents the adaptive learning rate. Let us consider the mean square error (MSE) as a cost function, and the goal of the learning process is to adjust the model's parameters in order to reduce the cost function. The following defines the selected cost function:

$$E_i(k) = \frac{1}{2} e_i^2(k). \tag{11}$$

The gradient-descent method states that the change in $\phi_j(k)$ can be written as follows:

$$\Delta\phi_j(k) = \frac{\partial E_i(k)}{\partial \phi_j(k)} \tag{12}$$

or

$$\Delta\phi_j(k) = \frac{\partial E_i(k)}{\partial e_i(k)} \times \frac{\partial e_i(k)}{\partial \phi_j(k)}. \tag{13}$$

Since the above equation can trap/stuck into the minima problem, so we switched to the Lyapunov stability method

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2022 at 10:20:10 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.



and used it to develop the parameter update equation. The positive-definite Lyapunov function is given as follows:

$$v_L(k) = \frac{1}{2}\Big[\alpha e_i^2(k) + \beta\phi_j^2(k)\Big] \tag{14}$$

where $\alpha > 0$ and $\beta > 0$. In the above equation, $v_L(k)$ is a function of identification error $e_i(k)$ and any $j$th MRENN model's weight $\phi_j(k)$. It can be observed that $v_L(k)$ value will be zero if both of its arguments $e_i(k)$ and $\phi_j(k)$ values are zero, and $v_L(k)$ value is nonzero and positive if any of its argument value is nonzero. The objective of defining $v_L(k)$ in terms of $e_i(k)$ and $\phi_j(k)$ is to find out an update equation for $\phi_j(k)$ so that $e_i(k)$ approach toward zero as the training progresses) and the following second Lyapunov-stability condition is satisfied:

$$\Delta v_L(k) = v_L(k) - v_L(k-1) < 0. \tag{15}$$

Now, on substituting the expression of $v_L(k)$ and $v_L(k)$ in the above equation will lead to the following equation:

$$\Delta v_L(k) = \frac{1}{2}\Big\{\Big[\alpha e_i^2(k) + \beta\phi_j^2(k)\Big] - \Big[\alpha e_i^2(k-1) + \beta\phi_j^2(k-1)\Big]\Big\}. \tag{16}$$

After rearranging the similar terms in the above equation, we will get

$$\Delta v_L(k) = \frac{1}{2}\Big\{\alpha\Big[e_i^2(k) - e_i^2(k-1)\Big] + \beta\Big[\phi_j^2(k) - \phi_j^2(k-1)\Big]\Big\} \tag{17}$$

or

$$\Delta v_L(k) = \frac{\alpha}{2}[e_i(k) + e_i(k-1)][e_i(k) - e_i(k-1)] + \frac{\beta}{2}[\phi_j(k) + \phi_j(k-1)][\phi_j(k) - \phi_j(k-1)]. \tag{18}$$

Let us define $\Delta e_i(k) = e_i(k) - e_i(k-1)$ and $\Delta\phi_j(k) = \phi_j(k) - \phi_j(k-1)$ then on substituting them in the above equation, we will get

$$\Delta v_L(k) = \frac{\alpha}{2}\{\Delta e_i(k)[\Delta e_i(k) + 2e_i(k)]\} + \frac{\beta}{2}\Big\{\Delta\phi_j(k)\Big[\Delta\phi_j(k) + 2\phi_j(k)\Big]\Big\} \tag{19}$$

or

$$\Delta v_L(k) = \frac{\beta}{2}\big(\Delta\phi_j(k)\big)^2\left[1 + \frac{\alpha}{\beta}\left[\frac{\Delta e_i(k)}{\Delta\phi_j(k)}\right]^2\right] + \beta\Delta\phi_j(k)\left[\phi_j(k) + \frac{\alpha}{\beta}e_i(k)\left[\frac{\Delta e_i(k)}{\Delta\phi_j(k)}\right]\right]. \tag{20}$$

If the change in parameter value is very small, then the above equation can be rewritten as follows:

$$\Delta v_L(k) = \frac{\beta}{2}\big(\Delta\phi_j(k)\big)^2\left[1 + \frac{\alpha}{\beta}\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]^2\right] + \beta\Delta\phi_j(k)\left[\phi_j(k) + \frac{\alpha}{\beta}e_i(k)\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]\right]. \tag{21}$$

*Theorem:* For the considered positive-definite Lyapunov function, $v_L(k) = (1/2)[\alpha e_i^2(k) + \beta\phi_j^2(k)] > 0$, condition $\Delta v_L(k) < 0$ is ensured if any $j$th parameter is updated as follows:

$$\Delta\phi_j(k) = \frac{\left\{\phi_j(k) + \frac{\alpha}{\beta}e_i(k)\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]\right\}}{1 + \frac{\alpha}{\beta}\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]^2}. \tag{22}$$

*Proof of Stability:* Let (21) be equated to $-\frac{1}{2}\lambda$ where $\lambda > 0$ for ensuring $\Delta v_L(k) < 0$. Thus

$$\Delta v_L(k) = \frac{\beta}{2}\big(\Delta\phi_j(k)\big)^2\left[1 + \frac{\alpha}{\beta}\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]^2\right] + \beta\Delta\phi_j(k)\left[\phi_j(k) + \frac{\alpha}{\beta}e_i(k)\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]\right] = -\frac{1}{2}\lambda \tag{23}$$

or

$$\Delta v_L(k) = \frac{\beta}{2}\big(\Delta\phi_j(k)\big)^2\left[1 + \frac{\alpha}{\beta}\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]^2\right] + \beta\Delta\phi_j(k)\left[\phi_j(k) + \frac{\alpha}{\beta}e_i(k)\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]\right] + \frac{1}{2}\lambda = 0. \tag{24}$$

Any generalized linear second-order equation can be written as follows:

$$az^2 + bz + c = 0. \tag{25}$$

The solution of the above quadratic equation can be obtained as follows:

$$z_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad z_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}. \tag{26}$$

On comparing (25) with (24), we can infer that $\Delta\phi_j(k) = z$

$$a = \frac{\beta}{2}\left[1 + \frac{\alpha}{\beta}\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]^2\right]$$

$$b = \beta\left[\phi_j(k) + \frac{\alpha}{\beta}e_i(k)\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]\right], \text{ and } c = \frac{1}{2}\lambda.$$

Now, there are three possibilities regarding the discriminant value: $b^2 - 4ac > 0$ or $b^2 - 4ac = 0$ or $b^2 - 4ac < 0$. If $b^2 - 4ac > 0$ is true, then its expression will be (on substituting expressions of $a$, $b$, and $c$)

$$\sqrt{\beta^2\left[\phi_j(k) + \frac{\alpha}{\beta}e_i(k)\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]\right]^2 - \beta\left[1 + \frac{\alpha}{\beta}\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]^2\right]\lambda} > 0. \tag{27}$$

Now, squaring both sides and rearranging the terms will give

$$\beta^2\left[\phi_j(k) + \tfrac{\alpha}{\beta}e_i(k)\left[\tfrac{\partial e_i(k)}{\partial\phi_j(k)}\right]\right]^2 > \beta\left[1 + \tfrac{\alpha}{\beta}\left[\tfrac{\partial e_i(k)}{\partial\phi_j(k)}\right]^2\right]\lambda \tag{28}$$

or

$$\lambda < \frac{\beta\left[\phi_j(k) + \frac{\alpha}{\beta}e_i(k)\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]\right]^2}{\left[1 + \frac{\alpha}{\beta}\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]^2\right]}. \tag{29}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2022 at 10:20:10 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.



Since $\alpha > 0$ and $\beta > 0$, so the right-hand side of the above equation is always a positive number. Thus, $\lambda <$ positive number. This further implies $\lambda$ can also assume a negative value which is not desirable since to satisfy $\Delta v_L(k) < 0$, $\lambda$ should only attain positive value [as required by (23)] and, hence, we should not find the expression for $\Delta\phi_j(k)$ using $b^2 - 4ac > 0$ condition. Now, if $b^2 - 4ac < 0$ is used to find the expression for $\Delta\phi_j(k)$, then it implies complex roots (which we do not want). Finally, if we use $b^2 - 4ac = 0$, then we can write

$$\sqrt{\beta^2\left[\phi_j(k) + \frac{\alpha}{\beta}e_i(k)\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]\right]^2 - \beta\left[1 + \frac{\alpha}{\beta}\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]^2\right]\lambda} = 0. \tag{30}$$

Squaring of both the sides of the above equation will lead to the following expression:

$$\beta^2\left[\phi_j(k) + \frac{\alpha}{\beta}e_i(k)\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]\right]^2 - \beta\left[1 + \frac{\alpha}{\beta}\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]^2\right]\lambda = 0. \tag{31}$$

So, the expression of $\lambda$ will come out to be

$$\lambda = \frac{\beta^2\left[\phi_j(k) + \frac{\alpha}{\beta}e_i(k)\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]\right]^2}{\beta\left[1 + \frac{\alpha}{\beta}\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]^2\right]} \tag{32}$$

or

$$\lambda = \frac{\beta\left[\phi_j(k) + \frac{\alpha}{\beta}e_i(k)\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]\right]^2}{\left[1 + \frac{\alpha}{\beta}\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]^2\right]}. \tag{33}$$

Since the right-hand side of (33) is always positive, so we can write

$$\frac{\beta\left[\phi_j(k) + \frac{\alpha}{\beta}e_i(k)\left[\frac{\partial e_i(k)}{\partial\phi(k)}\right]\right]^2}{\left[1 + \frac{\alpha}{\beta}\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]^2\right]} > 0 \tag{34}$$

which implies $\lambda > 0$ in (33) [as required by (23)]. Thus, we will use $b^2 - 4ac = 0$ condition to find out the corresponding expression for $\Delta\phi_j(k)$ so that we always get a positive value of $\lambda$ and, hence, the condition $\Delta v_L(k) < 0$ gets satisfied. Now, the unique root of (25) can be obtained as follows:

$$\Delta\phi_j(k) = \frac{-b}{2a} = \frac{-\beta\left[\phi_j(k) + \frac{\alpha}{\beta}e_i(k)\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]\right]}{\beta\left[1 + \frac{\alpha}{\beta}\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]^2\right]} = \frac{-\left[\phi_j(k) + \frac{\alpha}{\beta}e_i(k)\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]\right]}{\left[1 + \frac{\alpha}{\beta}\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]^2\right]}. \tag{35}$$

On substituting $\Delta\phi_j(k)$ value obtained in the above equation into (10) will give

$$\phi_j(k) = \phi_j(k-1) + \eta_i(k)\frac{\left[\phi_j(k) + \frac{\alpha}{\beta}e_i(k)\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]\right]}{\left[1 + \frac{\alpha}{\beta}\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\right]^2\right]}. \tag{36}$$

This completes the proof. ■

## V. Adaptive Learning Rate Scheme Based on Lyapunov Stability and BP Method

The learning rate value has a significant impact on how quickly the model is trained. The convergence of error (to a small value) is guaranteed if it is set to a small value, but the training time (to achieve convergence) will be very long. Conversely, if it is set to a greater number, training time will be reduced but instability may result. As a result, it is important to carefully vary the learning rate value so that it can react to the situation at hand. Let us define the positive-definite Lyapunov function as

$$v_L(k) = \frac{1}{2}e_i^2(k) \tag{37}$$

and the stability is guaranteed if the following condition is satisfied:

$$\Delta v_L(k) = v_L(k) - v_L(k-1) < 0 \tag{38}$$

or

$$\Delta v_L(k) = v_L(k) - v_L(k-1) = \frac{1}{2}\left\{e_i^2(k) - e_i^2(k-1)\right\} \tag{39}$$

or

$$\Delta v_L(k) = \frac{1}{2}[e_i(k) + e_i(k-1)][e_i(k) - e_i(k-1)] \tag{40}$$

or

$$\Delta v_L(k) = \frac{1}{2}[\Delta e_i(k) + 2e_i(k)][\Delta e_i(k)] \tag{41}$$

or

$$\Delta v_L(k) = \Delta e_i(k)\left[\frac{1}{2}[\Delta e_i(k)] + e_i(k)\right]. \tag{42}$$

The term $e_i(k)$ is expanded in the Taylor series format and is written as follows:

$$e_i(k) = e_i(k-1) + \frac{\partial e_i(k)}{\partial\phi_j(k)}\Delta\phi_j(k) + H.O.T \tag{43}$$

where H.O.T refers to the higher order terms and are being neglected (because their insignificant contributions). Thus, the simplified expression of (43) can be written as follows:

$$e_i(k+1) - e_i(k) = \Delta e_i(k) = \frac{\partial e_i(k)}{\partial\phi_j(k)}\Delta\phi_j(k). \tag{44}$$

As per the gradient-descent principle, the change in the value of any parameter of the proposed model can be written as

$$\Delta\phi_j(k) = -\eta_i(k)e_i(k)\frac{\partial e_i(k)}{\partial\phi_j(k)}. \tag{45}$$

Since $([\partial e_i(k)]/[\partial\phi_j(k)]) = -([\partial y_{mr}(k)]/[\partial\phi_j(k)])$, so the above equation can be rewritten as follows:

$$\Delta\phi_j(k) = \eta_i(k)e_i(k)\frac{\partial y_{mr}(k)}{\partial\phi_j(k)}. \tag{46}$$

On substituting (44) in to (42) will result in the following equation:

$$\Delta v_L(k) = \left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\Delta\phi_j(k)\right]\left[\frac{1}{2}\left[\frac{\partial e_i(k)}{\partial\phi_j(k)}\Delta\phi_j(k)\right] + e_i(k)\right]. \tag{47}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2022 at 10:20:10 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.



Now, substituting $([\partial e_i(k)]/[\partial \phi_j(k)]) = ([\partial (y_d(k) - y_{mr}(k))]/\ [\partial \phi_j(k))]) = -([\partial y_{mr}(k)]/[\partial \phi_j(k)])$ and (46) in to (47) will result in

$$\Delta v_L(k) = \frac{1}{2}\left[\frac{\partial y_{mr}(k)}{\partial \phi_j(k)}\right]^4 \eta_i(k)^2 e_i^2(k) - \left[\frac{\partial y_{mr}(k)}{\partial \phi_j(k)}\right]^2 \eta_i(k) e_i^2(k) \quad (48)$$

or

$$\Delta v_L(k) = -\eta_i(k) e_i^2(k)\left[\frac{\partial y_{mr}(k)}{\partial \phi_j(k)}\right]^2 \times \left\{1 - \frac{\eta_i}{2}\left[\frac{\partial y_{mr}(k)}{\partial \phi_j(k)}\right]^2\right\}. \quad (49)$$

The learning procedure will be stable if the condition $\Delta v < 0$ is satisfied. Hence, we can write

$$-\eta_i(k) e_i^2(k)\left[\frac{\partial y_{mr}(k)}{\partial \phi_j(k)}\right]^2 \left\{1 - \frac{\eta_i(k)}{2}\left[\frac{\partial y_{mr}(k)}{\partial \phi_j(k)}\right]^2\right\} < 0. \quad (50)$$

Multiplying both the sides of (50) by $-1$ will lead to the reversal of the inequality sign, that is

$$\eta_i(k) e_i^2(k)\left[\frac{\partial y_{mr}(k)}{\partial \phi_j(k)}\right]^2 \left\{1 - \frac{\eta_i(k)}{2}\left[\frac{\partial y_{mr}(k)}{\partial \phi_j(k)}\right]^2\right\} > 0. \quad (51)$$

The above equation will hold only if the following conditions are satisfied concurrently, that is:

$$\eta_i(k) > 0 \quad (52)$$

and

$$\left\{1 - \frac{\eta_i(k)}{2}\left[\frac{\partial y_{mr}(k)}{\partial \phi_j(k)}\right]^2\right\} > 0 \quad (53)$$

or

$$\left\{2 - \eta_i(k)\left[\frac{\partial y_{mr}(k)}{\partial \phi_j(k)}\right]^2\right\} > 0. \quad (54)$$

So, for (51) to hold the second condition [obtained from (54)] is

$$\eta_i(k) < \frac{2}{\left[\frac{\partial y_{mr}(k)}{\partial \phi_j(k)}\right]^2}. \quad (55)$$

Thus, the stability of the learning procedure is ensured if $\eta_i(k)$ values satisfies the following condition:

$$0 < \eta_i(k) < \frac{2}{\left[\frac{\partial y_{mr}(k)}{\partial \phi_j(k)}\right]^2} \quad (56)$$

and the recursive equation for tuning the $\eta_i(k)$ value can be written as follows:

$$\eta_i(k) = \frac{2}{\left[\frac{\partial y_{mr}(k)}{\partial \phi_j(k)}\right]^2}. \quad (57)$$

The MRENN model-based series–parallel identification configuration is shown in Fig. 2. It can be seen from the figure that the MRENN model only requires the unit-delayed value of the plant's output $y_d(k-1)$ and the present value of the external input $r(k)$ for computing its current output. The following Algorithm 1 describes the sequence of the implementation steps to carry out the MRENN training and validation stages.

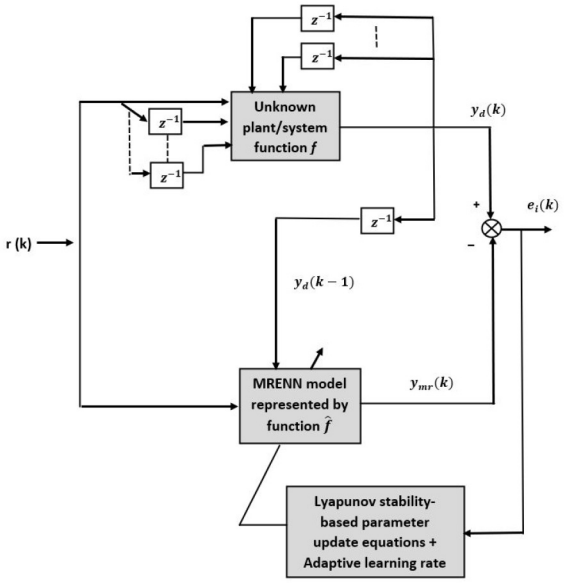



Fig. 2. Identification configuration based on the Lyapunov-stability-based MRENN model in series–parallel mode.

## VI. Simulation Results

The performance of the suggested model is assessed using two benchmark examples. Some error-based performance measures, such as root MSE (RMSE), mean average error (MAE), and mean absolute percentage error (MAPE), are used to assess the model's accuracy (in approximating the unknown dynamics of the plant). They are defined as follows:

$$\text{RMSE}(k) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_d(k) - y_{mr}(k))^2} \quad (58)$$

and

$$\text{MAE} = \frac{1}{n}\left\{\sum_{i=1}^{n}|y_d(k) - y_{mr}(k)|\right\} \quad (59)$$

and

$$\text{MAPE} = \frac{1}{n}\left[\sum_{i=1}^{n}\frac{|y_d(k) - y_m(k)|}{|y_d(k)|}\right] \times 100 \quad (60)$$

where the total count of input–output data samples during any iteration is denoted by $n$. Furthermore, in each of the simulation examples, the number of hidden-layer neurons (in each of the neural network models) is considered equal to 10 and the weights initial values are kept small. Also, the initial value of the adaptive learning rate $\eta_i(k)$ is set equal to 0.009 and the values of $\alpha$ and $\beta$ are set equal to 0.05.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2022 at 10:20:10 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.



**Algorithm 1** Pseudocode for Training and Validation of MRENN Model

1: Firstly, the number of the hidden layer neurons in the proposed MRENN model is set equal to 10 and the series-parallel identification configuration based on the MRENN model is set up.
2: Next, all the parameters of the MRENN model and $\alpha$ and $\beta$ values are initialized to a very small value (lying in a range of $(-1, 1)$). The initial value of the learning rate is also set close to zero.
3: Then the MRENN model and the plant are applied with a bounded external input $r(k)$ and the training is continued for the required number of epochs.
4: The identification program is executed with an iteration number $k = 1$ and the output of the MRENN model is calculated using $y_{mr}(k) = \left\{\sum_{i=1}^{m} s_i(k) w_i^o(k)\right\} + \left\{\sum_{i=1}^{p} x_i(k)\theta_i(k)\right\} + c_{21}(k) w^{oc}(k) - w_o^h(k)$ (Eq.6).
5: Identification error is then calculated using $e_i(k) = y_d(k) - y_{mr}(k)$. This error is then processed by the Lyapunov-stability-based parameter update equation (Eq.36) along with Eq.57 for determining the updated values of parameters and learning rate respectively.
6: Next incrementing the time index by 1 (that is, $k = k + 1$). This procedure will continue until all the epochs are exhausted/completed.
7: At the end of the training, the parameter values are stored. The validation step is then executed in which the trained MRENN model and the plant are subjected to a new external input and their corresponding output responses are recorded and compared to test whether the proposed model is trained well enough or not.

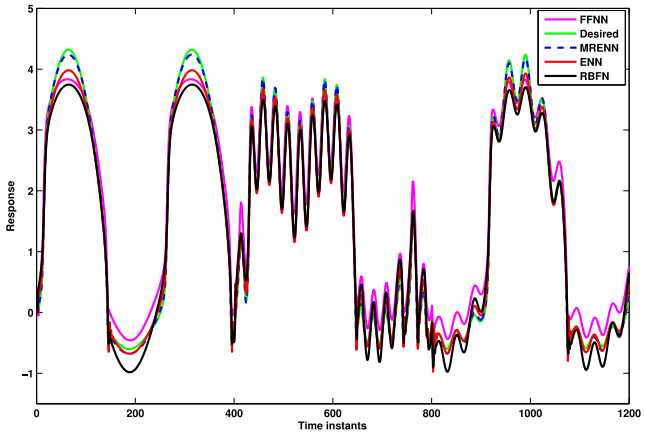



Fig. 3. Validation response obtained from the trained MRENN, ENN, RBFN, and FFNN models (Example 1).

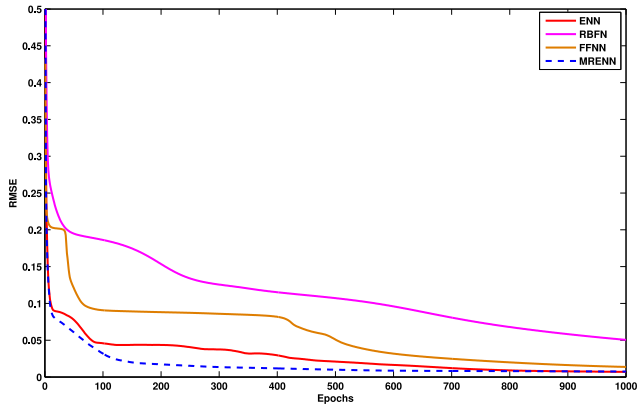



Fig. 4. RMSE plot obtained during the training of MRENN, ENN, RBFN, and FFNN models (Example 1).

### A. Example 1: Nonlinear Time-Delayed Identification

Let the dynamic difference equation of a nonlinear dynamical plant is given as follows [33]:

$$y_d(k) = \frac{y_d(k-1)y_d(k-2)\left[y_d(k-1)+2.5\right]}{\left[1+y_d^2(k-1)+y_d^2(k-2)\right]} + r(k) \quad (61)$$

where $r(k) = \sin(2\pi k/25)$ represents the externally applied input signal and is applied to the plant for generating the training data. The proposed MRENN model is applied with only two input signals, viz., the unit-delayed output value obtained from the plant and the current value of the externally applied input signal. The ENN, RBFN, and FFNN models are applied with the same number of inputs on which the output of the given plant depends upon. Thus, the output equations of MERNN, ENN, FFNN, and RBFN models are given as follows:

$$\begin{aligned} y_{mr}(k) &= \hat{f}_1\{y_d(k-1), r(k)\} \\ y_{\text{ENN}}(k) &= \hat{f}_2\{y_d(k-1), y_d(k-2), r(k)\} \\ y_{\text{FFNN}}(k) &= \hat{f}_3\{y_d(k-1), y_d(k-2), r(k)\} \\ y_{\text{RBFN}}(k) &= \hat{f}_4\{y_d(k-1), y_d(k-2), r(k)\}. \end{aligned} \quad (62)$$

The training is continued for 1000 epochs with 500-time steps in each epoch. At the end of the training, the parameters of the models are saved. These trained models and the given nonlinear plant are applied with a new external input $r(k)$ (validation step). It is defined as follows:

$$r(k) = \begin{cases} \sin\left(\frac{2\pi k}{250}\right) & \text{if } k \le 400 \\ 0.35\sin\left(\frac{\pi k}{210}\right) + 0.25\cos\left(\frac{2\pi k}{150}\right) & \\ +0.4\sin\left(\frac{4\pi k}{50}\right) & \text{if } 400 < k \le 800 \\ 0.8\sin\left(\frac{2\pi k}{300}\right) + 0.2sin\left(\frac{2\pi k}{35}\right) & \text{if } 800 < k \le 1200. \end{cases} \quad (63)$$

The output response obtained from these models is shown in Fig. 3. It can be observed that the response of the MRENN model is very close to the desired response as compared to the other three models. The RMSE and MAE plots are shown in Figs. 4 and 5, respectively, and the same conclusion can be drawn from these plots since the error can be seen to decrease at a rapid rate in the case of the MRENN model.

Finally, the various details regarding the error-based performance indicators values are shown in Table I. It can be easily observed that MRENN has given lower values (shown in bold letters) of each of the performance indicators which shows its superior performance over the other models.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2022 at 10:20:10 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.



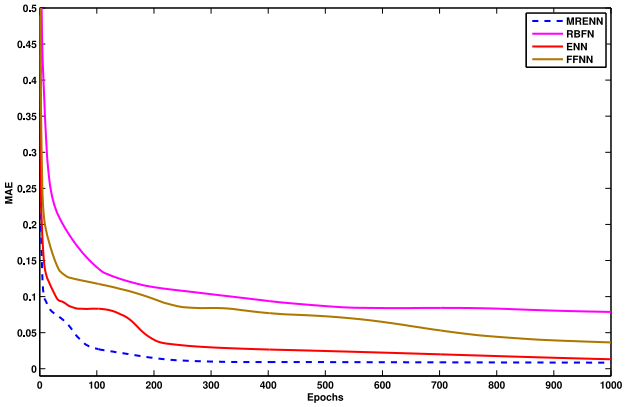



Fig. 5. MAE plot obtained during the training of MRENN, ENN, RBFN, and FFNN models (Example 1).

TABLE I
PERFORMANCE INDICATOR VALUES (EXAMPLE 1)

| S.no | Model | Epochs | RMSE | MAE | MAPE (%) |
|---|---|---|---|---|---|
| 1 | MRENN | 1000 | **0.0029** | **0.0037** | **2.03** |
| 2 | ENN | 1000 | 0.0089 | 0.0064 | 4.51 |
| 4 | FFNN | 1000 | 0.0601 | 0.0211 | 6.21 |
| 5 | RBFN | 1000 | 0.0897 | 0.0574 | 7.98 |

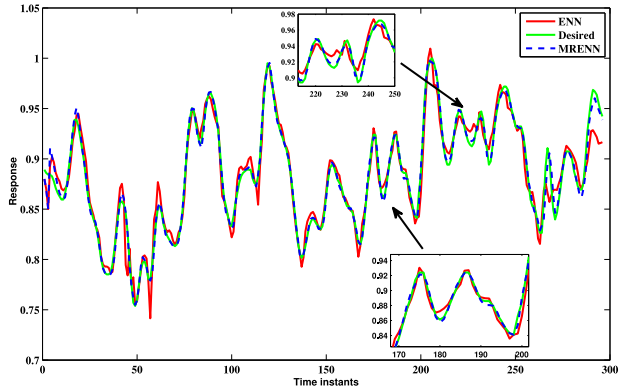



Fig. 6. Response obtained from MRENN and ENN models (Example 2).

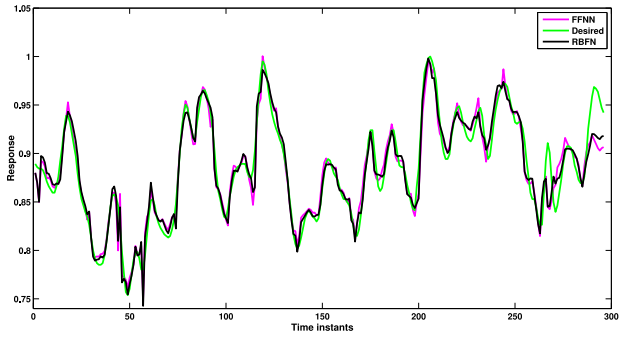



Fig. 7. Response obtained from FFNN and RBFN models (Example 2).

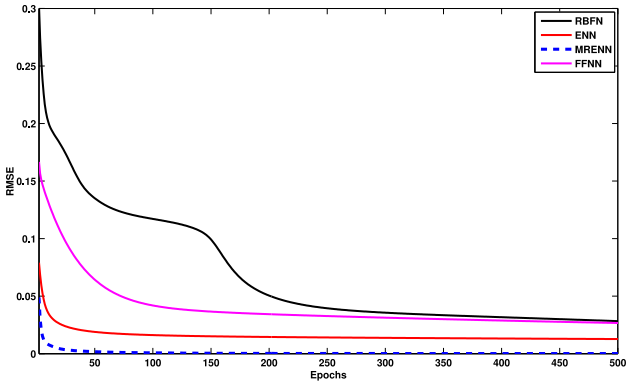



Fig. 8. RMSE plots obtained from MRENN, ENN, FFNN, and RBFN models (Example 2).

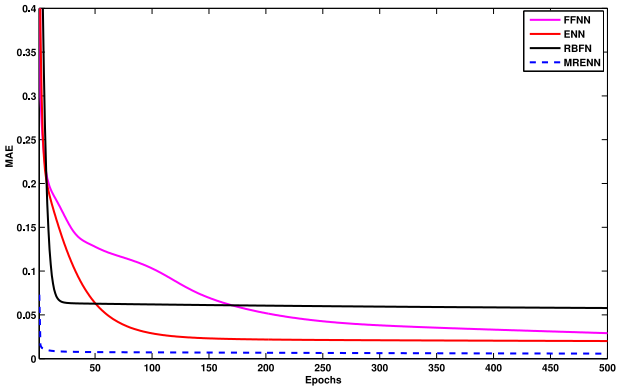



Fig. 9. MAE plots obtained from MRENN, ENN, FFNN, and RBFN models (Example 2).

### *B. Example 2: Identification of Box and Jerkin's Gas Furnace Time-Series Problem*

In this example, another benchmark time-series problem of Box–Jenkins gas furnace data is considered to test the performance of the proposed model. This time series is widely used by researchers to test their identification and modeling tools. The data set is recorded from the process of combustion of an air–methane gas mixture. The flow rate of the gas is maintained constant but the portion of the methane is changed randomly [which is acting as the input variable, $r(k)$]. The combustion of the gas in the furnace resulted in the emission of the $CO_2$ whose concentration represents the output, $y_d(k)$ (of the time series). A total of 296 input–output data pairs with a sampling interval of 9 s are recorded through the measurements. As per [38], the current value of $y_d(k)$ is influenced by six inputs, viz., $y_d(k-1)$, $y_d(k-2)$, $y_d(k-3)$, $r(k-1)$, $r(k-2)$, and $r(k-3)$. These six signals are applied as inputs to the ENN, FFNN, and RBFN models, but the proposed MRENN model is applied only with two input signals, viz., $y_d(k-1)$ and $r(k)$. The training is continued for 500 epochs and the output response obtained from MRENN and ENN models and FFNN and RBFN models (at the end of the last epoch) are shown in Figs. 6 and 7, respectively. From both the figures it can be seen that the output of the MRENN model is following the desired response more closely as

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2022 at 10:20:10 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 II
PERFORMANCE INDICATOR VALUES (EXAMPLE 2)

| S.no | Model | Epochs | RMSE | MAE | MAPE (%) |
|---|---|---|---|---|---|
| 1 | MRENN | 500 | **0.00021** | **0.0012** | **1.254** |
| 2 | ENN | 500 | 0.0048 | 0.0039 | 2.114 |
| 4 | FFNN | 500 | 0.0451 | 0.0110 | 3.574 |
| 5 | RBFN | 500 | 0.0498 | 0.0153 | 3.699 |

compared to other models. Furthermore, the performance of both FFNN and RBFN models is found to be quite similar. The corresponding RMSE and MAE plots are depicted in Figs. 8 and 9, respectively. Both error-based indicators are showing a rapid decrease in their values (as the training continued) with the MRENN model (with their final values listed in Table II) which again shows a superior performance shown by the MRENN model over the other considered models.

## VII. CONCLUSION

The MRENN model, an enhanced variation of the traditional ENN model, was created and used to approximate the unidentified dynamics of nonlinear complex systems. The performance of the MRENN model was also compared to that of the other well-known neural networks, such as ENN, RBFN, and FFNN, in order to assess its effectiveness. The simulation analysis demonstrated that the MRENN model outperformed the other models that were taken into consideration. This is owing to the weighted self-feedback connections that are present in the context nodes, the weighted direct connections between the input-layer signals and the output-layer neuron, and the additional weights that the additional context layer has contributed. The MRENN model's improved performance was the result of these improvements. Furthermore, the proposed model was applied with only two inputs, viz., unit-delayed output of the plant and the present value of external input. This, in turn, eliminated the need of knowing the order of the given plant. An adaptive BP learning algorithm was created based on the Lyapunov-stability method to ensure the stability of the whole system. An adaptive learning rate system was also built in order to enhance the training pace. Through the simulation examples, it was discovered that the MRENN model performs significantly better than the ENN, RBFN, and FFNN models. The future work will involve the development of the pruning strategy, which will aid in optimizing the count of neurons required in the MRENN model. Another objective will be design a controller based on the MRENN model and use it to regulate the nonlinear plant outputs.

## REFERENCES


[1] H. Zhao, S. Gao, Z. He, X. Zeng, W. Jin, and T. Li, "Identification of nonlinear dynamic system using a novel recurrent wavelet neural network based on the pipelined architecture," *IEEE Trans. Ind. Electron.*, vol. 61, no. 8, pp. 4171–4182, Aug. 2014.

[2] M. Alfaro-Ponce, A. A. Cruz, and I. Chairez, "Adaptive identifier for uncertain complex nonlinear systems based on continuous neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 3, pp. 483–494, Mar. 2014.

[3] C.-T. Lin, N. R. Pal, S.-L. Wu, Y.-T. Liu, and Y.-Y. Lin, "An interval type-2 neural fuzzy system for online system identification and feature elimination," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 7, pp. 1442–1455, Jul. 2015.

[4] J. Cervantes, W. Yu, S. Salazar, and I. Chairez, "Takagi–Sugeno dynamic neuro-fuzzy controller of uncertain nonlinear systems," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 6, pp. 1601–1615, Dec. 2017.

[5] C. Yang, Z. Li, R. Cui, and B. Xu, "Neural network-based motion control of an underactuated wheeled inverted pendulum model," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 11, pp. 2004–2016, Nov. 2014.

[6] H. V. H. Ayala, D. Habineza, M. Rakotondrabe, and L. dos Santos Coelho, "Nonlinear black-box system identification through coevolutionary algorithms and radial basis function artificial neural networks," *Appl. Soft Comput.*, vol. 87, Feb. 2020, Art. no. 105990.

[7] P. Agand, M. A. Shoorehdeli, and A. Khaki-Sedigh, "Adaptive recurrent neural network with Lyapunov stability learning rules for robot dynamic terms identification," *Eng. Appl. Artif. Intell.*, vol. 65, pp. 1–11, Oct. 2017.

[8] H.-G. Han, L. Zhang, Y. Hou, and J.-F. Qiao, "Nonlinear model predictive control based on a self-organizing recurrent neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 2, pp. 402–415, Feb. 2016.

[9] R. Kumar, S. Srivastava, and J. R. P. Gupta, "Comparative study of neural networks for control of nonlinear dynamical systems with Lyapunov stability-based adaptive learning rates," *Arabian J. Sci. Eng.*, vol. 43, no. 6, pp. 2971–2993, 2018.

[10] L. Jin, Z. Huang, Y. Li, Z. Sun, H. Li, and J. Zhang, "On modified multi-output Chebyshev-polynomial feed-forward neural network for pattern classification of wine regions," *IEEE Access*, vol. 7, pp. 1973–1980, 2018.

[11] K. Nouri, R. Dhaouadi, and N. B. Braiek, "Adaptive control of a nonlinear DC motor drive using recurrent neural networks," *Appl. Soft Comput.*, vol. 8, no. 1, pp. 371–382, 2008.

[12] J. Tang, C. Deng, and G.-B. Huang, "Extreme learning machine for multilayer perceptron," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 4, pp. 809–821, Apr. 2016.

[13] C.-H. Chang, "Deep and shallow architecture of multilayer neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2477–2486, Oct. 2015.

[14] Q. Shi, R. Katuwal, P. N. Suganthan, and M. Tanveer, "Random vector functional link neural network based ensemble deep learning," *Pattern Recognit.*, vol. 117, Sep. 2021, Art. no. 107978.

[15] B. Yang, J. Cao, R. Ni, and Y. Zhang, "Facial expression recognition using weighted mixture deep neural network based on double-channel facial images," *IEEE Access*, vol. 6, pp. 4630–4640, 2017.

[16] Y. Park and H. S. Yang, "Convolutional neural network based on an extreme learning machine for image classification," *Neurocomputing*, vol. 339, pp. 66–76, Apr. 2019.

[17] A. Perrusquía and W. Yu, "Identification and optimal control of nonlinear systems using recurrent neural networks and reinforcement learning: An overview," *Neurocomputing*, vol. 438, pp. 145–154, May 2021.

[18] J. Wang, L. Zhang, Q. Guo, and Z. Yi, "Recurrent neural networks with auxiliary memory units," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1652–1661, May 2018.

[19] Y.-Y. Lin, J.-Y. Chang, and C.-T. Lin, "Identification and prediction of dynamic systems using an interactively recurrent self-evolving fuzzy neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 2, pp. 310–321, Feb. 2013.

[20] J. Zhou, L. Dong, W. Guan, and J. Yan, "Impact load identification of nonlinear structures using deep recurrent neural network," *Mech. Syst. Signal Process.*, vol. 133, Nov. 2019, Art. no. 106292.

[21] X. Jin, J. Shao, X. Zhang, W. An, and R. Malekian, "Modeling of nonlinear system based on deep learning framework," *Nonlinear Dyn.*, vol. 84, no. 3, pp. 1327–1340, 2016.


Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2022 at 10:20:10 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.



[22] Y.-C. Cheng, W.-M. Qi, and W.-Y. Cai, "Dynamic properties of Elman and modified Elman neural network," in *Proc. Int. Conf. Mach. Learn. Cybern.*, vol. 2, 2002, pp. 637–640.
[23] N. A. S. Al-Jamali and H. S. Al-Raweshidy, "Modified Elman spike neural network for identification and control of dynamic system," *IEEE Access*, vol. 8, pp. 61246–61254, 2020.
[24] H.-W. Ge, F. Qian, Y.-C. Liang, W.-L. Du, and L. Wang, "Identification and control of nonlinear systems by a dissimilation particle swarm optimization-based Elman neural network," *Nonlinear Anal. Real World Appl.*, vol. 9, no. 4, pp. 1345–1360, 2008.
[25] J. Ni, C. K. Ahn, L. Liu, and C. Liu, "Prescribed performance fixed-time recurrent neural network control for uncertain nonlinear systems," *Neurocomputing*, vol. 363, pp. 351–365, Oct. 2019.
[26] S. Ganjefar and M. Tofighi, "Single-hidden-layer fuzzy recurrent wavelet neural network: Applications to function approximation and system identification," *Inf. Sci.*, vol. 294, pp. 269–285, Feb. 2015.
[27] H.-W. Ge, W.-L. Du, F. Qian, and Y.-C. Liang, "Identification and control of nonlinear systems by a time-delay recurrent neural network," *Neurocomputing*, vol. 72, nos. 13–15, pp. 2857–2864, 2009.
[28] X. Li, Z. Chen, and Z. Yuan, "Simple recurrent neural network-based adaptive predictive control for nonlinear systems," *Asian J. Control*, vol. 4, no. 2, pp. 231–239, 2002.
[29] J. A. R. Vargas, W. Pedrycz, and E. M. Hemerly, "Improved learning algorithm for two-layer neural networks for identification of nonlinear systems," *Neurocomputing*, vol. 329, pp. 86–96, Feb. 2019.
[30] F. Bonassi and R. Scattolini, "Recurrent neural network-based internal model control design for stable nonlinear systems," *Eur. J. Control*, vol. 65, May 2022, Art. no. 100632.
[31] X. Yao, Z. Wang, and H. Zhang, "Prediction and identification of discrete-time dynamic nonlinear systems based on adaptive echo state network," *Neural Netw.*, vol. 113, pp. 11–19, May 2019.
[32] H.-G. Han, Z.-Y. Chen, H.-X. Liu, and J.-F. Qiao, "A self-organizing interval type-2 fuzzy-neural-network for modeling nonlinear systems," *Neurocomputing*, vol. 290, pp. 196–207, May 2018.
[33] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Trans. Neural Netw.*, vol. 1, no. 1, pp. 4–27, Mar. 1990.
[34] Z. Zeng and J. Wang, "Design and analysis of high-capacity associative memories based on a class of discrete-time recurrent neural networks," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 6, pp. 1525–1536, Dec. 2008.
[35] H. D. Patino and D. Liu, "Neural network-based model reference adaptive control system," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 30, no. 1, pp. 198–204, Feb. 2000.
[36] Q. Liu and J. Wang, "A one-layer recurrent neural network for constrained nonsmooth optimization," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 5, pp. 1323–1333, Oct. 2011.
[37] Y. W. Wong, K. P. Seng, and L.-M. Ang, "Radial basis function neural network with incremental learning for face recognition," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 4, pp. 940–949, Aug. 2011.
[38] S. Purwar, I. N. Kar, and A. N. Jha, "On-line system identification of complex systems using Chebyshev neural networks," *Appl. Soft Comput.*, vol. 7, no. 1, pp. 364–372, 2007.

**Rajesh Kumar** received the B.Tech. degree in applied electronics and instrumentation engineering from Manav Rachna College of Engineering, Faridabad (Maharshi Dayanand University, Rohtak), India, in 2011, the M.Tech. degree in electrical engineering (control systems) from the National Institute of Technology Kurukshetra, Mirzapur, India, in 2013, and the Ph.D. degree in the area of control and identification of nonlinear systems using intelligent tools from Netaji Subhas Institute of Technology (University of Delhi), New Delhi, India, in 2019.

He has published number of research papers in various high-quality international SCI/SCIE journals. His current research interests include artificial neural networks, intelligent modeling and control, system identification, stability analysis, and optimization.

Dr. Kumar is also serving as the reviewer for several reputed international journals.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2022 at 10:20:10 UTC from IEEE Xplore. Restrictions apply.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/362016540

# Multilabel Facial Emotion Classification

**Conference Paper** · July 2022

DOI: 10.6084/m9.figshare.20318439

CITATIONS
0

READS
16

**2 authors**, including:

Mansi Panchal
Delhi Technological University
**1** PUBLICATION **0** CITATIONS

SEE PROFILE

**Some of the authors of this publication are also working on these related projects:**



Multilabel Facial Emotion Classification View project

All content following this page was uploaded by Mansi Panchal on 15 July 2022.

The user has requested enhancement of the downloaded file.

# Multilabel Facial Emotion Classification

Mansi Bhardwaj
Department Of Applied Mathematics
Delhi Technological University
Delhi, India
mansi.bhardwaj3rd@gmail.com

Mansi Panchal
Department Of Applied Mathematics
Delhi Technological University
Delhi, India
mansi130698@gmail.com

*Abstract*— **A Multi label emotion recognition system is an important field of research and also has lots of applications. Examples include classify physically disabled people (deaf, dumb, and bedridden) and Autism children's emotional expressions based on facial landmarks. Understanding different human emotions is of great relevance as to gain insights into human cognition and affect as well as for the design of computational models and perceptual interfaces. Us as human beings expresses our feelings through distinct emotions yet most part, research studies have been limited to six basic categories—happiness, surprise, sadness, anger, fear, and disgust which restricts us from the real world problems. Here, in this paper we classified an important group of expressions, which we call compound emotion categories. Compound emotions are that can be visualized as combination of basic component categories to create new ones. For instance, happily surprised and angrily surprised are two distinct compound emotion categories**.

***Keywords— Compound emotions, facial emotions, First order statistics****(FOS)* ***, gray level co-occurrence matrix (****GLCM****), Local Binary Pattern(****LBP****)***

## I. INTRODUCTION

FACES are keys to individual emotion, they play a major role in communication and interaction that makes machine understanding, perception, and modeling of human expression, an important problem in computer vision.

Mehrabian, a famous psychologist, found from his research that the emotional data that humans classify as emotions are distributed in sections. He found that only 7% of the emotional data total is passed by language, and 38% is transported by our language auxiliary, The associate editor coordinating the review of this manuscript and approving it for publication was Rosalia Maglietta . which differs from culture to culture, such as the rhythm of speech, tone, pitch, etc. So far, the highest percentage of emotional data shown by facial expression is 55% [5].

There is a significant amount of research on facial expressions in computer vision. Perhaps the most fundamental problem in this area is how to categorize the facial expressions to extract information about the underlying emotional states. In this paper we have considered the problem of classifying the different compound emotions on the basis of the relative change in people's facial expressions. Mostly research is limited to these set of the 6 basic emotions 'Neutral', 'Happy', 'Sad', 'Angry', 'Surprised ' 'Fearful' [4].

In the present work, we demonstrate that the production and visual perception of compound emotion class that constitutes the combination of the given basic emotions. The problem of multilabel arises as the repertoire of facial expressions typically used by humans is better described using a rich set of basic and compound categories rather than a small set of basic elements. So the problem of multi - label emotion classification is of great relevance as human's expresses there feeling s and emotions not just through the basic emotions instead in real world the human emotional can be expressed through several combinational emotions too.

The work is categorized in 3 major parts, first step is to analyze the training dataset and extract the region of interest (ROI) which is the most active region on the face when a person changes its facial expressions. Second step is to detect multiple image features so system can differentiate the facial landmarks in different emotional state of a person's face. And finally in the third step the main approach for multi- label classification of the sets of emotions is to compare the difference in the value set of the relevant features for different emotion class and then labeling them as a category in the compound emotions.

## II. OBJECTIVE

The scenario is about classifying different compound emotions that are combinations of the basics emotions into different classes. While past research had identified facial expressions associated with a single label that are not enough to express the complete feeling of a person in this paper we have studied an important class of facial expressions observed the compound emotions.

For example the facial expression of happy surprise when we feel joyful in a surprised way, as, for example, at a surprise birthday party. So here we have identified 7 classes of the compound emotions which are consistently produced across cultures, suggesting that the number of facial expressions is more in numbers than previously believe.

The hypothesis is that human emotions are not restricted to the class of the basic six emotions rather there exist an important class of compound emotions that are more expressive and informative about a person's mental state and mood which we humans expresses almost regularly. It is of major use to detect such class of emotions. To date, most research has focused on the study of six emotions typically seen in most cultures are happiness, surprise, sadness, anger, fear, and neutral while this paper considers the class of compound emotions.

## III. RELATED WORK

So recognition of multi label facial emotions is a challenging problem with many applications. . Previous

work of the problem has typically focused by categorizing a set of such predetermined facial motions as in FACS (Facial Action Coding System), that is all visually distinguishable facial movements.

Ekman and Friesen [1] developed the Facial Action Coding System (FACS) for describing subtle changes in facial expressions. FACS consists of 49 action units, including those for head and eye positions. Thirty of these are anatomically related to contraction of a specific set of facial muscles. Although there are a small number of atomic action units, more than 7,000 combinations of action units have been observed. FACS provides the necessary detail with which to describe facial expression.
Another approach applied to solve the problem was proposed as division the face into low, mid, and high levels[2] .At the low level we take regions corresponding to the face, mouth, eyebrows, and eyes and model the rigid and non-rigid motions of these regions using a collection of parameterized flow models. and non-rigid facial motions using a collection of local parameterized optical flow models.

These approaches are bit complex as action units [3] is a difficult problem because there are no quantitative definitions and they appear in complex combinations while in this paper, rather than using the prior method of action units to recognize any respective emotion we have approached the problem by a different procedure, here we firstly detected the facial landmarks which further differentiated the face into 8 facial regions which are the most sensitive areas when facial emotion changes thus are of greater relevance .The following 8 facial regions are mouth , inner mouth , right eye ,left eye ,right eyebrow ,left eyebrow, nose , Jaw. Then the respective features from each of the face regions are considered as the tools to classify the face emotions.

## IV. METHODOLOGY

This section explains the entire step-by-step approach employed in this work, which primarily consists of the following three phases: The initial step in the process is to identify the important facial landmarks that give key facial features or regions. The question now is why facial landmarks are so important in our case.

Face landmark detection is a subset of the shape prediction issue. A shape predictor attempts to pinpoint key points of interest along a shape given an input image (and typically a ROI that describes the item of interest)[10]. Our goal in the context of facial landmarks is to use shape prediction methods to discover essential facial structures on the face [6].Detecting facial landmarks is therefore a two step process:

- Identify the face in the photo.
  Face landmarking is the recognition and localisation of specific keypoints on the face, which is perhaps the most critical stage in the face processing process. The issue is to locate the location of a number of landmarks on a face based on a photograph of the face. The eye corners, nose tip, nostril corners, mouth corners, terminal points of the eyebrow arcs, jaw, and other landmarks are commonly employed.
  The dlib library's pre-trained facial landmark detector is used to estimate the location of 68 (x, y)-coordinates that correspond to facial structures.

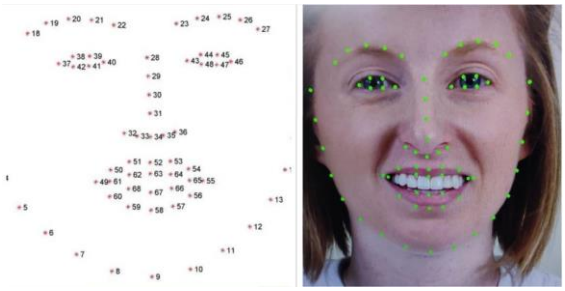


Figure 1. Detecting facial landmarks

- On the face ROI, identify the important facial structures.
  Detecting the face's Region of Interest entails segmenting the face into the parts we're interested in, which are the most active on the face when a person alters their facial expressions. Both the shape and layout features rely on good segmentation techniques.

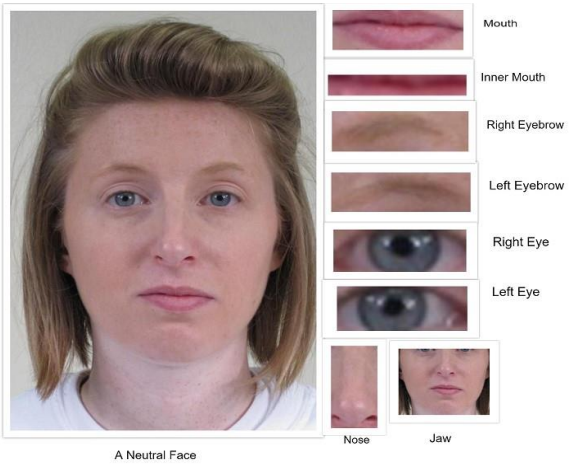



Figure 2 : Extracting Facial Features From Face

The extraction of important features from the split region of interest is the focus of the second step. Face processing begins with feature extraction, which is the first step in detecting and recognizing faces [9].

Feature extraction is a technique for extracting the visual content of photographs so that they may be indexed and retrieved. Feature extraction is a term that refers to a piece of data that is useful in completing a computational task for a certain application.

The purpose is to turn the image into quantitative data that can be further processed for emotion labelling, classification, and recognition. The Image feature classes applied to all the ROI of faces in the dataset are listed in the table below.

| Feature class | Features |
|---|---|
| FOS Feature | FOS_Mean,FOS_Variance, FOS_Median, FOS_Mode, FOS_Skewness, FOS_Kurtosis, FOS_Energy, FOS_Entropy, FOS_MinimalGrayLevel, FOS_MaximalGrayLevel, FOS_CoefficientOfVariation FOS_10Percentile, FOS_25Percentile,FOS_75Percentile, FOS_90Percentile, FOS_HistogramWidth |
| GLCM Feature | GLCM_ASM, GLCM_Contrast, GLCM_Correlation, GLCM_SumOfSquaresVariance, GLCM_InverseDifferenceMoment, GLCM_SumAverage, GLCM_SumVariance, GLCM_SumEntropy, GLCM_Entropy, GLCM_DifferenceVariance, GLCM_DifferenceEntropy, GLCM_Information1 ,GLCM_Information2, GLCM_MaximalCorrelationCoefficient |
| LBP Feature | energy, entropy n meanfeature |

Table 1 : Feature class with there

*A.* First Order Statistics

First order feature extraction is a retrieval approach based on image histogram features. The Histogram depicts the likelihood of a given value of grayscale pixel degree occurring in a picture. Several first-order parameters can be derived from the numbers provided by the histogram, including mean, variance, skewness, kurtosis, and entropy.

*B.* GLCM Feature

The GLCM functions calculate how often pairs of pixels with given values and in a specified spatial relationship appear in an image, create a GLCM, and then extract statistical measures from this matrix to characterise the texture of an image[8]. After you've created the GLCMs with graycomatrix, you can use graycoprops to get many statistics from them. These statistics give information on an image's texture. The statistics are listed in the table below. GLCM Contrast(Measures the local variations in the gray-level co-occurrence matrix. ), GLCM Correlation(Measures the joint probability occurrence of the specified pixel pairs. ), GLCM_ Energy(Provides the sum of squared elements in the GLCM. ), GLCM Correlation(Measures the joint probability occurrence of the specified pixel pairs. ), GLCM Correlation( Also known as the angular second moment or homogeneity.) and so on.

*C.* LBP Feature

Local Binary Pattern (LBP) is a basic yet effective texture operator that labels pixels in an image by thresholding each pixel's vicinity. This method is commonly used to investigate their local qualities and to discover the characteristics of specific image sections.

| Index | feature values | feature class2 |
|---|---|---|
| 1 | 134.5020438 | FOS based features |
| 2 | 1243.54216 | FOS based features |
| 3 | 133 | FOS based features |
| 4 | 132 | FOS based features |
| 5 | -0.252935347 | FOS based features |
| 6 | 2.588867051 | FOS based features |
| 7 | 0.00844632 | FOS based features |
| 8 | 4.914049627 | FOS based features |
| 9 | 24 | FOS based features |
| 10 | 205 | FOS based features |
| 11 | 0.085742657 | FOS based features |
| 12 | 90 | FOS based features |
| 13 | 110 | FOS based features |
| 14 | 162 | FOS based features |
| 15 | 182 | FOS based features |
| 16 | 92 | FOS based features |
| 17 | 0.000572526 | GLCM based features |
| 18 | 1444.773973 | GLCM based features |
| 19 | 0.410776373 | GLCM based features |
| 20 | 1230.692253 | GLCM based features |
| 21 | 0.164423322 | GLCM based features |
| 22 | 258.8987459 | GLCM based features |
| 23 | 3477.995038 | GLCM based features |
| 24 | 7.787598227 | GLCM based features |
| 25 | 11.97297921 | GLCM based features |
| 26 | 0.000367446 | GLCM based features |
| 27 | 5.01695243 | GLCM based features |
| 28 | -0.307968139 | GLCM based features |
| 29 | 0.978515737 | GLCM based features |
| 30 | 4.152016136 | GLCM based features |
| 31 | 0.001996626 | GLCM based features |
| 32 | 2093.328607 | GLCM based features |
| 33 | 0.852582818 | GLCM based features |
| 34 | 18.55723427 | GLCM based features |
| 35 | 0.600926965 | GLCM based features |
| 36 | 12.99940654 | GLCM based features |
| 37 | 2144.897065 | GLCM based features |
| 38 | 0.400581843 | GLCM based features |
| 39 | 3.559631582 | GLCM based features |
| 40 | 0.001691831 | GLCM based features |
| 41 | 4.537667866 | GLCM based features |
| 42 | 0.500995358 | GLCM based features |
| 43 | 0.037840544 | GLCM based features |
| 44 | 2.227745037 | GLCM based features |
| 45 | 5.334359481 | GLCM based features |
| 46 | -1.630513282 | GLCM based features |
| 47 | 10.25361091 | GLCM based features |
| 48 | -2.268206449 | LBP based features |
| 49 | 16.20969096 | LBP based features |
| 50 | -2.71380685 | LBP based features |

Table 2 : Numerical feature values of ROI (mouth)

The third step focuses on dealing with the numerical data we obtain from the feature extraction procedure, with the first objective being to identify differences in the respective features and create a dataset that can be used in model training. We used the difference of characteristics determined in the second phase from each basic emotion to each compound emotion, i.e. the difference of features of pleasant surprise emotion from six basic emotions. Following that, we used various classifiers to examine our model and determine its accuracy rate.

## V. EXPERIMENT & RESULTS

In this section, we tested multiple classifiers to see how accurate our experiment was. We utilised the free source machine learning framework "MEKA" for this. It helps with the creation, testing, and assessment of multi-label and multi-target classifiers.

A data instance can be connected with numerous labels in the multi-label problem. This is in contrast to the typical task of single-label classification (i.e., multi-class, or binary), which assigns a single class label to each instance. The multi-label context is gaining popularity, and it may be used to a wide range of domains, including text, music, photos, and video, as well as bioinformatics.

The result of our experiment with different classifier as follows :

| Classifier Name | Accuracy Rate (%) |
|---|---|
| BPNN | 73% |
| Binary Relevance Method (BR Method) | 83.4% |
| Probabilistic Classifier Chain (PCC) | 87% |
| MCC | 87.1% |
| MULAN | 92.2% |

Table 3 : Accuracy rate for different classifiers

It can be observed that to increase the accuracy of our experiment we have used four types of classifier and with MULAN [7] we get the higher accuracy.

Next, we have performed the above classifier taking cross validation as evaluation parameter and the result as follows :

| Classifier Name | Accuracy Rate (%) |
|---|---|
| BPNN | 75% |
| Binary Relevance Method (BR Method) | 84.5% |
| Probabilistic Classifier Chain (PCC) | 88% |
| MCC | 88% |
| MULAN | 92.2% |

Table 4 : Accuracy rate with cross validation parameter

Again we can observed from Table 4 that using cross-validation as a evaluation parameter, there is a minimal increment of accuracy rate in all classifier except MULAN.
The results presented in this research reveal that, in case of facial expression emotion classifier, its features have valuable information about emotions that cannot be extracted from the visual information. This redundant information is very valuable to improve the performance of the emotion recognition system. Most importantly the texture based features have done wonders to read the minute details from the face regions whenever feelings are expressed through a class of facial emotions.

## VI. CONCLUSION

To examine the class of compound emotions, this study looked at several parts of the face that make up the prominent features. Detecting facial emotion is an intrinsically challenging problem due to individual variability in facial features, as well as ethnic and cultural differences. Detecting compound emotions is even more difficult because the distinction between dominant and complimentary emotions is often thin. The class of compound emotions is frequently misclassified, including by humans. However, the findings of this work suggest that a careful investigation of a face's region of interest (ROI) can clear up the majority of these misunderstandings.

This paper analyzes the winner's approach- segmenting the face into the 8 crucial sub parts that are most effective regions when a person's facial emotion changes and detecting the image texture based features that can differentiate the facial details in a particular emotional state . The proposed method includes three major contributions :

- Spot the facial landmarks and deriving the suitable region of interest(ROI)
- Detecting the features for every sub region of face
- Analyzing the difference in the features and classifying and labeling the emotion class

The results presented in this research show that it is feasible to recognize human complex emotional state rather than just the basic emotion class which is ineffective to express the true feelings of the person and catches limited information visually.

## VII. REFERENCES


[1] P. Ekman and W.V. Friesen, Facial Action Coding System. Palo Alto, Calif.: Consulting Psychologists Press, Inc., 1978.

[2] Michael J. Black , Yaser Yacoob "Tracking and Recognizing Rigid and Non-Rigid Facial Motions using Local Parametric Models of Image Motion" Proceedings of IEEE International Conference on Computer Vision June 1995

[3] Ying-li Tian , Takeo Kanade1 and Jeffrey F. Cohn "Recognizing Lower Face Action Units for Facial Expression Analysis" IEEE Transactions on Pattern Analysis and Machine Intelligence, Feb. 2001

[4] Chirag Dalvi, Manish Rathod , Shruti Patil , Shilpa Gite AND Ketan Kotecha "A Survey of AI-Based Facial Emotion Recognition: Features, ML & DL Techniques, Age-Wise Datasets and Future Directions" IEEE access November 30, 2021

[5] A. Mehrabian and S. R. Ferris, ''Inference of attitudes from nonverbal communication in two channels,'' J. Consulting Psychol., vol. 31, no. 3, pp. 248–252, 1967, doi: 10.1037/h0024648.

[6] G. M. BEUMER, , QIAN TAO, M. BAZEN, VELDHUIS" A LANDMARK PAPER IN FACE RECOGNITION.Seventh IEEE International Conference on Automatic Face and Gesture Recognition (FG 2006), 10-12 April 2006.

[7] Grigorios Tsoumakas, Eleftherios Spyromitros-Xioufi, Jozef Vilcek "MULAN: A Java library for multi-label learning" July 2011 Journal of Machine Learning Research.

[8] P. Mohanaiah, P. Sathyanarayana, L. GuruKumar "Image Texture Feature Extraction Using GLCM

Approach" International Journal of Scientific and Research Publications, Volume 3, Issue 5, May 2013.

[9] M. M, A. M (2021) Facial geometric feature extraction based emotional expression classification using machine learning algorithms. PLoS ONE 16(2): e0247131.

[10] Priya Sahaa, Debotosh Bhattacharjeeb, Barin Kumar Dec, Mita Nasipurib "An Approach to Detect the Region of Interest of Expressive FaceImages " International Conference on Information and Communication Technologies (ICICT 2014).

View publication stats

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. 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/TIM.2022.3186355



# Multi-level Classification and Detection of Cardiac Arrhythmias with High-Resolution Superlet Transform and Deep Convolution Neural Network

Prashant Mani Tripathi, Ashish Kumar, Manjeet Kumar, *Member, IEEE*, and Rama Komaragiri, *Senior Member, IEEE*

*Abstract*-- **Atrial fibrillation and ventricular fibrillation are the two most common cardiac arrhythmia. These cardiac arrhythmias cause heart strokes and other heart complications leading increased risk of heart failure. Early and accurate cardiac arrhythmia detection is vital to preventing various heart-related diseases. Electrocardiogram (ECG) is a popular and reliable method to detect cardiac arrhythmia and heart-related diseases. However, sometimes it becomes difficult and time-consuming to interpret the ECG signal, even for a cardiac expert. This study proposes a deep learning-based method to effectively classify and detect cardiac arrhythmias (atrial fibrillation and ventricular fibrillation) using a time-frequency spectrogram of ECG records. The proposed framework utilizes superlet transform (SLT) to transform the one-dimensional (1-D) ECG signal into a two-dimensional (2-D) time-frequency spectrogram. The last layer of the pre-trained convolutional neural networks, namely, AlexNet, GoogLeNet, and DenseNet, are modified and then used to classify the ECG records into healthy heart, atrial fibrillation, and ventricular fibrillation. The proposed method is tested using ECG signals from the MIT-BIH arrhythmia database, MIT-BIH malignant database, Fantasia database, AF termination challenge database, and CU ventricular tachyarrhythmia database. The proposed method with DenseNet-201 architecture provides the best performance with an overall test accuracy of 96.2%. The proposed method reduces the system complexity as it does not require noise removal, hand-crafted features, and beat detection. Further, the proposed system automatically classifies the ECG signal into healthy heart, atrial fibrillation, and ventricular fibrillation.**

*Index Terms*-- **Atrial fibrillation, AlexNet, cardiac arrhythmia, deep learning, DenseNet, GoogLeNet, superlet transform, ventricular fibrillation.**

## 1. Introduction

Cardiovascular disease (CVD) is the primary cause of death and a key contributor to disabilities in humans. CVD is a cluster of diseases that affect the cardiovascular system. CVD includes congenital heart disease, coronary heart disease, stroke, rheumatic heart disease, and peripheral arterial disease, to name a few. CVD being severe and lethal, cause a considerable burden on the healthcare system [1]. The CVD cases are estimated to increase from 271 million to 549 million in 32 years (From 1990 to 2022) [2]. The number of casualties due to CVDs are estimated to increase from 12.1 million (1990) to 19.2 million (2022) [2]. World Health Organisation reported that CVDs alone are responsible for one-third of global deaths [3]. In developing countries, casualties due to CVDs are more than 75% [3], [4]. India has the highest loss in possible productive years of life among various countries due to CVD-related mortality of people aged 35-64 years [5]. Compared to the western population, CVDs affect the Indian population at least a decade earlier and in the most productive years of life [6]. In India, CVDs are estimated to increase from 25.7 million in 1990 to 78.541 million in 2030 [7]. In India, by 2016, the total deaths and disability-adjusted life years due to CVD increased approximately 200 % from 1990 [8]. In India, rapid progression, high mortality rate, and early age of onset of CVDs are becoming a huge concern and drawing research attention.

There are various methods to detect CVDs. Electrocardiogram (ECG) is a ubiquitous and extensively used method to detect CVDs. ECG is a non-invasive diagnostic technique capable of detecting arrhythmia, acute coronary syndrome, and intraventricular conduction disturbances [9]. ECG is a plot of the electrical activity within the heart. The sinoatrial node generates the electrical signal in the heart that passes through the atrioventricular node and Purkinje fibers to reach the ventricle [9]. Accurate and rapid interpretation of an extended duration ECG signal monitoring of a subject creates a considerable burden on cardiologists. The availability of an intelligent, self-learning and prognostic ECG analysis system can assist cardiologists in quickly interpreting an ECG signal. Such a system facilitates early and accurate detection of heart abnormalities [10], [11].

An automatic ECG signal analysis system consists of four steps: pre-processing, feature extraction, feature selection, and classification. The feature selection is a vital process that significantly influences the performance of the ECG classifier [12]. In literature, various approaches are proposed to classify

Prashant Mani Tripathi is with the Department of Electronics and Communication Engineering, Bennett University, Greater Noida, 201310, India (e-mail: pmanitripathi@gmail.com).

Ashish Kumar is with the School of Electronics Engineering, Vellore Institute of Technology, Chennai, 600127 India (e-mail: akumar.1june@gmail.com).

Manjeet Kumar is with Department of Electronics and Communication Engineering, Delhi Technological University, Bawana, Delhi, 110042 India. (e-mail: manjeetchhillar@gmail.com).

Rama Komaragiri is with Department of Electronics and Communication Engineering, Bennett University, Greater Noida, 201310, India (e-mail: rama.komaragiri@gmail.com).

© 2022 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 July 04,2022 at 05:47:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. 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/TIM.2022.3186355



ECG signals. A heartbeat recognition system using a support vector machine (SVM) is proposed in [13], [14]. SVM is widely used to classify ECG signals to detect arrhythmia. In [13], Hermite characterization and higher-order statistics extract various features. Pasolli and Melgani [14] presented an ECG signal classification method based on margin sampling, posterior probability, query by committee principle, and SVM. In [14], the performance of active learning is improved by randomly selecting the initial training set. In [15], the SVM classifier and morphological features of an ECG signal are used to detect ventricular arrhythmia. In [16], Czabanski et al. used SVM to detect atrial fibrillation in long-duration ECG signals. In [17], a long-short-term memory (LSTM) based autoencoder network with SVM is used to classify arrhythmias in the ECG signal. In [18], Raj and Ray classified ECG signals into sixteen classes using discrete orthogonal Stockwell transform and SVM based on particle swarm optimization. Martis et al. in [19] proposed a method based on probabilistic neural networks and discrete wavelet transform to classify heartbeat. In [19], the features from an ECG signal are extracted by implementing a method based on principal component analysis and linear discriminant analysis. The dimensionality of the features is reduced by using independent component analysis.

Various classifiers like recurrent neural network (RNN), random forest tree, k-nearest neighbor (KNN), and cluster analysis are employed to classify various features in an ECG signal [20]. In [21], RNN combined with the eigenvector method is used to classify the heartbeats in an ECG signal. As per the recommendations of the Association for the Advancement of Medical Instrumentation and the inter-patient scheme, Li and Zhou [22] classified the ECG signals using wavelet packet decomposition and random forest. Yeh et al. [23] suggested a cluster-based analysis method that accurately classifies normal and abnormal heartbeats. KNN is also a popular classifier for arrhythmia classification [24]. Sahu and Ray [25] proposed a KNN based method to detect myocardial infarction and achieved an accuracy of 99.82% [25]. In [26], KNN uses phase characterization and singular value features to detect ventricular arrhythmia in ECG signals. Das and Ari [27] used S-transform to compute features of an ECG signal, and the features are classified into five classes using artificial neural networks and SVM. Dokur and Olmez [28] proposed a hybrid neural network-based method to classify heartbeats. Albuquerque et al. [29] introduced an optimum path forest classifier for automatic arrhythmia detection.

Deep learning is one of the most effective methods to analyze and interpret data in many scientific fields, including healthcare [24]. Deep learning emulates the way humans acquire certain types of knowledge by screening, detecting, and predicting diseases [30]. Deep learning algorithms possess better learning capabilities than traditional machine learning algorithms. Deep learning-based networks extract independent features with fewer data pre-processing, making them suitable for analyzing high-dimensional data [30]. Deep learning classifiers based on RNN, dense neural network (DNN), LSTM, and convolutional neural network (CNN) are used to detect arrhythmias. Lai et al. [31] presented a DNN to detect ventricular arrhythmia in the short segment ECG signal. Xu et al. [32] suggested a DNN architecture to classify the raw ECG signal into different arrhythmias and normal heartbeat types. Further, in [32], DNN is used for classification and feature extraction. Hammad et al. [33] proposed a multi-tier deep learning-based model with genetic algorithm-based optimization to classify cardiac abnormalities in the ECG signal.

Machine learning approaches based on CNN are gaining importance in ECG signal analysis. Meng et al. in [34] proposed a deep CNN to automatically classify ECG signals from wearable devices. Ma et al. [35] proposed an automatic system to detect atrial fibrillation by integrating the results from CNN in SVM. Kiranyaz et al. [36] proposed a 1D-CNN-based system to classify patient-specific real-time ECG records. Acharya et al. [37] proposed a 1D-CNN model to classify the heartbeat. In [38], a method is proposed based on short-time Fourier transform (STFT) and 2-D CNN to classify arrhythmia. The STFT in [38] converts the time-domain ECG signal into a time-frequency spectrogram. Using transfer learning, Salem et al. [38] suggested classifying and identifying some ECG patterns. In [39], an RNN classifier consisting of two bidirectional LSTM layers is proposed to identify morphology/rhythm abnormalities in an ECG signal. In [40], Jin et al. proposed a method based on a dual-level attentional deep neural network that combines convolutional layers, bidirectional LSTM layers, and attention mechanisms to detect multi-label arrhythmias in ECG signals.

Essa and Xie [41] used two deep learning models to classify heartbeats into different arrhythmias. The first model (CNN and LSTM) is used to extract local features and temporal dynamics. The second model based on LSTM uses RR intervals and higher-order statistics to classify abnormal heartbeats. In [42], homeomorphically irreducible tree graph pattern-based feature extraction function and SVM are used to classify ECG records into seven classes. The overall accuracy of 92.95% is reported in [42].

This paper proposes a deep CNN to classify three cardiac conditions: healthy heart, atrial fibrillation, and ventricular fibrillation. This work transforms one-dimensional input ECG signals into a two-dimensional time-frequency spectrogram using superlet transform (SLT). In SLT, a set of wavelets with increasingly constrained bandwidth provide a time-frequency spectrogram of an electroencephalogram signal [43]. In the case of an electroencephalogram signal, SLT generates the sharpest TF resolution over the entire frequency range [43]. The ability of SLT to provide fine temporal and frequency details of a signal is used to analyze other biological signals like ECG signals. These abovementioned properties make SLT a suitable choice for ECG signal analysis.

The proposed method does not require extra noise filtering and manual feature extraction steps, thus reducing the overall circuit complexity. A database of 2100 images (700 images for each class) representing three classifications, namely, healthy heart, atrial fibrillation, and ventricular fibrillation, are used to train and validate (85% training data + 15% validation data) the classifier. A separate test database of 210 images (seventy images for each class) is used to test the classifiers. The two-dimensional time-frequency spectrograms are applied to three pre-trained classifiers: Alex Net, GoogLeNet, and DenseNet-201. The use of a pre-trained classifier reduces the training period of the classifier. The classifier automatically

© 2022 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 July 04,2022 at 05:47:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. 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/TIM.2022.3186355



extracts the relevant feature maps from the input images through the convolutional and pooling layer. Based on the extracted features, the classifier identifies the class of the test image. The key contributions of this paper are as follows.

1) A high-resolution technique using Superlet is proposed to identify cardiac arrhythmias for the first time in this field of application.
2) ECG signal contains transient oscillations events in the form of QRS complexes for a very less time. The proposed high-resolution superlet technique led to the detection of QRS complexes and quicker classification.
3) A system is proposed to identify the ECG records of a human subject into three cardiac arrhythmias, namely: healthy heart, atrial fibrillation, and ventricular fibrillation.
4) The proposed method identifies the QRS complex peaks of ECG signals to classify the ECG signals without extra noise removal steps, which lowers the hardware complexity.

The organization of this paper is as follows: Section 2 describes the mathematical background of an SLT. Section 3 provides the materials and methods employed in this work. Results and simulation are discussed in section 4. Section 5 includes discussions. Section 6 is dedicated to conclusions and future scope.

## 2. Mathematical Background of Superlet Transform

STFT and continuous wavelet transform (CWT) have some limitations. A trade-off occurs in STFT and CWT between time and frequency resolution. The STFT provides a good frequency resolution at higher frequencies, but the temporal resolution is poor. The STFT is not suitable for dealing with signals which are non-stationary or contain fast-transient features. The CWT can analyze signals at different frequencies with different resolutions. CWT provides a good temporal resolution at higher frequencies, but the frequency resolution degrades and becomes redundant. However, the frequency resolution is good at low frequencies, but temporal resolution degrades. SLT can overcome the resolution problems of STFT and CWT by combining good frequency resolution with proper temporal resolution. Unlike STFT and CWT, SLT provides a new spectral estimation that reveals the transient oscillation events of signals [43]. The SLT can provide a time-frequency spectrogram with high resolution in time and frequency localization compared to STFT and CWT. This property makes SLT suitable for analyzing transient oscillation events like ECG and electroencephalogram signals.

Superlet (SL) utilizes multiple wavelets to produce a better TF resolution and is less leaky than a single wavelet. Wavelets with fewer cycles provide a good temporal resolution with a degraded frequency resolution. On the other hand, wavelets with a large number of cycles provide a low temporal resolution at a better frequency resolution. SLT combines good temporal and appropriate frequency resolutions to provide a super-resolution [43]. SLT increases the frequency resolution using wavelets having many cycles and ample time spread parameter in the higher-frequency range while maintaining the temporal resolution at low frequencies. In TF analysis, the Morlet wavelet is popular. Eq. (1) defines the mother wavelet of the Morlet wavelet.

$$\Phi(t) = \left( e^{-jft} - e^{-\frac{1}{2}f^2} \right) e^{-\frac{1}{2}t^2} \tag{1}$$

In SL, the mother wavelet is a modified Morlet wavelet defined by eq. (2).

$$\psi_{f,p}(t) = \frac{1}{B_p\sqrt{2\pi}} e^{-\frac{t^2}{2B_p^2}} e^{j2\pi ft} \tag{2}$$

The displacement parameter or time spread parameter (in second) $B_p$ is given by eq. (3). $f$ is the central frequency, and $p$ is the number of base cycles.

$$B_p = \frac{p}{D \times f} \tag{3}$$

The parameter $B_p$ controls the time variance of the wavelet. $B_p$ has an inverse relationship with the frequency, implying that a small $B_p$ corresponds to broader frequency response. The value of $B_p$ is selected so that the wave covers $p$ cycles within $D$ standard deviations of a Gaussian envelope. $D$ and $p$ are design parameters. Normalization of the wavelet is an essential factor because it directly influences the ability of the time-frequency spectrogram to express the various properties of data [43]. The virtue of normalization is the detection of those events which are self-similar across the scale. In normalization, the events with equal peak amplitude and shape are compressed to give a similar intensity representation. In this study, to analyze an ECG signal, the wavelet is normalized to the unit integral of the modulus. For a mother wavelet, the number of cycles $p$ is fixed and selected so that the wavelet is admissible (to remove its mean).

By combining the multiple wavelets, SL provides a better TF representation. SL is a group of a finite number of wavelets with the same central frequency extending over multiple bandwidths.

$$SL_{f,m} = \left\{ \psi_{f,m} \mid p = p_1, p_2, p_3, ....., p_m \right\} \tag{4}$$

Here, $m$ represents the order of SL. $p_1, p_2, \cdots, p_m$ indicate the number of cycles in the individual wavelet in the set. The order of SL represents the number of wavelets in the set. In SL, the number of cycles in a wavelet is determined using additive or multiplicative. In this study, the number of cycles in a wavelet is determined using the multiplicative relation defined in eq. (5).

$$p_i = i \times p \tag{5}$$

Here, $i=1, 2, 3 \cdots m$. An SL of order one has one wavelet with $p_1$ cycles. SL combines multiple wavelets to detect localized time-frequency (TF) packets [43]. In the higher-frequency band, SL decreases the redundancy of representation.

The SL of a signal $s(t)$ provides a response that is the geometrical mean of the responses of individual wavelets as given in eq. (6) [43]:

$$R[\mathrm{SL}_{f,m}] = \sqrt[m]{\prod_{i=1}^{m} R[\psi_{f,p_i}]} = \sqrt[m]{\prod_{i=1}^{m} \left| \sqrt{2}\frac{1}{a}\int_{-\infty}^{\infty} s(\tau)\psi_{f,p_i}{}^{*}\left(\frac{\tau - t}{a}\right) d\tau \right|} \tag{6}$$

Here $R[\psi_{f,p_i}]$ is the response to the signal $s(t)$ provided by $i^{th}$

© 2022 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 July 04,2022 at 05:47:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. 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/TIM.2022.3186355



wavelet in the set and defined in eq. (7).

$$R[\psi_{f,p_i}] = s(t) * \psi_{f,p_i}(t) = \frac{1}{a}\int_{-\infty}^{\infty} s(\tau)\psi_{f,p_i}^{*}\left(\frac{\tau - t}{a}\right)d\tau \qquad (7)$$

The $i^{th}$ order wavelet $\Psi_{pi}(t)$ in an SL is given by eq. (8).

$$\psi_{p_i}(t) = \frac{Df}{p \cdot i \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{Dft}{p \cdot i}\right]^2} e^{j2\pi ft} \qquad (8)$$

The SL estimates the oscillation packets present in the signal at central frequency *f*. The SL transform (SLT) is like a CWT. In CWT, one uses the wavelets, while in SLT, SLs are used [43]. CWT is the SLT with SLs of order 1. However, for higher-order SLTs, the signal representation is sharp and less redundant than CWT.

**Algorithm 1:** Generation of SL scalogram

1. Input signal y(τ)
2. Selection of mother wavelet $\psi_{f,p}(t) = \frac{1}{B_p\sqrt{2\pi}} e^{-\frac{t^2}{2B_p^2}} e^{j2\pi ft}$, here $B_p = \frac{p}{D \times f}$
3. Generation of SL: $\mathrm{SL}_{f,m} = \left\{\psi_{f,p} \middle| p = p_1, p_2, p_3, ....., p_m\right\}$
4. Multiplicative SL scalogram $R_{y,p_i,m}(t,a) = \prod_{i=1}^{m}\left|\sqrt{2}\frac{1}{a}\int_{-\infty}^{\infty} y(\tau)\psi_i^{*}\left(\frac{\tau - t}{a}\right)d\tau\right|^{\frac{2}{m}}$
5. Obtained SL scalogram is the geometric mean of the modified scalogram of individual wavelets in the set

The SLT is suitable for the narrowband, and adaptive SLT (ASLT) is preferred in a wide frequency range [43]. In an ASLT, a low order is used for lower frequencies, and the order increases with frequency for higher frequencies. As the central frequency increases, then the order of SL is automatically adjusted. This ability of ASLs compensates for the increase in wavelet bandwidth with an increase in frequency. To provide an enhanced time-frequency representation in ASLT, the increase in order as a function of frequency is given by eq. (9) [43].

$$\mathrm{ASL}_f = \mathrm{SL}_{f,m}\big|_m = a(f) \qquad (9)$$

Here, *a(f)* is a uniformly increasing function of central frequency with integer values given by eq. (10).

$$a(f) = m_{\min} + (m_{\max} - m_{\min}) \times \left\lceil \frac{f - f_{\min}}{f_{\max} - f} \right\rceil \qquad (10)$$

Here, $m_{\max}$ and $m_{\min}$ are the orders of those SLs that correspond to the largest and smallest central frequencies, $f_{\max}$ and $f_{min,}$ respectively.

## 3. Material and Method

**(i) Database**

In this work, the ECG records to be classified are derived from the five standard ECG database, namely AF termination challenge database, CU ventricular tachyarrhythmia database, MIT-BIH malignant ventricular database, Fantasia database, and MIT-BIH arrhythmia database. All the databases are freely available at PhysioNet [44].

**The AF termination challenge database** is created to support the development of automatic methods to detect the spontaneous termination of atrial fibrillation [45]. Each record containing two ECG signals represents a one-minute segment of atrial fibrillation sampled at 128 samples per second in this database. This database consists of eighty records from sixty different subjects. This database consists of a learning dataset with thirty records. The remaining fifty records are grouped into two test datasets consisting of thirty and twenty unique records.

**CU ventricular tachyarrhythmia database** is a collection of thirty-five eight-minute ECG records [46]. The ECG records are recorded from those subjects who experienced an episode of ventricular fibrillation, ventricular tachycardia, and ventricular flutter. All the records are digitized at 250Hz with a 12-bit resolution over a 10 V amplitude range.

**The MIT-BIH malignant database** contains twenty-two records of thirty minutes each. This database records contain the patterns of ventricular fibrillation, ventricular tachycardia, and ventricular flutter [47]. The sampling frequency for digitization is 250 Hz with a 12-bit resolution.

**The Fantasia database** contains ECG records of the subjects watching the Fantasia movie [48]. The Fantasia database has forty records of a duration of 120 minutes each. Out of the forty ECG records, twenty records are from young (21-34 years) subjects. The remaining twenty records are from elderly subjects aged between 68 and 85 years [46]. In the Fantasia database, the records are digitized at a sampling frequency of 250 Hz.

**The MIT-BIH arrhythmia database** is a joint effort of Boston's Beth Israel hospital and MIT [49]. The MIT-BIH arrhythmia database includes forty-eight ECG records from forty-seven subjects. Each ECG signal record is a dual-channel thirty minutes long recording. A sampling frequency of 360 Hz per channel with 11-bit resolution over a 10 mV range is used for digitizing these records. Each record is annotated independently by two or more cardiac experts.

**(ii) Proposed Method**

The flow diagram of the proposed method is presented in Fig. 1. The ECG records from the AF termination challenge database, CU ventricular tachyarrhythmia database, MIT-BIH malignant ventricular database, Fantasia database, and MIT-BIH arrhythmia database are used to validate the proposed method. Different databases are selected to increase the number of ECG records in each class, namely healthy heart, atrial fibrillation, and ventricular fibrillation. AF termination challenge database provides the ECG records that reflect the episode of atrial fibrillation. CU ventricular tachyarrhythmia database and MIT-BIH malignant ventricular database provide ECG records reflecting the episodes of ventricular fibrillation. The Fantasia and MIT-BIH arrhythmia databases provide ECG signals with a healthy heart condition. The heart condition is categorized into healthy heart, atrial fibrillation, and ventricular fibrillation classes using ECG records from these databases.

All the records in this work are taken across lead-II. In the proposed method, the ECG records from different databases are first segmented into segments of ten-second duration. This

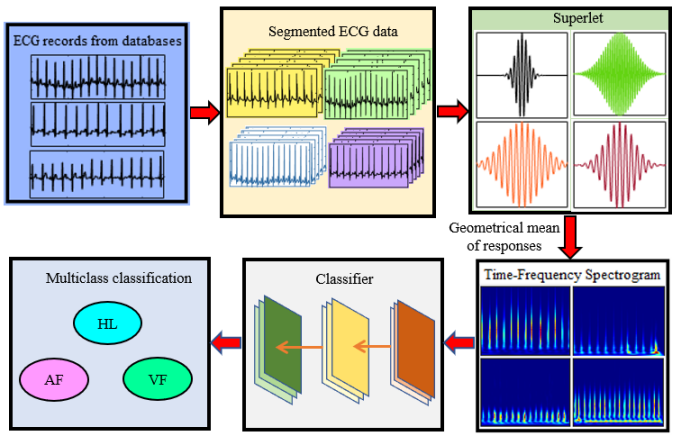
© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 04,2022 at 05:47:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. 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/TIM.2022.3186355



segmentation process increases the number of ECG records of healthy heart, atrial fibrillation, and ventricular fibrillation. In the next step, SLT is applied to segmented ten-second ECG records. Each wavelet in SL provides a time-frequency spectrogram in response to the applied ten-second ECG record. When all the responses of the wavelets in SL are combined geometrically, it provides a sharper and less redundant time-frequency spectrogram to the ECG record. Thus, SLT converts the 1-D ECG record into a 2-D image. A total of 2310 2-D images are obtained using SLT that belong to healthy heart, atrial fibrillation, and ventricular fibrillation.

Each class has 700 images, out of which 105 are used to validate the classifier. Apart from training and validation images, a set of 210 images (70 from each class) is used to test the classifier. The obtained 2-D images of ten-second duration ECG records (time-frequency spectrogram images) are used to classify the ECG records into three classes, namely: healthy heart, atrial fibrillation, and ventricular fibrillation. In this study, deep learning-based networks are utilized to classify ECG records. AlexNet, GoogLeNet, and DenseNet-201 are the three pre-trained convolutional neural networks used in this study. The last layers of the three networks (AlexNet, GoogLeNet, and DenseNet-201) are replaced with new layers to classify the ECG records into three classes healthy heart, atrial fibrillation, and ventricular fibrillation. Various parameters of the pre-trained networks, such as batch size, learn rate, validation frequency, and others, are optimized to provide the best classification results.




**HL:** Healthy heart, **AF:** atrial fibrillation, **VF:** ventricular fibrillation
**Fig.1.** Block diagram of the proposed method

### (iii) Classifier

**AlexNet** was developed by Alex Krizhevsky, Ilya Sutskever, and Geoff Hinton in 2012 [50]. AlexNet was designed to categorize 1.2 million high-resolution images into 1000 classes in the ImageNet LSVRC-2010 contest. The input to the AlexNet is an image, and the output is the class for the object in an image. The AlexNet architecture is presented in Fig. 2.

AlexNet consists of a convolution layer, rectified linear unit layer, pooling layer, and fully connected layer. Specifically, AlexNet has five convolutional layers along with three fully connected layers. The use of a rectified linear unit in AlexNet makes the learning process speedy and easy [50]. The pooling layer in the network aggregates the information in neighboring groups in the same kernel. The information in neighboring groups aggregated by pooling layers does not overlap [50], [51]. In AlexNet, an image with a resolution of 227 × 227 × 3 is applied to the first convolutional layer with 96 (11 × 11) kernels with a stride of four. The input is passed through the normalization layer, pooling layers, and the first convolutional layer. The output of the first convolution layer is then given as an input to the second convolutional layer. The second convolutional layer is equipped with 256 kernels of size 5 × 5 × 48. The third convolutional layer, with 384 kernels of size 3 × 3 × 256, connects the second and fourth convolutional layers.

The output of the third convolutional layer is applied to the fourth convolutional layer with 384 kernels of size 3 × 3 × 192. The fifth convolutional layer with 256 kernels of size 3 × 3 × 192 accepts its input from the fourth convolution layer. All the three fully connected layers of the AlexNet with 4096 neurons each perform the same operation. They perform a weighted sum operation with an added bias term. The third fully connected layer produces an output that, on passing through the softmax activation function, provides the prediction of the network.

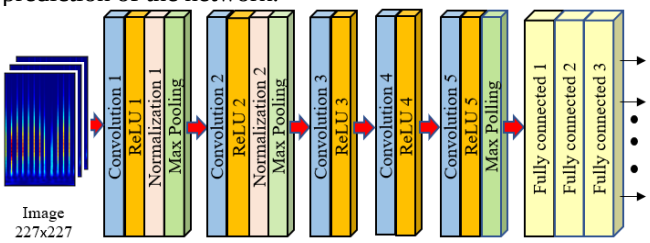



**Fig.2.** The architecture of Alex Net

**GoogLeNet** is twenty-two layers deep pre-trained CNN. GoogLeNet, based on the inception architecture, is different from AlexNet [52]. The GoogLeNet was designed so that the network could run on limited computational resources [52]. GoogLeNet uses different techniques such as global average pooling and 1 × 1 convolution to create a deeper architecture. In inception architecture, 1 × 1 convolution is used to reduce

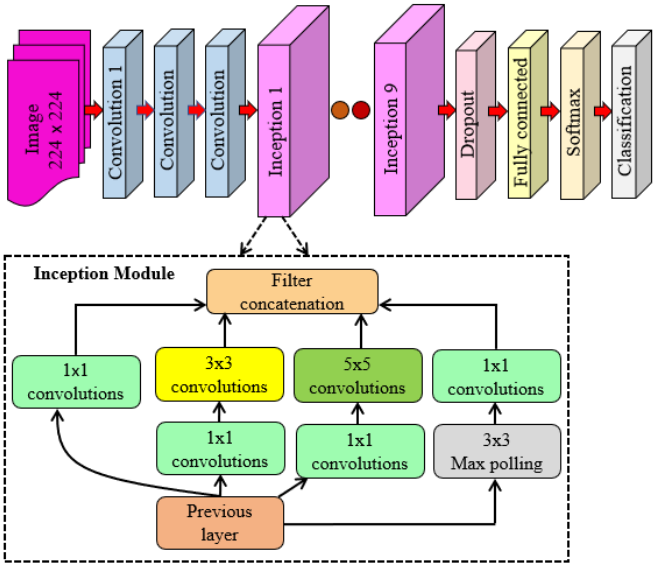



**Fig.3.** GoogLeNet architecture

© 2022 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 July 04,2022 at 05:47:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. 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/TIM.2022.3186355



the biases and weights of the networks [52]. At the end of GoogLeNet architecture, the global average pooling method takes a feature map of size 7 × 7 and averages it to 1 × 1.

The architecture of GoogLeNet is demonstrated in Fig. 3. In the GoogLeNet architecture, two auxiliary classifiers are attached to the output of the inception module. These auxiliary classifiers are used during training. They contribute weighted loss (weighted by 0.3) to the total loss of the network. But during inference time, auxiliary classifiers are discarded. The detailed structure of auxiliary classifiers is as below [52]:

- An average pooling layer with a filter size of 5 × 5 and a stride of 3.
- A 1 × 1 convolution with 128 filters is used to reduce dimension and rectify linear activation.
- A fully connected layer having 1024 outputs and rectified linear activation.
- A dropout layer with a dropout ratio of 70%
- A softmax layer as a classifier that predicts the same 1000 classes as the primary classifier GoogLeNet accepts an image of size 224 × 224.

**DenseNet** is a pre-trained CNN in which all the layers are directly connected in a forward feed fashion [53]. The direct connection between layers ensures maximum information flow [54]. Each layer obtains its input from the feature maps of all the preceding layers. The layer's feature maps act as an input for all successive layers. A DenseNet with *P* layers has *P*(*P*+1)/2 connections, whereas traditional convolutional networks only have *P* connections [53]. In DenseNet, the dense connection reduces the problems like gradient vanishing, and exploding and facilitates feature reuse [38], [54].

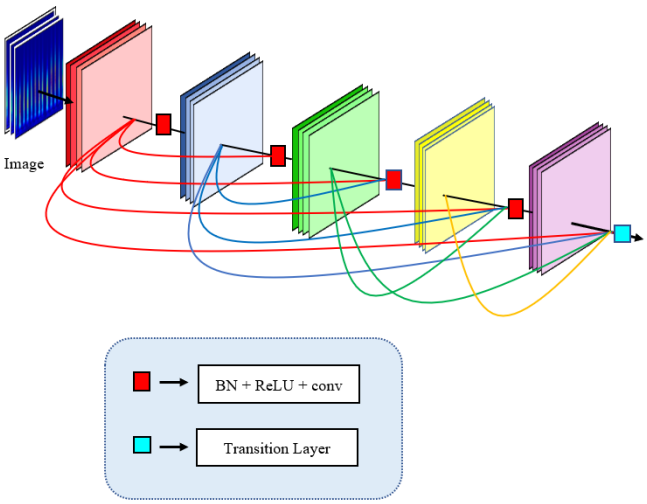



**Fig.4.** Architecture of DenseNet

DenseNet combines the property of identity mappings, deep supervision, and diversified depth [53]. The improved flow of information and gradient makes DenseNet easy to train [53]. In the proposed method, DenseNet-201, a version of DenseNet, is used to classify ECG records. The number "201" indicates the depth of the network. The pre-trained DenseNet-201 has trained to classify images into 1000 classes. The schematic of DenseNet-201 architecture is shown in fig. 4.

## 4. Simulation and Results

### (i) Validation Parameters

The performance metrics used to validate the proposed method includes accuracy (*Acc*), sensitivity (*sens*), specificity (*spec*), precision (*prec*), and F1-score (*F1-s*) described in Eq. (11) through eq. (15) [55].

$$Acc(\%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{11}$$

$$sens(\%) = \frac{TP}{TP + FN} \times 100 \tag{12}$$

$$spec(\%) = \frac{TN}{TN + FP} \times 100 \tag{13}$$

$$prec(\%) = \frac{TP}{TP + FP} \times 100 \tag{14}$$

$$F1-s(\%) = \frac{2 \times prec \times sens}{prec + sens} \times 100 \tag{15}$$

Here, *TP* (true positive) means the correct classification of positive class; *TN* (true negative) means a correct classification of negative class; *FP* (false positive) means an incorrect classification of positive class; *FN* (false negative) means an incorrect classification of negative class.

### (ii) Time-frequency representation of ECG records using SLT

The SLs have excellent capability to reveal high-frequency bursts in a single trial. The SL acting like a band-pass filter with sharp and narrow bandwidth, concentrates the frequency resolution with increasing order. The wavelets normalization in SL suppresses the redundancy of time-frequency representation. The normalization of wavelets in SL facilitates instantaneous power detection at a scale-independent frequency [56]. Due to this normalization, high-frequency oscillations (QRS complexes) can be detected in the ECG records. An MIT-BIH arrhythmia record-100 and its SLT are shown in Fig. 5(a) and 5(b), respectively. The duration of the record-100 is ten seconds.

SLT detects all the QRS complex peaks present in the record-100. Few detected QRS complex peaks of MIT-BIH record-100 are depicted in Fig. 5(c). Similarly, the MIT-BIH record-209 and its time-frequency spectrogram are shown in Fig. 6(a) and 6(b). SLT has accurately detected all the QRS complex peaks present in the MIT-BIH record-209. Few detected QRS complex peaks are shown in Fig. 6(c). Table 1 shows the variation of detection parameters with standard deviation (D) for record MIT-BIH record-100. For D = 0.33, all the thirteen QRS complex peaks of record-100 are detected successfully. For D = 1.2, the SLT precisely identified the timestamp of eight QRS complex peaks, and five QRS complex peaks were detected at different timestamps.

The variation in QRS complex peaks detection with *p* and *f* is shown in Fig. 7. The number of correctly detected QRS complex peaks presented in the ten-second MIT-BIH record-100 are different for different base cycles and central frequencies. Table 2 summarizes the QRS complex peaks detection variation with base cycle and central frequency at a

© 2022 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 July 04,2022 at 05:47:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. 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/TIM.2022.3186355



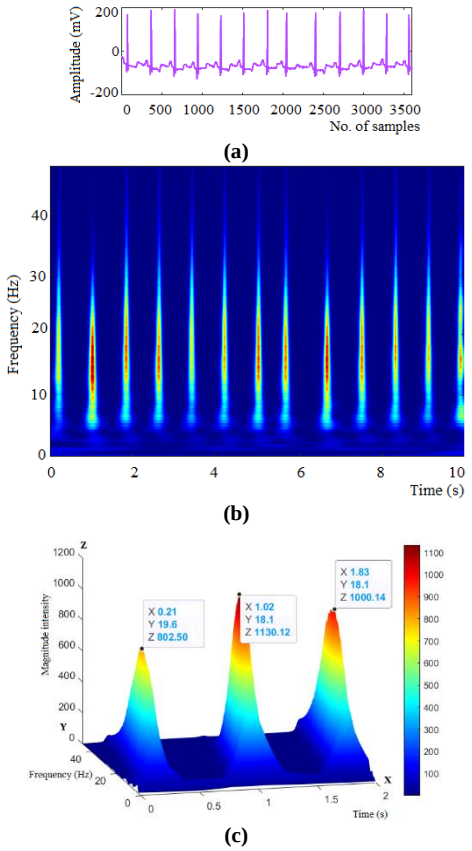



**Fig. 5. (a)** The MIT-BIH arrhythmia record-100 **(b)** Time-frequency power spectrum of record-100 **(c)** Few detected QRS complex peaks in record-100

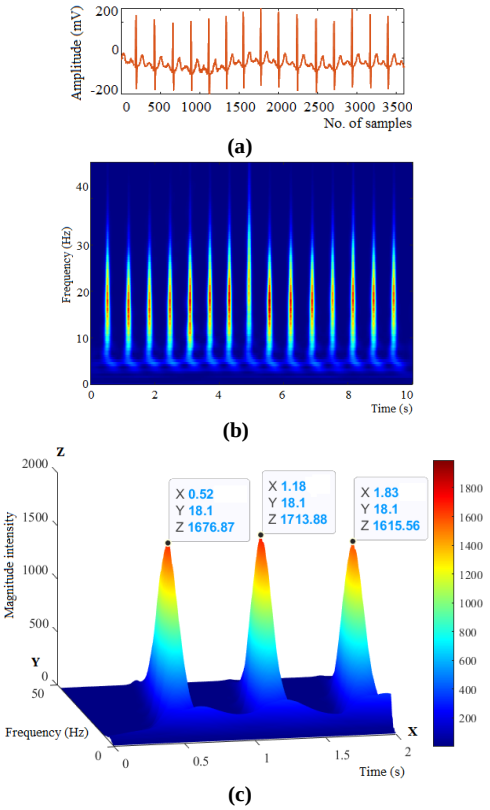



**Fig. 6. (a)** The MIT-BIH arrhythmia record-209 **(b)** Time-frequency power spectrum of record-209 **(c)** Few detected QRS complex peaks in record-209

**Table 1.** Variation of detection parameter with standard deviation (D) for MIT-BIH record-100

| Standard deviation (D) | Time spread parameter *(Bp)*(second) | No. of QRS complex peaks detected at the actual timestamp | No. of QRS complex peaks detected at a different timestamp |
|---|---|---|---|
| 1.2 | 2.5 | 8 | 5 |
| 0.5 | 3 | 10 | 3 |
| 0.33 | 4.5 | 13 | 0 |
| 0.27 | 5.5 | 12 | 1 |
| 0.2 | 7.5 | 12 | 1 |

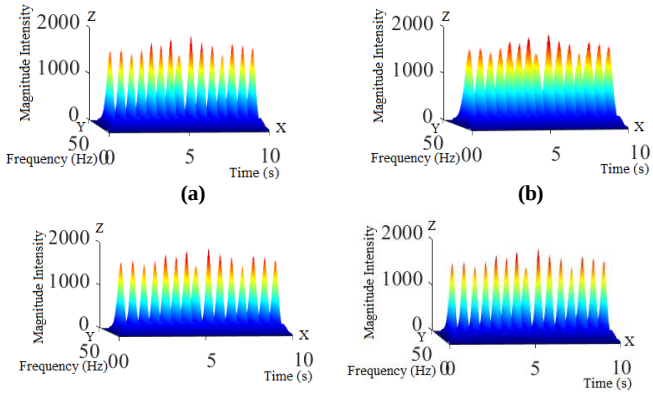



**Fig. 7.** Variation of detection parameter with base cycle and central frequency for MIT-BIH record-100 **(a)** $p$=3, $f$=1 Hz **(b)** $p$=6, $f$=2 Hz **(c)** $p$=9, $f$=3 Hz **(d)** $p$=12, $f$=4 Hz

**Table 2.** Variation of detection parameter with base cycle and central frequency for MIT-BIH record-100

| Number of base cycles (*p*) | Central frequency (*f*) (Hz) | No. of QRS complex peaks detected at the correct location | No. of QRS complex peaks detected at a different location |
|---|---|---|---|
| 3 | 1 | 13 | 0 |
| 6 | 2 | 8 | 5 |
| 9 | 3 | 11 | 2 |
| 12 | 4 | 8 | 5 |

constant D = 0.33. With $p$ = 3 and $f$ = 1 Hz, all the peaks of record-100 are detected precisely.

Figure 8 shows the efficacy of SLT in dealing with noises. The time-frequency spectrogram of a noisy ECG record containing muscle artifact is computed using STFT, CWT, and SLT. The resulting spectrograms are shown in Figures 8(a) to 8(d), respectively.

The time-frequency spectrogram of STFT and CWT is contaminated with noise in the ECG signal. The superior denoising capability of SLT removes denoises the noisy ECG signal, thus showing better denoising capabilities without any extra filters. SLT performs equally well with other noises like baseline wander, electrode motion artifact, and powerline interference. Hence, the SLT reduces the hardware complexity. The SLT provides a sharp and high-resolution spectrogram even in the presence of noises.

In the proposed method, the time-frequency spectrogram of ECG records by the SL is utilized as an image to classify ECG records. The ECG records belonging to different classes from various standard databases are subjected to SL. With the help of SL, a database of 2100 images (700 images for each class) is prepared. For classification, deep learning methods are used.

© 2022 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 July 04,2022 at 05:47:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. 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/TIM.2022.3186355



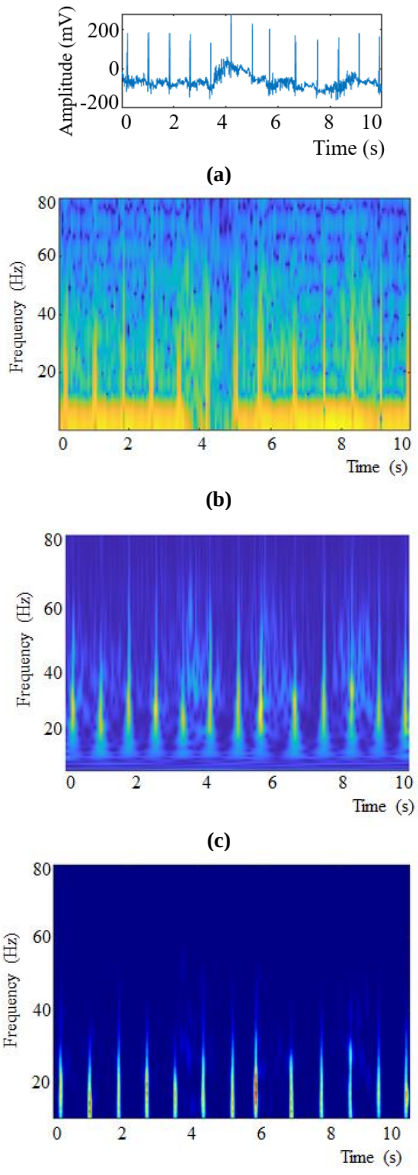



**Fig. 8. (a)** Noisy ECG record and its time-frequency representation provided by **(b)** STFT **(c)** CWT **(d)** SLT

The deep learning method extracts features from the images and efficiently recognizes the class of the image. This study considers three pre-trained classifiers for classification: AlexNet, GoogLeNet, and DenseNet-201. The classifiers are trained with 85% of data and validated with 15% of data in the database. A separate set of data with 210 images (with 70 images belonging to each class) are used for testing the performance of classifiers.

**(iii) Simulation results with classifiers**

The simulations are performed on a machine with AMD Ryzen 5 4500U CPU @ 2.38 GHz, 8 GB of RAM, and a 64-bit Windows 10 operating system.

**(a) AlexNet** with twenty-five layers and twenty-four connections is used. The images in the database are resized to 227 × 227 × 3, as per the image size requirements of AlexNet. In the AlexNet, the last three layers are changed while the rest layers are unchanged. The parameters of AlexNet need a little adoption according to new images. In this work, the root mean square propagation optimizer is used to train the transferred AlexNet. The learn rate is $10^{-5}$, with a batch size of thirty-two. The classifier provides an overall validation accuracy of 92.7% and a testing accuracy of 92.9%. The confusion matrix of AlexNet for validation and test database is represented in Fig 9(a) and 9(b), respectively.

From the confusion matrix of the test database, sixty-six images are correctly classified as atrial fibrillation, which corresponds to 31.4% of all 210 images. Two images of a healthy heart are incorrectly classified as atrial fibrillation. Similarly, three images belonging to ventricular fibrillation are incorrectly classified as atrial fibrillation—the proposed method correctly classified sixty-seven and sixty-two images for a healthy heart and ventricular fibrillation class. To obtain more insights from the AlexNet based model, the performance measures are calculated from the confusion matrix shown in Table 3. In confusion matrices, HL, AF, and VF represent healthy heart, atrial fibrillation, and ventricular fibrillation. In Table 3, AN, GN, and DN represent the AlexNet, GoogLeNet, and DenseNet-201.

The performance metrics such as *Acc*, *sens*, *spec*, *prec*, and *F1-s* are evaluated on test database and listed in Table 3. For atrial fibrillation, the *Acc*, *sens*, *spec*, *prec*, and *F1-s* are 95.7%, 94.2%, 96.4%, 92.9%, and 93.5% respectively. For healthy heart, these performance metrics are 94.7%, 95.7%, 94.2%, 89.3%, and 92.3% respectively. Similarly, for ventricular fibrillation, these performance metrics are 95.2%, 88.5%, 98.5%, 96.8%, and 92.5% respectively.

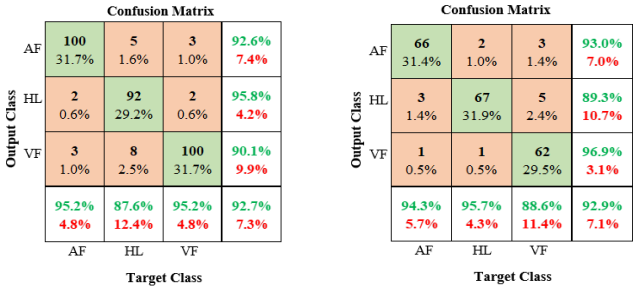



**Fig. 9. (a)** Confusion matrix of validation database **(b)** Confusion matrix of test database by AlexNet

**(b) The GoogLeNet** with 144 layers and 170 connections is used to classify the ECG records in the proposed method. The last three layers in the GoogLeNet architecture are replaced to categorize the new images. In GoogLeNet, the images are

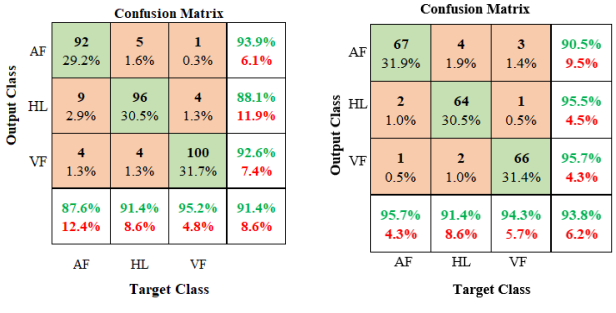



**Fig.10. (a)** Confusion matrix of validation database **(b)** Confusion matrix of test database by transferred GoogLeNet

© 2022 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 July 04,2022 at 05:47:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. 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/TIM.2022.3186355



**Table 3**. Detection summary (on test database) of models using SLT and transferred AlexNet, GoogLeNet, and DenseNet-201

| Class | Acc (%) | | | sens (%) | | | spec (%) | | | prec (%) | | | F1-s (%) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AN | GN | DN | AN | GN | DN | AN | GN | DN | AN | GN | DN | AN | GN | DN |
| **AF** | 95.7 | 95.2 | 97.1 | 94.2 | 95.7 | 100 | 96.4 | 95.0 | 95.7 | 92.9 | 90.5 | 92.1 | 93.5 | 93.0 | 95.8 |
| **HL** | 94.7 | 95.7 | 98.0 | 95.7 | 91.4 | 95.7 | 94.2 | 97.8 | 99.2 | 89.3 | 95.5 | 98.5 | 92.3 | 93.4 | 97.0 |
| **VF** | 95.2 | 96.6 | 97.1 | 88.5 | 94.2 | 92.8 | 98.5 | 97.8 | 99.2 | 96.8 | 95.6 | 98.4 | 92.5 | 94.8 | 95.6 |

resized to 224 × 224. The learn rate is $10^{-5}$, with a batch size of thirty-two. The GoogLeNet provides an overall accuracy of 91.4% on the validation database, and the overall accuracy is 93.8% for the test database. The obtained confusion matrixes from GoogLeNet on the validation and test databases are shown in Fig. 10(a) and Fig. 10(b).

From the confusion matrix plot in Fig. 10(b), sixty-seven images are correctly classified as atrial fibrillation. Sixty-four images are correctly classified as a healthy heart. Similarly, sixty-six images are correctly classified as ventricular fibrillation. Four healthy heart images are incorrectly classified as atrial fibrillation, corresponding to 1.9% of 210 test images. Three images related to ventricular fibrillation are incorrectly classified as atrial fibrillation. The performance metrics for each class on the test database are listed in Table 3.

For atrial fibrillation, the *Acc*, *sens*, *spec*, *prec*, and *F1-s* are 95.2%, 95.7%, 95.0%, 90.5%, and 93.0% respectively. For healthy heart class these performance metrics are 95.7%, 91.4%, 97.8%, 95.5% and 93.4% respectively. For atrial fibrillation, the *Acc*, *sens*, *spec*, *prec*, and *F1-s* are 95.2%, 95.7%, 95.0%, 90.5%, and 93.0% respectively. For healthy heart class these performance metrics are 95.7%, 91.4%, 97.8%, 95.5% and 93.4% respectively.

**(c) DenseNet-201** accepts an image size 224 × 224 × 3 as an input. The last three layers are updated in the pre-trained DenseNet-201 to accommodate the three categories of images. The learn rate and batch sizes are $10^{-5}$ and thirty-two, respectively. The root mean square propagation optimizer is used to train the transferred DenseNet-201. The confusion matrix plot is represented in Fig. 11(a) and Fig. 11(b) for the validation and test database. 96.2% of images are correctly classified, whereas 3.8% are wrongly classified. The classifier correctly detects seventy images of atrial fibrillation. Sixty-seven images of a healthy heart are correctly classified.

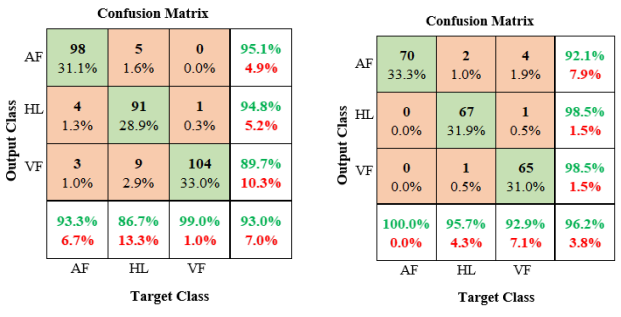



**Fig.11. (a)** Confusion matrix of validation database **(b)** Confusion matrix of test database by transferred DenseNet-201

Similarly, sixty-five images on ventricular fibrillation are correctly classified by the proposed method. Two healthy heart images are incorrectly identified as atrial fibrillation. Similarly, one healthy heart image is misclassified as ventricular fibrillation. Also, five (1+4) ventricular fibrillation images are incorrectly classified as the healthy heart (1) and atrial fibrillation (4), respectively.

## 5. Discussion

From the above results, the DenseNet-201 provides very good results. DenseNet-201 being the best among the three provides an overall accuracy of 96.2%. Other performance measures of DenseNet-201 are superior compared to AlexNet, and GoogLeNet. DenseNet-201 provides the best accuracy for each class compared to AlexNet and GoogLeNet. The AlexNet has shown the lowest performance with an overall accuracy of 92.9% on the test database. The performance of GoogLeNet is quite good compared to AlexNet, with an overall test accuracy of 93.8%.

The detection results (on the test database) of all the three models utilizing SLT and AlexNet, GoogLeNet, and DenseNet-201, respectively, are summarized in Table 3.

Although the DenseNet-201 has provided good results, its training time and structure complicacy are higher than AlexNet and GoogLeNet. The efficiency of the models used in this study is compared in Table 4.

**Table 4.** Performance comparison of different models (SLT + transferred networks)

| Model | Number of layers | Training Time (hour) | Testing Time (second) |
|---|---|---|---|
| AlexNet | 25 | 0.6 | 3.8 |
| GoogLeNet | 144 | 2.6 | 8.1 |
| DenseNet-201 | 708 | 13.4 | 10.4 |

**Table 5.** Performance comparison of various methods for atrial fibrillation detection

| Authors | Method | sens (%) | spec (%) | prec (%) | F1-s (%) |
|---|---|---|---|---|---|
| Mathunjwa et al. (2021) [57] | Recurrent plot + CNN | 86.79 | NR | 73.61 | 79.65 |
| Lai et al. (2020) [58] | Patch-based lead + CNN | 93.1 | 93.4 | NR | NR |
| Ross-Howe and Tizhoosh (2019) [59] | Spectrogram+ DenseNet + SVM | 88.38 ± 0.02 | 95.14 ± 0.01 | NR | NR |
| Fan et al. (2018) [60] | MS-CNN | 92.41 | 97.70 | 85.43 | 88.78 |
| **Proposed Method** | **SLT + DenseNet-201** | **100** | **95.7** | **92.1** | **95.8** |

In Table 4, the models (combination of SLT and

© 2022 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 July 04,2022 at 05:47:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. 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/TIM.2022.3186355



transferred networks) are compared based on the number of layers, training time, and testing time. The training time and testing time of AlexNet are the lowest, while for DenseNet-201, it is the highest.

In Table 5, the proposed method is compared with existing methods for detecting atrial fibrillation. The proposed method outperforms the compared methods. The proposed method provides the best result by correctly detecting all the images of atrial fibrillation without any noise removal or pre-processing steps. The proposed method shows a sensitivity of 100% for atrial fibrillation, which means it would not miss any subject suffering from atrial fibrillation.

The performance of the proposed method for ventricular fibrillation is compared in Table 6. In Table 7, the proposed method is compared with other arrhythmia detection methods. The proposed method is also evaluated with STFT and CWT. The proposed method with STFT provides an overall accuracy (on the test database) of 81.4%, 83.8%, and 91.0% for AlexNet, GoogLeNet, and DenseNet-201, respectively. Similarly, with the CWT, the results on the test database are 92.4%, 92.2%, and 94.8% for AlexNet, GoogLeNet, and DenseNet-201, respectively. The results show that the proposed method with SLT outperforms STFT and CWT.

**Table 6.** Performance comparison of various methods for ventricular fibrillation detection

| Authors | Method | Acc (%) | sens (%) | spec (%) | F1-s (%) |
|---|---|---|---|---|---|
| Lee et al. (2013) [15] | Morphological features + SVM | NA | 95.12 | 96.67 | NA |
| Tripathy et al. (2018) [61] | DTFT + LS-SVM | 89.81 | 86.38 | 93.97 | NA |
| Mohanty et al. (2018) [62] | TF and statistical features + cubic SVM | 92.23 | 79.43 | 81.44 | NA |
| **Proposed Method** | **SLT + DenseNet-201** | **97.1** | **92.8** | **99.2** | **95.6** |

***DTFT:*** digital Taylor-Fourier transform, ***LS-SVM:*** least square support vector machine

**Table 7**. Performance comparison of the proposed method with other existing methods

| Authors | Method | Number of classes | Acc (%) |
|---|---|---|---|
| Llamedo and Martinez (2012) [63] | LDC + EMCA | 3 | 95.42 |
| Essa and Xie (2021) [41] | Median filter + CNN-LSTM + RRHOS-LSTM | 4 | 95.81 |
| Baygin et al. (2021) [42] | HIT + SVM | 7 | 92.95 |
| **Proposed method** | **STFT + DenseNet-201** | **3** | **91.0** |
| **Proposed method** | **CWT + DenseNet-201** | **3** | **94.8** |
| **Proposed method** | **SLT + DenseNet-201** | **3** | **96.2** |

***LDC***: Linear discriminant classifier, **EMCA:** expectation-maximization clustering algorithm**,** ***RRHOS*****:** RR intervals and higher-order statistics, ***HIT:*** homeomorphically irreducible tree

## 6. Conclusions and Future Scope

This paper proposes a method to detect healthy heart, atrial fibrillation, and ventricular fibrillation from ECG records. In the proposed method, the combination of SLT and deep neural network architectures successfully detect cardiac arrhythmia. An SLT with high resolution provides the time-frequency spectrogram of the ECG records. Results show that the SLT can correctly detect biological events. The SLT has successfully detected all the QRS complex peaks of the ECG records from the database. In the study, the SLT converts one-dimensional ECG records into two-dimensional RGB images by providing the time-frequency spectrogram. These images are fed to pre-trained convolutional neural network architectures, including AlexNet, GoogLeNet, and DenseNet-201. The use of a pre-trained network provides better results with less training. The last three layers of the networks are modified to meet the need for classification. The DenseNet-201 architecture among all three networks produces the best performance with an overall accuracy of 96.2%. The results show that the DenseNet-201 successfully detected all the images under atrial fibrillation categories. In other cases, the performance of DenseNet is reasonable compared to AlexNet and GoogLeNet. In the proposed method, denoising of ECG signals is not required, which reduces the computational requirements.

A large database can result in a very good classification accuracy of the proposed method. The availability of a large database is a limitation of the proposed method. Also, the large database increases the training time of the pre-trained classifiers. Each addition of new data to the new database requires retraining. Another limitation of the proposed method is selecting the order and time-spread parameters in SL, affecting classification accuracy.

With more training data, the proposed method may be implemented for clinical practice and by classifying more types of arrhythmias. Also, the combination of SL and other advanced pre-trained networks such as Xception, VGG, and ResNet-50 can be used to test the proposed SL based model.

## Compliance with Ethical Standards

**Conflict of Interest:** All Authors of this work declare no conflict of interest.

**Ethical approval:** This article does not contain any studies with animals or human participants performed by any authors.

**References**

[1] S. Li, F. Li, S. Tang and F. Luo, "Heart Sounds Classification Based on Feature Fusion Using Lightweight Neural Networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-9, 2021, Art no. 4007009, doi: 10.1109/TIM.2021.3109389.

[2] Gregory A. Roth et al., "Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study," *Journal of the American College of Cardiology*, vol. 76, no. 25, pp. 2982-3021, Dec. 2020. doi: 10.1016/J.JACC.2020.11.010

[3] https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1 (accessed Jan. 31, 2022).

[4] J. Huang, B. Chen, B. Yao and W. He, "ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network," *IEEE Access*, vol. 7, pp. 92871-92880, 2019, doi: 10.1109/ACCESS.2019.2928017.

[5] K. Srinath Reddy, Bela Shah, Cherian Varghese, and Anbumani Ramadoss, "Responding to the threat of chronic diseases in India," *The Lancet*, vol. 366, no. 9498, pp. 1744-1749, Nov. 2005.

[6] Dorairaj Prabhakaran, Panniyammakal Jeemon, and Ambuj Roy,

© 2022 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 July 04,2022 at 05:47:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. 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/TIM.2022.3186355



"Cardiovascular diseases in India: current epidemiology and future directions," *Circulation*, vol. 133, no. 16, pp. 1605-1620, Apr. 2016

[7] D. Prabhakaran et al., "India State-Level Disease Burden Initiative CVD Collaborators. The changing patterns of cardiovascular diseases and their risk factors in the states of India: the Global Burden of Disease Study 1990–2016," *Lancet Glob Health*, vol. 6, no. 12, pp. e1339-e1351, Dec. 2018.

[8] Indian Council of Medical Research, Public Health Foundation of India, Institute for Health Metrics and Evaluation. GBD India compares data visualization. http://vizhub.healthdata.org/gbdcompare/ India (accessed Jan. 31, 2022).

[9] Prashant Mani Tripathi, Ashish Kumar, Rama Komaragiri, and Manjeet Kumar, "A Review on Computational Methods for Denoising and Detecting ECG Signals to Detect Cardiovascular Diseases," *Archives of Computational Methods in Engineering*, pp. 1-40, 2021, doi**:**10.1007/s11831-021-09642-2.

[10] A. Kumar, R. Ranganatham, M. Kumar and R. Komaragiri, "Hardware Emulation of a Biorthogonal Wavelet Transform-Based Heart Rate Monitoring Device," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 5271-5281, 15 Feb.15, 2021, doi: 10.1109/JSEN.2020.3034742.

[11] Prashant Mani Tripathi, Ashish Kumar, Rama Komaragiri, and Manjeet Kumar, "Watermarking of ECG signals compressed using Fourier decomposition method," *Multimedia Tools and Applications* 11492, 2021, doi**:** 10.1007/s11042-021-11492-w.

[12] Y. Li, Z. Zhang, F. Zhou, Y. Xing, J. Li, and C. Liu, "Multi-Label Classification of Arrhythmia for Long-Term Electrocardiogram Signals With Feature Learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-11, 2021, Art no. 2512611, doi: 10.1109/TIM.2021.3077667.

[13] S. Osowski, L. T. Hoai, and T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 4, pp. 582-589, Apr. 2004, doi: 10.1109/TBME.2004.824138.

[14] E. Pasolli and F. Melgani, "Active Learning Methods for Electrocardiographic Signal Classification," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 6, pp. 1405-1416, Nov. 2010, doi: 10.1109/TITB.2010.2048922.

[15] S. H. Lee, H. Ko and Y. Yoon, "Classification of Ventricular arrhythmia using a support vector machine based on morphological features," 2013 35th Annual International Conference of the *IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5785-5788, 2013, doi: 10.1109/EMBC.2013.6610866.

[16] Robert Czabanski et al., "Detection of atrial fibrillation episodes in long-term heart rhythm signals using a support vector machine." *Sensors*, vol. 20, no. 3, pp. 765, 30 Jan. 2020.

[17] B. Hou, J. Yang, P. Wang and R. Yan, "LSTM-Based Auto-Encoder Model for ECG Arrhythmias Classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1232-1240, Apr. 2020, doi: 10.1109/TIM.2019.2910342.

[18] S. Raj and K. C. Ray, "ECG Signal Analysis Using DCT-Based DOST and PSO Optimized SVM," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 3, pp. 470-478, March 2017.

[19] Roshan Joy Martis, U. Rajendra Acharya, and Lim Choo Min, "ECG beat classification using PCA, LDA, ICA, and discrete wavelet transform," *Biomedical Signal Processing and Control*, vol.8, no. 5 pp. 437-448, Sep.2013, doi:10.1016/J.BSPC.2013.01.005.

[20] Özal Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," *Computers in Biology and Medicine*, vol. 96, pp. 189-202, May 2018.

[21] Elif Derya Übeyli, "Combining recurrent neural networks with eigenvector methods for classification of ECG beats," *Digital Signal Processing*, vol. 19, no. 2, pp. 320-329, March 2009.

[22] Taiyong Li, and Min Zhou, "ECG classification using wavelet packet entropy and random forests," *Entropy*, vol. 18, no. 8, pp. 285, Aug. 2016, doi:10.3390/E18080285.

[23] Yun-Chi Yeh, Che Wun Chiou, and Hong-Jhih Lin, "Analyzing ECG for cardiac arrhythmia using cluster analysis," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1000-1010, Jan. 2012.

[24] Edward B. Panganiban, Arnold C. Paglinawan, Wen Yaw Chung, and Gilbert Lance S. Paa, "ECG diagnostic support system (EDSS): A deep learning neural network-based classification system for detecting ECG abnormal rhythms from low-powered wearable biosensors," *Sensing and Bio-Sensing Research*, vol. 31, pp. 100398, Feb. 2021, doi:10.1016/J.SBSR.2021.100398.

[25] G. Sahu and K. C. Ray, "An Efficient Method for Detection and Localization of Myocardial Infarction," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-12, 2022, Art no. 4001312, doi: 10.1109/TIM.2021.3132833.

[26] N. Sinha and A. Das, "Discrimination of Life-Threatening Arrhythmias Using Singular Value, Harmonic Phase Distribution, and Dynamic Time Warping of ECG Signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-8, 2021, Art no. 2504508, doi: 10.1109/TIM.2020.3045190.

[27] Manab Kumar Das and Samit Ari, "Electrocardiogram beat classification using S-transform based feature set," *Journal of Mechanics in Medicine and Biology*, vol. 14, no. 5, 2014.

[28] Zümray Dokur and Tamer Ölmez, "ECG beat classification by a novel hybrid neural network," *Computer Methods and Programs in Biomedicine*, vol. 66, no. 2-3, pp. 167-181, Sep 2001.

[29] Victor Hugo C. De Albuquerque et al.,"Robust automated cardiac arrhythmia detection in ECG beat signals," *Neural Computing and Applications*, vol. 29, no. 3, pp. 679-693, July 2018.

[30] Jin-Yu Sun, Hui Shen, Qiang Qu, Wei Sun, and Xiang-Qing Kong, "The application of deep learning in electrocardiogram: Where we came from and where we should go?," *International Journal of Cardiology*, vol. 337, pp. 71-78, Aug. 2021.

[31] D. Lai, X. Fan, Y. Zhang and W. Chen, "Intelligent and Efficient Detection of Life-Threatening Ventricular Arrhythmias in Short Segments of Surface ECG Signals," *IEEE Sensors Journal*, vol. 21, no. 13, pp. 14110-14120, 1 July1, 2021, doi: 10.1109/ JSEN. 2020.3031597.

[32] S. S. Xu, M. -W. Mak and C. -C. Cheung, "Towards End-to-End ECG Classification With Raw Signal Extraction and Deep Neural Networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1574-1584, July 2019.

[33] M. Hammad, A. M. Iliyasu, A. Subasi, E. S. L. Ho and A. A. A. El-Latif, "A Multitier Deep Learning Model for Arrhythmia Detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-9, 2021, Art no. 2502809, doi: 10.1109/TIM.2020.3033072.

[34] L. Meng, K. Ge, Y. Song, D. Yang and Z. Lin, "Long-term Wearable Electrocardiogram Signal Monitoring and Analysis Based on Convolutional Neural Network," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-11, 2021, Art no. 2507711, doi: 10.1109/TIM.2021.3072144.

[35] C. Ma, S. Wei, T. Chen, J. Zhong, Z. Liu and C. Liu, "Integration of Results From Convolutional Neural Network in a Support Vector Machine for the Detection of Atrial Fibrillation," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-10, 2021, Art no. 2504610, doi: 10.1109/TIM.2020.3044718.

[36] S. Kiranyaz, T. Ince and M. Gabbouj, "Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664-675, March 2016, doi: 10.1109/TBME.2015.2468589.

[37] U. Rajendra Acharya et al., "A deep convolutional neural network model to classify heartbeats." *Computers in Biology and Medicine*, vol. 89, pp. 389-396, Oct. 2017.

[38] J. Huang, B. Chen, B. Yao and W. He, "ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network," *IEEE Access*, vol. 7, pp. 92871-92880, 2019.

[38] M. Salem, S. Taheri and J. Yuan, "ECG Arrhythmia Classification Using Transfer Learning from 2- Dimensional Deep CNN Features," 2018 *IEEE Biomedical Circuits and Systems Conference* (BioCAS), 2018, pp. 1-4, doi: 10.1109/BIOCAS.2018.8584808.

[39] Ahmed Mostayed, Junye Luo, Xingliang Shu, and William Wee, "Classification of 12-lead ECG signals with Bi-directional LSTM network," *arXiv preprint arXiv:1811.02090*, 2018

[40] Y. Jin et al., "A Novel Interpretable Method Based on Dual-Level Attentional Deep Neural Network for Actual Multilabel Arrhythmia Detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-11, 2022, Art no. 2500311.

[41] E. Essa and X. Xie, "An Ensemble of Deep Learning-Based Multi-Model for ECG Heartbeats Arrhythmia Classification," *IEEE Access*, vol. 9, pp. 103452-103464, 2021, doi: 10.1109/ACCESS.2021.3098986.

[42] Baygin, Turker Tuncer, Sengul Dogan, Ru-San Tan, and U. Rajendra Acharya, "Automated arrhythmia detection with homeomorphically irreducible tree technique using more than 10,000 individual subject ECG records," *Information Sciences*, vol. 575, pp. 323-337, Oct. 2021, doi: 10.1016/J.INS.2021.06.022.

© 2022 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 July 04,2022 at 05:47:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. 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/TIM.2022.3186355



[43] Vasile V. Moca, Harald Bârzan, Adriana Nagy-Dăbâcan, and Raul C. Mureșan,"Time-frequency super-resolution with superlets." *Nature communications*, vol. 12, no. 1, pp. 1-18, Article number: 337, doi: 10.1038/s41467-020-20539-9

[44] Ary L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215-e220, 2000.

[45] G. E. Moody, "Spontaneous termination of atrial fibrillation: a challenge from physionet and computers in cardiology 2004," *Computers in Cardiology*, 2004, pp. 101-104, doi: 10.1109 /CIC. 20 0 4.1442881.

[46] Nolle, F. M., F. K. Badura, J. M. Catlett, R. W. Bowser, and M. H. Sketch, "CREI-GARD, a new concept in computerized arrhythmia monitoring systems," *Computers in Cardiology*, vol. 13, pp. 515-518, 1986.

[47] Scott David Greenwald, "The development and analysis of a ventricular fibrillation detector," Ph.D. diss., Massachusetts Institute of Technology, 1986.

[48] N. Iyengar, C. K. Peng, Raymond Morin, Ary L. Goldberger, and Lewis A. Lipsitz, "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 271, no. 4, pp. R1078-R1084, 1996.

[49] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45-50, May-June 2001, doi: 10.1109/51.932724.

[50] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems* 25 (2012): 1097-1105.

[51] S. Sun et al., "Fault Diagnosis of Conventional Circuit Breaker Contact System Based on Time–Frequency Analysis and Improved AlexNet," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-12, 2021, Art no. 3508512.

[52] C. Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-9, 2015.

[53] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger, "Densely connected convolutional networks," In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700-4708, 2017.

[54] T. Li, W. Jiao, L. -N. Wang and G. Zhong, "Automatic DenseNet Sparsification," *IEEE Access*, vol. 8, pp. 62561-62571, 2020.

[55] A. Kumar, V. K. Mehla, H. Tomar, M. Kumar and R. Komaragiri, "Classification of Normal and Abnormal ECG signals using Support Vector Machine and Fourier Decomposition Method," 2020 *IEEE International Symposium on Smart Electronic Systems (iSES)*, 2020, pp. 161-166, doi: 10.1109 /iSES5 0453 .20 20 .00044.

[56] Jonathan M. Lilly, "Element analysis: a wavelet-based method for analysing time-localized events in noisy time series," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 473, no. 2200, Apr. 2017.

[57] Bhekumuzi M. Mathunjwa, Yin-Tsong Lin, Chien-Hung Lin, Maysam F. Abbod, and Jiann-Shing Shieh, "ECG arrhythmia classification by using a recurrence plot and convolutional neural network," *Biomedical Signal Processing and Control*, vol. 64, pp. 102262, Feb. 2021, doi: 10.1016/J.BSPC.2020.102262.

[58] D. Lai, Y. Bu, Y. Su, X. Zhang and C. -S. Ma, "Non-Standardized Patch-Based ECG Lead Together With Deep Learning Based Algorithm for Automatic Screening of Atrial Fibrillation," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 6, pp. 1569-1578, June 2020, doi: 10.1109/JBHI.2020.2980454.

[59] S. Ross-Howe and H. R. Tizhoosh, "Atrial Fibrillation Detection Using Deep Features and Convolutional Networks," 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 2019, pp. 1-4, doi: 10.1109/BHI.2019.8834583.

[60] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun and Y. Li, "Multiscaled Fusion of Deep Convolutional Neural Networks for Screening Atrial Fibrillation From Single Lead Short ECG Recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 6, pp. 1744-1753, Nov. 2018, doi: 10.1109/JBHI.2018.2858789.

[61] M. Llamedo and J. P. Martinez, "An Automatic Patient-Adapted ECG Heartbeat Classifier Allowing Expert Assistance," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 8, pp. 2312-2320, Aug. 2012.

[62] Monalisa Mohanty, Santanu Sahoo, Pradyut Biswal, and Sukanta Sabut, "Efficient classification of ventricular arrhythmias using feature selection and C4. 5 classifier," *Biomedical Signal Processing and Control*, vol. 44, pp. 200-208, July 2018.

[63] Rajesh et al., "Detection of life-threatening ventricular arrhythmia using digital Taylor Fourier transform." Frontiers in physiology, vol. 9, pp. 722, June 2018, doi: 10.3389/FPHYS.2018.00722.

© 2022 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 July 04,2022 at 05:47:05 UTC from IEEE Xplore. Restrictions apply.

2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) | 978-1-6654-1120-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAECT54875.2022.9807840

# Optimal Reactive Power Dispatch Using Giza Pyramids Construction Algorithm

Ganesh Kumar Jaiswal
*Electrical Engineering*
*Delhi Technological University*
Delhi, India
ganesh10m213@gmail.com

Uma Nangia
MIEEE, MIE, LMISTE
*Electrical Engineering*
*Delhi Technological University*
*Delhi, India*
umanangia@dce.ac.in

N. K. Jain
MIEEE, MIE, LMISTE
*Electrical Engineering*
*Delhi Technological University*
Delhi, India
vnarender84@yahoo.com

***Abstract*— The proper monitoring and management of reactive power is critical for power system's stability, reliability and security. The challenge of reactive power scheduling, which is a mixed integer and nonlinear problem, is critical for improving system efficiency and keeping voltages acceptable limits. While optimizing a specific objective function, it satisfies all equality and inequality constraints. This study used a new ancient-inspired meta-heuristic technique called the Giza Pyramids Construction (GPC) Optimization algorithm to solve the Optimal Reactive Power Dispatch (ORPD) problem. The proposed single-objective deterministic ORPD approach is validated and compared to previous reported research works for power loss minimization utilizing two cases of IEEE 30-bus and IEEE-57 bus systems, respectively.**

***Keywords*— *Giza Pyramids Construction (GPC), Success History Based Adaptive Differential Evolution (SHADE-EC), Jellyfish Search (JS), Artificial Electric Field Algorithm (AEFA)***

## I. INTRODUCTION

ORPD (optimal reactive power dispatch), which is a subproblem of Optimal Power Flow(OPF) [1, 2], is a well-known engineering optimization problem that tries to manage power systems in a stable and effective manner. ORPD is an basic tool in power system management and controlling process because of the significant impact on security, reliability, and economic operations. ORPD is modelled numerically as a nonlinear problem with well-defined objectives and constraints. One of the main objectives of an ORPD problem is to minimise the objective functions by locating control variables within an acceptable operating range. Voltage of generators, transformer tap ratios and reactive power output of compensating devices are the control variables. Unlike the OPF problem, the ORPD problem considers active power generation as a set of known variables. As a result, because the ORPD problem has fewer control variables to optimise, it is easier to solve than the OPF problem. ORPD's main goals are to minimise transmission losses [3], improve voltage profile [4], lower transmission costs [5], and improve power system voltage stability [6].

In response to the challenges of choosing the most appropriate control variables, researchers have proposed various optimization methods as well as constraint handling strategies for the ORPD problem. The development and implementation of meta-heuristic methods in the solution of the ORPD problem has recently yielded better results as compared to stochastic search optimization methods such as Genetic Algorithm (GA) [7], Evolutionary Programming (EP) [8], Tabu search (TS) [9] and Differential Evolution (DE) [10] algorithm. Meta-heuristic methods adopt specific mechanisms for escaping local optima and therefore have been used widely in the last 20 years to address the ORPD problem. Particle Swarm Optimization (PSO) [11,12], Harmony Search Algorithm (HSA) [13], Moth-Flame Optimization (MFO) [14], Gravitational Search Algorithm (GSA) [15], Artificial Bee Colony (ABC) [16], Social Spider Algorithm (SSA) [17] are a few examples of these methods. However, there is no single strategy is available, that can resolve all optimization issues. As a result, the ORPD problem can still be addressed by applying a newly discovered optimization technique. This research proposes using the GPC optimization algorithm, a novel ancient-inspired meta-heuristic method, to solve the ORPD problem.

Significant contributions of this study are:

- Solving the problem of ORPD as a single objective optimization problem using a new application for the Giza Pyramids Construction (GPC) optimization technique.
- The efficacy of the Giza Pyramids optimization (GPC) based ORPD method is compared to the ORPD problem-solving methodologies.

The following is how the remaining of the manuscript is organised: Section II shows the mathematical construction of ORPD. The applied GPC algorithm for solving ORPD problem is described in Section III. The results are summarised and examined in Section IV. The conclusion reached as a result of this investigation is presented in Section V followed by the references.

## II. PROBLEM FORMULATION

### A. General form of ORPD

Minimize:

$$f\,(U,V) \tag{1}$$

Subject to:

$$g\,(U,V) \le 0 \tag{2}$$

&

$$h\,(U,V) = 0 \tag{3}$$

Where $f$ is the fitness function.
Vector U refers to the control variables.
Vector V denotes the state variables.
ORPD's objective function is represented by $f\,(u,v)$.
$g\,(U,V)$: a set of constraints on inequality
$h\,(U,V)$: a set of constraints on equality

- *Control Variables*

The control variables are expressed in the form of vector as:

$$U = [V_{G1} \dots\dots V_{GNG}, Q_{C1} \dots\dots Q_{CNC}, T_1 \dots\dots T_{NT}] \tag{4}$$

Where $V_{G1} \dots\dots V_{GNG}$ is the PV bus (generator bus) voltage magnitude.

$Q_{C1} \dots\dots Q_{CNC}$ is shunt compensation at the buses.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:42:28 UTC from IEEE Xplore. Restrictions apply.

$T_1 \dots \dots . . T_{NT}$ is the position of the transformer taps.
$NG$: Quantity of generators
$NC$: Quantity of VAR compensators
$NT$: Quantity of transformers taps

- *Stable Variables*

In vector form, the state variables can be given as:

$$V = [V_{L1} \dots \dots, V_{LNL}, Q_{G1} \dots \dots Q_{GNG}, S_{L1} \dots \dots S_{LNL}] \quad (5)$$

Where $V_{L1} \dots \dots, V_{LNL}$ represents magnitude of voltage at load bus.
$Q_{G1} \dots \dots Q_{GNG}$ represents reactive power generation at PV bus.$S_{L1} \dots \dots S_{LNL}$ represents transmission line apparent power flow.

*B. Objective Function*

The main purpose of the ORPD problem is to find the best settings for various control parameters to reduce power losses during power system network control and operation. The following objective function for power loss minimization is used in this paper to address the single objective ORPD problem.

$$F_1 = \min \; P_{Loss} = \sum_{i=1}^{NTL} G_{ij} \; (V_i^2 + V_j^2 - 2V_i \, V_j \, \text{Cos} \, \delta_{ij}) \quad (6)$$

Where $V_i$ & $V_j$ is the voltage at ith & jth bus respectively, $G_{ij}$ is the conductance of line i-j, δij is the difference in the phase between ith & jth bus. NTL refers to the total number of transmission line.

*C. Constraints*

The ORPD problem's equality and inequality constraints are listed below.

*Equality Constraints:*

$$P_i = P_{Gi} - P_{Di} - P_{Loss}$$
$$Q_i = Q_{Gi} - Q_{Di} - Q_{Loss}$$

Where $P_{Gi}$ is active and $Q_{Gi}$ $is$ reactive power generation at ith bus. $P_{Di}$ is active and $Q_{Di}$ reactive power demand at ith bus. $P_{Loss}$ is active and $Q_{Loss}$ reactive power losses.

*Inequality Constraints :*

Generator Constraints:

$$V_{Gi}^{min} \le V_{Gi} \le V_{Gi}^{max}, \quad \forall i \quad i \in NG$$
$$P_{Gi}^{min} \le P_{Gi} \le P_{Gi}^{max}, \quad \forall i \quad i \in NG$$
$$Q_{Gi}^{min} \le Q_{Gi} \le Q_{Gi}^{max}, \quad \forall i \quad i \in NG$$

Where $V_{Gi}$ is voltage, $P_{Gi}$ is active power generation and $Q_{Gi}$ is reactive power generation at generator bus.
Transformer Constraints:

$$T_i^{min} \le T_i \le T_i^{max}, \quad \forall i \quad i \in NT$$

Where $T_i$ is tap setting of transformers.
Shunt VAR Compensator Constraints:

$$Q_{Ci}^{min} \le Q_{Ci} \le Q_{Ci}^{max}, \quad \forall i \quad i \in NC$$

Where $Q_{Ci}$ is shunt var compensations.
Security Constraints:

$$V_{Li}^{min} \le V_{Li} \le V_{Li}^{max}, \quad \forall i \quad i \in NL$$
$$S_{Li} \le S_{Li}^{max}, \quad \forall i \quad i \in NTL$$

Where $V_{Li}$ depicts load bus voltage and $S_{Li}$ denotes transmission line loading.

## III. Overview of Giza Pyramids Construction Algorithm

GPC [18] is a newly designed meta-heuristic optimization method based on ancient-inspired ideology. Restrictions and a lack of hardware and software capacities in ancient times led to the development of some form of optimization capabilities in ancient civilization. The construction of pyramids was optimized in ancient Egyptian society, which considerably decreased construction costs and provided optimized workforce management. The stones were obtained from the surrounding area, and ramps were employed to transport the stone blocks from their initial positions to the installation location. The initial velocity, ramp gradient and friction force all such factors have an effect on stone block movement. The workers themselves are constantly altering places in trying to gain control of the stone block, and the ones who do their jobs the best are known as Pharaoh's special agents. The solutions are determined from the resultant of the stone block and worker's position. During the construction process, some workers are replaced with efficient workers with a fixed probability in order to equalize the strength of the works for moving a stone block.

Equation (7) determines the stone block displacement on the ramp from its previous position.

$$d = \frac{v_0^2}{2g(\sin \theta + \mu_k \cos \theta)} \quad (7)$$

Where, $v_0$ denotes the stone block's initial velocity, $g$ denotes the earth's gravity, $\theta$ represents angle formed by the ramp with the horizontal and $\mu_k$ denotes the kinetic friction coefficient. The new position of the workers carrying the stone blocks is determined by Equation (8). For the workers the friction effect is not considered.

$$x = \frac{v_0^2}{2g \sin \theta} \quad (8)$$

New position (new solution) that is obtained from the resultant of equations (7) and (8) is given by equation (9)

$$\vec{p} = (\vec{p}_i + d) \times x \, \vec{\epsilon}_i \quad (9)$$

Where $\vec{p}_i$ represents the present position, $d$ illustrates the displacement of a stone block, $x$ represents the magnitude of workers motion and $\vec{\epsilon}_i$ is an arbitrary vector with a Normal, Uniform, or Lévy distribution.

By default, the algorithm performs the substitution operation with a 50% probability. If the initial solutions are $\emptyset = (\varphi_1, \varphi_2, \dots \dots, \varphi_n)$ and the produced solutions employing equation (8) are $\boldsymbol{\psi} = (\psi_1, \psi_2, \dots \dots, \psi_n)$, then the new solutions on substitution will be $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots \dots, \xi_n)$. The substitution condition is described in Equation (10) as

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:42:28 UTC from IEEE Xplore. Restrictions apply.

$$\xi_k = \begin{cases} \psi_{k,} & if\ rand[0,1] \le 0.5 \\ \varphi_k, & otherwise \end{cases} \quad (10)$$

Figure 1 depicts the flow chart for the Giza Pyramids Optimization method.

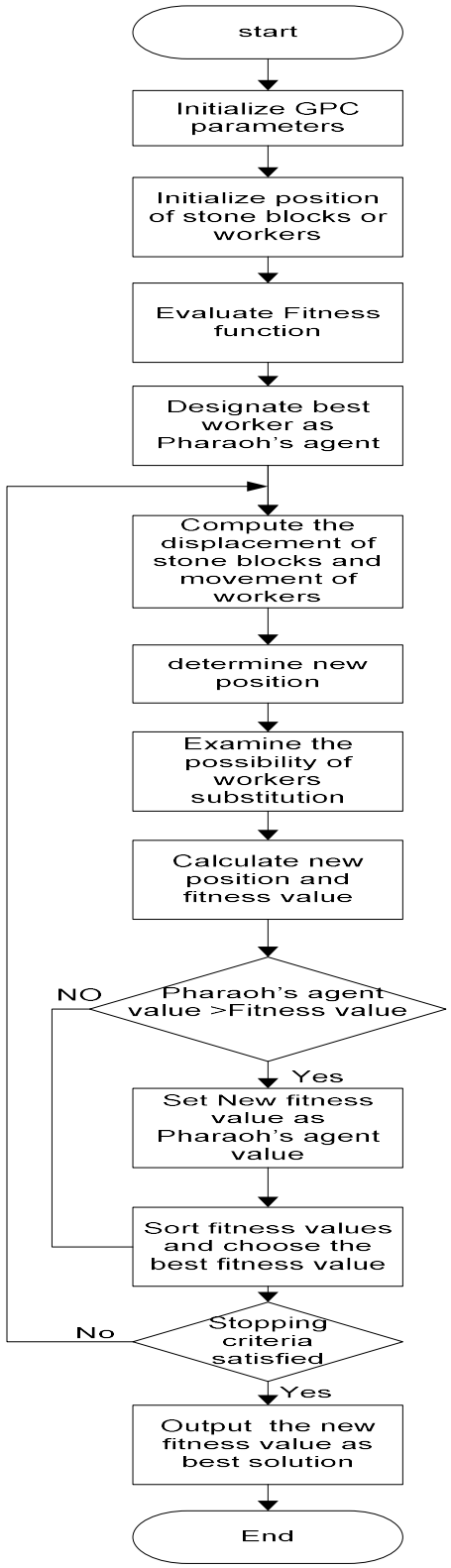



Fig. 1. Flow Chart of the applied Giza Pyramids Construction optimization

TABLE I. GENERATOR BUS DATA FOR IEEE 30 BUS SYSTEM

| Bus No. | $P_{Max}$ | $P_{Min}$ | $Q_{Max}$ | $Q_{Min}$ | Settings of Thermal Units, $P_G$ (MW) | |
|---|---|---|---|---|---|---|
| | (MW) | (MW) | (MVAr) | (MVAr) | Case 1 | Case 1A |
| 1 | 200 | 50 | 150 | -20 | Swing | Swing |
| 2 | 80 | 20 | 60 | -20 | 75 | 80 |
| 5 | 50 | 15 | 62.5 | -15 | 40 | 50 |
| 8 | 35 | 10 | 48.7 | -15 | 30 | 20 |
| 11 | 30 | 10 | 40 | -10 | 25 | 20 |
| 13 | 40 | 12 | 44.7 | -15 | 30 | 20 |

TABLE II. GENERATOR BUS DATA FOR IEEE 57 BUS SYSTEM

| Bus No. | $P_{Max}$ | $P_{Min}$ | $Q_{Max}$ | $Q_{Min}$ | Settings of Thermal Units, $P_G$ (MW) | |
|---|---|---|---|---|---|---|
| | (MW) | (MW) | (MVAr) | (MVAr) | Case 1 | Case 1A |
| 1 | 576 | 0 | 200 | -140 | Swing | Swing |
| 2 | 100 | 30 | 50 | -17 | 50 | 0 |
| 3 | 140 | 40 | 60 | -10 | 60 | 40 |
| 6 | 100 | 30 | 25 | -8 | 50 | 0 |
| 8 | 550 | 100 | 200 | -140 | 400 | 450 |
| 9 | 100 | 30 | 9 | -3 | 50 | 0 |
| 12 | 410 | 100 | 155 | -150 | 300 | 310 |

## IV. SIMULATION RESULTS

This section contains simulation results, extensive analysis, and comparisons for the ORPD problem, which is analyzed on the two cases of IEEE 30 & 57 bus systems respectively. Tables I and II contain the generating data for the IEEE 30 & 57 bus systems, respectively. These cases consider the different active power setting for generators, which are described in [19]. Also, it has been shown in Tables I and II. The ORPD problem is considered as mixed integer optimization problems where in shunt VAR compensators and transformer taps are considered as discrete variables. The shunt VAR compensators varies from 0 to 5 MVAr in a discrete step of 200 kVAr and the transformer tap settings varies from 0.9 to 1.1 pu. in a step of 0.02 p.u. The GPC technique was used with a population of 40 workers and a maximum of 1000 iterations. The goal of this work was to explore transmission loss minimization for two cases (Cases 1 & 1A) of the IEEE-30 bus system and two cases (Cases 2 & 2A) of the IEEE-57 bus system, respectively. The outcome of the applied procedure is compared to previously published methods such as SHADE-EC, JS, AEFA and Rao-3 algorithms.

- IEEE 30-bus test system

This system consists of 6 generating units, 9 shunt VAR compensators and, 4 regulating transformers with off-nominal tap ratio. Table III lists all of the control variables as well as their allowable limits for assessing the objective function for

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:42:28 UTC from IEEE Xplore. Restrictions apply.

the two cases of the IEEE 30-bus system. The acquired results for the control variables and the objective function are also reported. Table IV illustrates a comparison of the proposed study's results with some previously reported ORPD results.

- IEEE 57-bus test system

The system contains 7 synchronous generators, 17 regulating transformers with off-nominal tap ratio and, 3 reactive power compensators. Table V includes all of the control variables and their allowable limits for investigating the objective function for the two cases of the IEEE 57-bus situations. Table VI illustrates a comparison of the applied study's results with some previously reported ORPD results.

From the Table V and VI we can conclude that the transmission losses attained by the applied GPC algorithm are lower than those obtained by the AEFA, JS, SHADE-EC, and Rao-3 algorithms for the two cases of IEEE-30 and IEEE-57 bus systems. It's also worth noting that all of the optimum control variables are within permissible bounds. Figures 1,2,3 and 4 show the convergence characteristics of the applied ORPD approach in Cases (1), (1A), (2) and, (2A) respectively.

TABLE III. VALUES OF CONTROL VARIABLES AND OBJEVTIVE FUNCTIONS FOR IEEE 30 BUS SYSTEM

| **Parameters** | **Control Variables Settings** | | | **Case 1** | **Case 1A** |
|---|---|---|---|---|---|
| | | ***Min*** | ***Max*** | ***Min $P_{Loss}$*** | |
| Generator Voltage (p.u.) | V1 | 0.95 | 1.1 | 1.10 | 1.10 |
| | V2 | 0.95 | 1.1 | 1.10 | 1.09 |
| | V5 | 0.95 | 1.1 | 1.08 | 1.08 |
| | V8 | 0.95 | 1.1 | 1.08 | 1.08 |
| | V11 | 0.95 | 1.1 | 1.10 | 1.10 |
| | V13 | 0.95 | 1.1 | 1.10 | 1.10 |
| Transformer Tap Position (p.u.) | T11 | 0.9 | 1.1 | 1.02 | 1.05 |
| | T12 | 0.9 | 1.1 | 0.98 | 0.91 |
| | T15 | 0.9 | 1.1 | 1.02 | 0.99 |
| | T36 | 0.9 | 1.1 | 1.01 | 0.97 |
| Capacitor Bank (MVAr) | QC10 | 0 | 5 | 3.54 | 3.21 |
| | QC12 | 0 | 5 | 4.55 | 4.85 |
| | QC15 | 0 | 5 | 4.59 | 3.89 |
| | QC17 | 0 | 5 | 5.00 | 5.00 |
| | QC20 | 0 | 5 | 4.21 | 4.79 |
| | QC21 | 0 | 5 | 4.96 | 5.00 |
| | QC23 | 0 | 5 | 2.82 | 2.91 |
| | QC24 | 0 | 5 | 4.95 | 5.00 |
| | QC29 | 0 | 5 | 3.21 | 2.09 |
| Objective Functions | $P_{Loss}$ (MW) | N.A. | | 4.1253 | 4.5488 |

TABLE IV. RESULTS COMPARISON FOR THE IEEE-30 BUS SYSTEM

| Case No. | Algorithms | Power Loss (MW) |
|---|---|---|
| Case 1 | SHADE-EC | 4.4126 [21] |
| | AEFA | 4.4711 [22] |
| | JS | 4.4289 [22] |
| | Rao-3 | 4.4124 [22] |
| | GPC | **4.1253** |
| Case 1A | SHADE-EC | 4.8612 [21] |
| | AEFA | 4.9393 [22] |
| | JS | 4.8635 [22] |
| | Rao-3 | 4.8612 [22] |
| | GPC | **4.5488** |

TABLE V. VALUES OF CONTROL VARIABLES AND OBJEVTIVE FUNCTIONS FOR IEEE 57 BUS SYSTEM

| **Parameters** | **Control Variables Settings** | | | **Case 2** | **Case 2A** |
|---|---|---|---|---|---|
| | | ***Min*** | ***Max*** | ***Min $P_{Loss}$*** | |
| Generator Voltage (p.u.) | V1 | 0.95 | 1.1 | 1.10 | 1.10 |
| | V2 | 0.95 | 1.1 | 1.10 | 1.10 |
| | V3 | 0.95 | 1.1 | 1.09 | 1.09 |
| | V6 | 0.95 | 1.1 | 1.09 | 1.08 |
| | V8 | 0.95 | 1.1 | 1.10 | 1.10 |
| | V9 | 0.95 | 1.1 | 1.09 | 1.09 |
| | V12 | 0.95 | 1.1 | 1.08 | 1.08 |
| Transformer Tap Position (p.u.) | T19 | 0.9 | 1.1 | 1.10 | 1.10 |
| | T20 | 0.9 | 1.1 | 1.10 | 1.09 |
| | T31 | 0.9 | 1.1 | 1.09 | 1.10 |
| | T35 | 0.9 | 1.1 | 1.09 | 1.10 |
| | T36 | 0.9 | 1.1 | 1.10 | 1.10 |
| | T37 | 0.9 | 1.1 | 1.05 | 1.06 |
| | T41 | 0.9 | 1.1 | 1.03 | 1.05 |
| | T46 | 0.9 | 1.1 | 1.07 | 1.08 |
| | T54 | 0.9 | 1.1 | 1.10 | 1.10 |
| | T58 | 0.9 | 1.1 | 0.98 | 0.96 |
| | T59 | 0.9 | 1.1 | 0.99 | 0.97 |
| | T65 | 0.9 | 1.1 | 1.01 | 0.99 |
| | T66 | 0.9 | 1.1 | 0.99 | 0.97 |
| | T71 | 0.9 | 1.1 | 1.03 | 1.03 |
| | T73 | 0.9 | 1.1 | 1.10 | 1.10 |
| | T76 | 0.9 | 1.1 | 1.07 | 1.08 |
| | T80 | 0.9 | 1.1 | 1.05 | 1.04 |
| Capacitor Bank (MVAr) | QC18 | 0 | 20 | 16.29 | 16.09 |
| | QC25 | 0 | 20 | 16.74 | 16.17 |
| | QC53 | 0 | 20 | 13.45 | 13.54 |
| Objective Functions | $P_{Loss}$ (MW) | N.A. | | 17.7484 | 22.5401 |

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:42:28 UTC from IEEE Xplore. Restrictions apply.

TABLE VI. RESULTS COMPARISON FOR THE IEEE-57 BUS SYSTEM

| Case No. | Algorithms | Power Loss (MW) |
|---|---|---|
| Case 2 | SHADE-EC | 18.400 [21] |
| | AEFA | 20.610 [22] |
| | JS | 18.594 [22] |
| | Rao-3 | 18.1494 [22] |
| | GPC | **17.7484** |
| Case 2A | SHADE-EC | 23.3031 [21] |
| | AEFA | 25.648 [22] |
| | JS | 23.356 [22] |
| | Rao-3 | 23.3040 [22] |
| | GPC | **22.5401** |

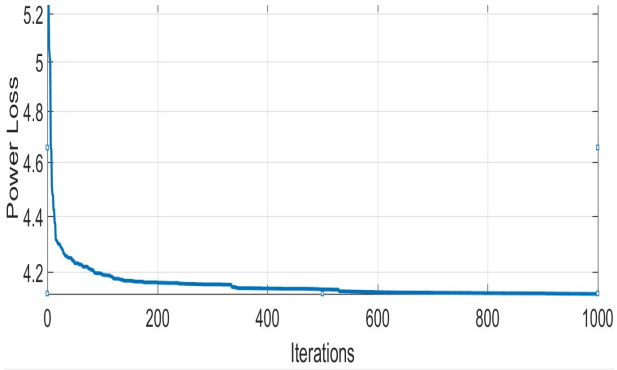



Fig. 2. GPC algorithm convergence characteristics for Case 1

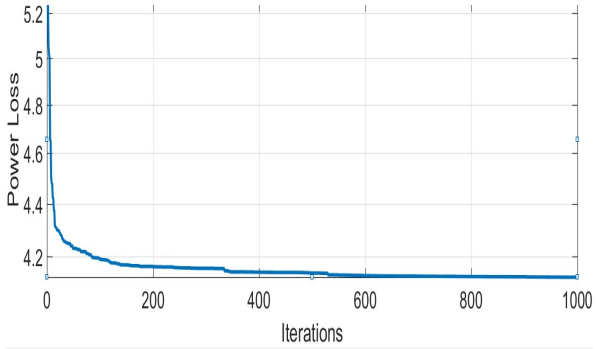



Fig. 3. GPC algorithm convergence characteristics for Case 1A

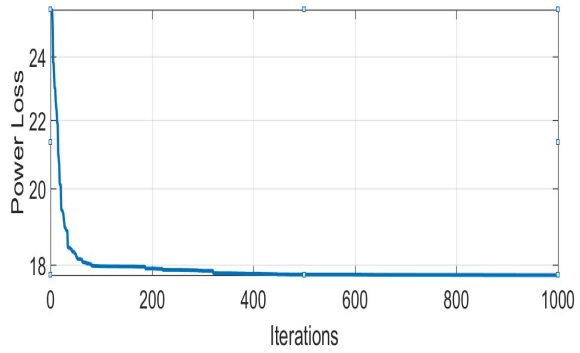



Fig. 4. GPC algorithm convergence characteristics for Case 2

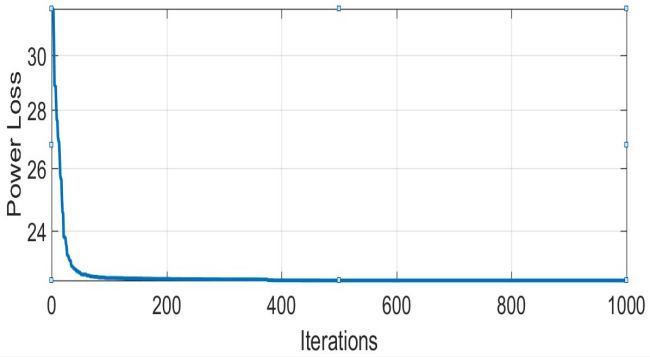



Fig. 5. GPC algorithm convergence characteristics for Case 2A

## V. CONCLUSIONS

This work introduced the Giza Pyramids Construction (GPC) Optimization algorithm, a new ancient-inspired meta-heuristic technique for addressing the ORPD issue using generator voltages, transformer tap settings, and shunt VAR sources as control variables to achieve the lowest transmission losses. The GPC approach's findings were compared to those described in current literature, demonstrating that the GPC algorithm can solve ORPD more efficiently in terms of search capability and resilience. In the future, the study could be expanded to include multi-objective optimization utilising a multi-objective evolutionary algorithm (MOEA), as well as the inclusion of voltage stability as a goal.

## REFERENCES


[1] AlRashidi, M.R. and El-Hawary, M.E., "Applications of computational intelligence techniques for solving the revived optimal power flow problem," Electric Power Systems Research, 79(4), 2009, pp.694-702.

[2] Chen G, Lu Z, Zhang Z, "Improved krill herd algorithm with novel constraint handling method for solving optimal power flow problems," Energies 2018, 11(1):76.

[3] H.W. Dommel, W.F. Tinney, "Optimal power flow solutions," IEEE Transactions on power apparatus and systems (10) 1968,pp. 1866–1876.

[4] B.A. Robbins, A.D. Domínguez-García, "Optimal reactive power dispatch for voltage regulation in unbalanced distribution systems," IEEE Transactions on Power Systems 31 (4), 2016,pp. 2903–2913.

[5] J.G. Vlachogiannis, K.Y. Lee, "Quantum-inspired evolutionary algorithm for real and reactive power dispatch," IEEE transactions on power systems 23 (4) (2008),pp. 1627–1636.

[6] D. Devaraj, J.P. Roselyn, "Genetic algorithm based reactive power dispatch for voltage stability improvement," International Journal of Electrical Power & Energy Systems 32 (10) (2010),pp. 1151–1156.

[7] Q.H. Wu, Y.J. Cao, J.Y. Wen, "Optimal reactive power dispatch using an adaptive genetic algorithm," Int. J. Electr. Power Energy Syst. 20 (8) (1998),pp. 563–569.

[8] Q. H. Wu and J. T. Ma, "Power system optimal reactive power dispatch using evolutionary programming," in IEEE Transactions on Power Systems, vol. 10, no. 3, Aug. 1995, pp. 1243-1249.

[9] Yutian, L. and Li, M., "Reactive power optimization based on Tabu search approach," Automation of Electric Power Systems, 2. 2000.

[10] Chi, R., Li, Z., Chi, X., Qu, Z. and Tu, H.B., "Reactive Power Optimization of Power System Based on Improved Differential Evolution Algorithm," Mathematical Problems in Engineering, 2021.

[11] J.G. Vlachogiannis, K.Y. Lee, "A comparative study on particle swarm optimization for optimal steady-state performance of power systems," Power Syst. IEEE Trans. 21 (4), 2006, pp.1718–1728.

[12] Reddy PL, Yesuratnam G, "PSO based optimal reactive power dispatch for voltage profile improvement," Power, communication and information technology conference (PCITC), 2015, IEEE, pp. 361–366.


Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:42:28 UTC from IEEE Xplore. Restrictions apply.

[13] A.H. Khazali, M. Kalantar, “Optimal reactive power dispatch based on harmony search algorithm,” Int. J. Electr. Power Energy Syst. 33 (3) 2011, pp. 684–692.

[14] Mei, R.N.S., Sulaiman, M.H., Mustaffa, Z. and Daniyal, H., “Optimal reactive power dispatch solution by loss minimization using moth-flame optimization technique,” Applied Soft Computing, 59, 2017, pp. 210-222.

[15] Duman, S., Sönmez, Y.U.S.U.F., Güvenç, U. and Yörükeren, N., “Optimal reactive power dispatch using a gravitational search algorithm,” IET generation, transmission & distribution, 6(6), 2012, pp.563-576.

[16] Rayudu K, Yesuratnam G, Jayalaxmi A (2016) Artificial bee colony algorithm for optimal reactive power dispatch to improve voltage stability. In: Circuit, power and computing technologies (ICCPCT), 2016 international conference on IEEE, pp. 1–7.

[17] Nguyen, T.T. and Vo, D.N., “Improved social spider optimization algorithm for optimal reactive power dispatch problem with different objectives,” Neural Computing and Applications, 32(10), 2020, pp.5919-5950.

[18] S. Harif, J. Mohammadzadeh, M. Khalilian and S. Ebrahimnejad, “Giza Pyramids Construction: an ancient-inspired metaheuristic algorithm for optimization,” Evolutionary Intelligence, 2021, 14: pp.1743–1761.

[19] Power Systems Test Case Archive, Univ. Washington, Seattle, WA, USA, 2020.

[20] A. Rajan and T. Malakar, ‘‘Exchange market algorithm based optimum reactive power dispatch,’’ Appl. Soft Comput., vol. 43, Jun. 2016, pp. 320–336.

[21] P. Biswas, P. N. Suganthan, R. Mallipeddi, and G. A. J. Amaratunga, ‘‘Optimal reactive power dispatch with uncertainties in load demand and renewable energy sources adopting scenario-based approach,’’ Appl. Soft Comput., vol. 75, Feb. 2019, pp. 616–632.

[22] Mohamed H. Hassan1 , Salah Kamel , Mahmoud A. El-Dabah , Tahir Khurshaid And José Luis Domínguez-García, “Optimal Reactive Power Dispatch With Time-Varying Demand and Renewable Energy Uncertainty Using Rao-3 Algorithm,” IEEE Access, VOLUME 9, 2021, pp. 23264-23283.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2022 at 04:42:28 UTC from IEEE Xplore. Restrictions apply.

Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

# Optimization of process parameters for 3D printing process using Taguchi based grey approach

Ramesh Raju [a,*], M. Murali M.K. Varma [b], Pushp Kumar Baghel [b]

[a] Department of Mechanical Engineering, Sree Vidyanikethan Engineering College, Tirupati 517102, AP, India
[b] Department of Mechanical Engineering, Noida International University, Plot 1, Sector 17A, Yamuna Expressway, Gautam Budh Nagar, Uttar Pradesh 203201, India

ARTICLE INFO

*Article history:*
Available online xxxx

*Keywords:*
Additive Manufacturing
3D printing
Optimization
Taguchi Grey Approach
PLA Materials

ABSTRACT

Additive manufacturing (AM) is the method for fabricating the components effectively. The components fabricated by AM process has wider applications in various domains. Fusion deposition modeling is one of the AM methodology that has been used for fabricating the components. Biodegradability of PLA makes the material as best alternate for plastics and carbon nano particles has been addedwith PLA which has the possibility using this material for various engineering applications. Taguchi's experimental design is concept used for devising the experimentation with the help of selected independent process variables. Single aspects optimization has been performed by Taguchi's approach. Tensile strength, toughness and roughness of the fabricated samples are considered as performance measures. Grey system theory is the methodology engaged for adopting the multi aspects optimization. The best possible process variables are identified for multiple experimentation measures. The effect combinatorial variables were identified by using the interaction analysis.
Copyright © 2022 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Advanced Materials and Modern Manufacturing.

## 1. Introduction

3-D printing is one among the advanced method of additive manufacturing (AM) for making various range of products. The method comprises of printing continuous layers one among another. AM is most commonly adopted method in various engineering applications [1]. Techniques of AM have been developed comprehensively with various types of materials and it is a beneficial approach for making prototypes, tools and other components [2]. Fusion Deposition Modelling [FDM] is a methodology which using similar to traditional method such as injection moulding. This method of making components involves molten plastic which is extruded from a hot nozzle for making the part layer one above another [3]. FDM is the most commonly engaged method due to its ease of use and economic benefits and also it is the method to fabricate functional parts without any difficulty in achieving geometrical requirements [4–6]. Polyactic Acid (PLA) is one of the most auspicious material that is biodegradable and it can be the suitable alternate for traditionally available plastics. Also the exceptional tensile characteristics of these materials laid a pathway as best opted for medical and other mild structural applications. Packaging industries have extensive usage of PLA material due to its biodegradable characteristics [7–11]. Response Surface Methodology has been engaged for analyzing the process parameters and also for deriving the relationship amidst the considered input and output variables with the help of empirical equations. Also the effect of process parameters on desired output variables are also assessed [12–14].An investigation has been performed for optimizing the mechanical characteristics using Taguchi based grey approach. The interaction effect among the process parameters have been derived and it is significant process parameters were identified [15].

AM is one of the most favorable method for making PLA with needed structure. Since the traditional approaches for not offering required performance expected by various manufacturers. Carbon nanoparticles are most considered as reinforcement due to its biocompatibility for various industrial applications [16–19]. Taguchi's approach is one of the important and proficient method for experiment design and it is most suitable for optimizing the parameters with single objectives. Multiple process parameters have been optimized with the help of Multi Criteria Decision Making (MCDM)

* Corresponding author.
*E-mail address:* mrramesh2002@gmail.com (R. Raju).

https://doi.org/10.1016/j.matpr.2022.07.163
2214-7853/Copyright © 2022 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Advanced Materials and Modern Manufacturing.

Please cite this article as: R. Raju, M. Murali M.K. Varma and P. Kumar Baghel, Optimization of process parameters for 3D printing process using Taguchi based grey approach, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2022.07.163

approaches. Grey theory is one of MCDM approach which will be effectively utilized for optimizing multi variables [20–25].

Based on available literature, it is decided that there is a need of manufacturing the PLA components with FDM approach. Also investigational analysis on desired performance measures by using hybrid Grey-Taguchi method needs major consideration. The present investigational analysis, an attempt has been made to execute an exertion to ascertain the characteristics of major processing variables and evaluating the material properties of in-situ PLA polymer materials with the aid of Taguchi based Grey approach.

## 2. Materials and methods

PLA is one of the environmental friendly material and also having better biodegradability. In this present investigation, PLA is considered as work material and fed through the FDM setup which is illustrated in Fig. 1. PLA which is taken as wire are fused and made to deposit as layer one among another. The thickness and size of the layer have been predetermined and fed into the FDM setup. The submicron sized carbon particles have been selected for achieving the desired mechanical characteristics. During the time of fabricating the layers, an attachment also provided for depositing the carbon particles on the surface. The desired work samples are made in accordance with "ASTM D638 TYPE V" standard.

Taguchi's experimental design is adopted for reducing the complexity while designing the experimental trials. It is a proven approach for making effective experimental design and also for analyzing the process parameters. Based on the literature three factors with three levels are selected as shown in Table 1 and as per Taguchi's approach an L27 Orthogonal Array has been selected for conducting experiments. By varying the process parameters as per selected OA, 27 parts are printed. The temperature of extruder and bet are maintained as 240 °C and 100 °C respectively, throughout the experimentation. Tensile strength, toughness and surface roughness of printed samples are considered as performance measures in this investigation.

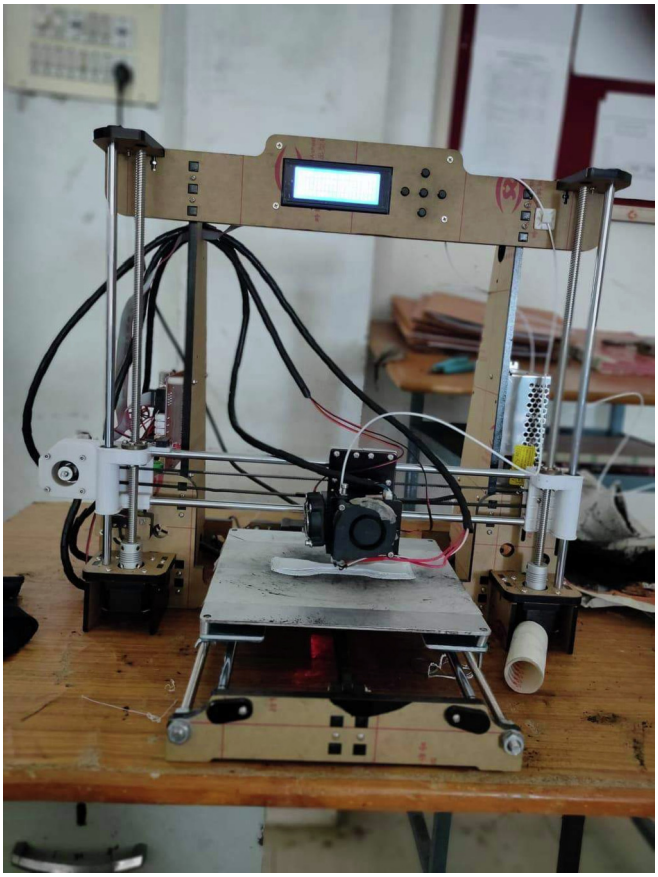


**Fig. 1.** Fused deposition Machine setup.

**Table 1**
Processing variables against different levels.

| Symbols | Variables | Levels | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| A | Printing Speed (mm/sec) | 70 | 80 | 90 |
| B | C Deposition (mg) | 5 | 10 | 15 |
| C | No. of Layers | 1 | 2 | 3 |

## 3. Results and discussion

To explore implication of the independent variables against preferred output variables, experimentation has been done as per $L_{27}$OA, during FDM of PLA material.

The effective FDM process is achieved through determining the best possible process variables, an endeavor has been taken in this investigational analysis. In the FDM process, higher tensile strength, higher toughnessand minimum surface roughness are the signs of customary routine procedures. So tensile strength and toughness are deemed as higher the healthier criterion and the same passion the lower the healthier criterion for surface roughness. The layout for conducting experiments with observation are shown in Table 2.

## 4. Response analysis on tensile strength

The response analysis on tensile strength has been conducted and illustrated in Fig. 2. The figure indicates that the tensile

**Table 2**
Experimental layout with observations.

| Sl. No | Input variables | | | Output variables | | |
|---|---|---|---|---|---|---|
| | Printing Speed (mm/sec) | C Deposition (mg) | No. of Layers | Tensile Strength (Mpa) | Toughness (J/m$^3$) | Surface Roughness (microns) |
| 1 | 70 | 5 | 1 | 7.77 | 1.98 | 0.005 |
| 2 | 70 | 5 | 2 | 8.52 | 2.03 | 0.006 |
| 3 | 70 | 5 | 3 | 8.18 | 1.78 | 0.005 |
| 4 | 70 | 10 | 1 | 8.65 | 2.11 | 0.005 |
| 5 | 70 | 10 | 2 | 9.02 | 2.12 | 0.007 |
| 6 | 70 | 10 | 3 | 8.58 | 2.05 | 0.006 |
| 7 | 70 | 15 | 1 | 8.42 | 1.87 | 0.006 |
| 8 | 70 | 15 | 2 | 8.55 | 2.10 | 0.005 |
| 9 | 70 | 15 | 3 | 8.48 | 1.88 | 0.006 |
| 10 | 80 | 5 | 1 | 8.88 | 2.10 | 0.006 |
| 11 | 80 | 5 | 2 | 9.02 | 2.52 | 0.005 |
| 12 | 80 | 5 | 3 | 8.74 | 2.31 | 0.006 |
| 13 | 80 | 10 | 1 | 9.15 | 2.64 | 0.005 |
| 14 | 80 | 10 | 2 | 7.88 | 1.95 | 0.006 |
| 15 | 80 | 10 | 3 | 8.98 | 2.58 | 0.005 |
| 16 | 80 | 15 | 1 | 9.11 | 2.05 | 0.006 |
| 17 | 80 | 15 | 2 | 8.85 | 2.45 | 0.006 |
| 18 | 80 | 15 | 3 | 8.69 | 2.22 | 0.005 |
| 19 | 90 | 5 | 1 | 7.14 | 1.77 | 0.007 |
| 20 | 90 | 5 | 2 | 7.72 | 2.25 | 0.006 |
| 21 | 90 | 5 | 3 | 7.68 | 1.80 | 0.006 |
| 22 | 90 | 10 | 1 | 8.44 | 2.11 | 0.006 |
| 23 | 90 | 10 | 2 | 8.15 | 2.42 | 0.005 |
| 24 | 90 | 10 | 3 | 8.02 | 1.88 | 0.006 |
| 25 | 90 | 15 | 1 | 7.92 | 1.85 | 0.005 |
| 26 | 90 | 15 | 2 | 7.56 | 1.85 | 0.006 |
| 27 | 90 | 15 | 3 | 9.55 | 2.82 | 0.004 |

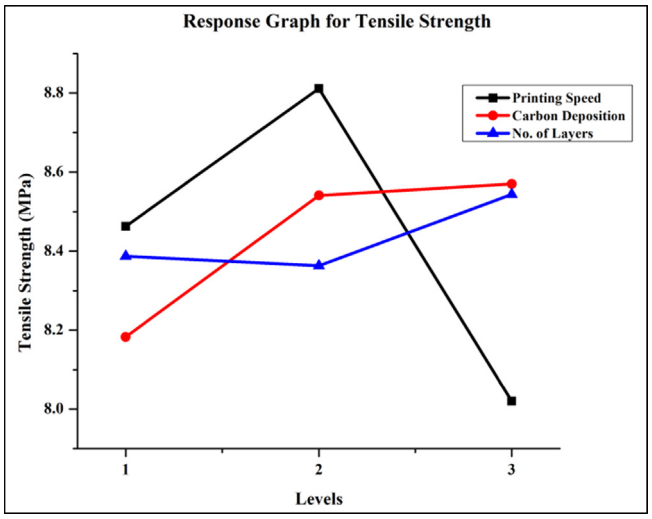



**Fig. 2.** Response graph for tensile strength.

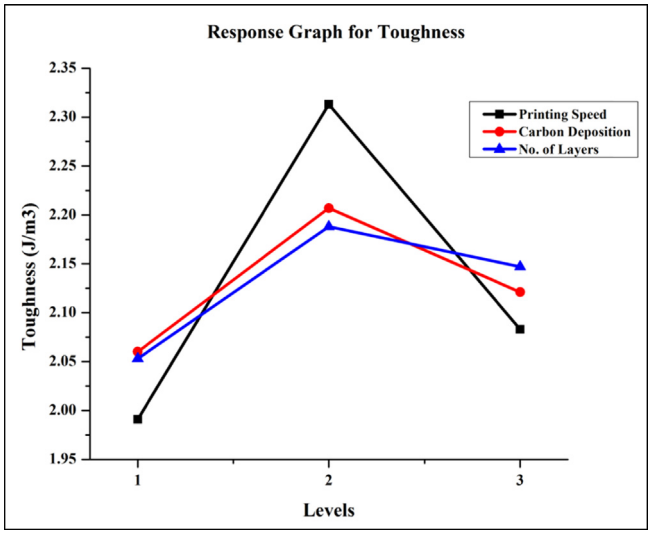



**Fig. 3.** Response graph for toughness.

**Table 3**
Mean - Tensile strength Analysis.

| Response Table for Means - Tensile Strength | | | |
|---|---|---|---|
| Levels | Printing Speed | Carbon Deposition | No. of Layers |
| 1 | 8.463 | 8.183 | 8.387 |
| 2 | 8.811 | 8.541 | 8.363 |
| 3 | 8.02 | 8.57 | 8.544 |
| Delta | 0.791 | 0.387 | 0.181 |
| Rank | 1 | 2 | 3 |

strength of 3D printed component having increasing trend with increase in number of layers and carbon deposition.

It is obvious from the observation that the layer and carbon deposition makes considerable increase in the strength of material. Tensile strength is predominantly influenced by printing speed. Intermediate level of printing speed makes better impact strength on the components. The optimum process parameters have been arrived and tabulated in the Table 3. The values also indicate the best combination for getting improved tensile strength as A2B3C3.

## 5. Response analysis on toughness

Fig. 3 represents the response analysis for toughness of the component prepared by AM process.

It is obvious from the illustrations that the toughness of components are influenced by printing speed. The intermediate levels of printing speed, carbon deposition and number of layers produces better and improved toughness on the printed components. The optimum process parameters have been accomplished by Taguchi's single response analysis. The outcomes of the present analysis are depicted in Table 4. It is clear from the outcomes that the best suitable combination for attaining better toughness is A2B2C2.

## 6. Response analysis on surface roughness

The response analysis for roughness of fabricated samples has been conducted and the outcomes are graphically illustrated in Fig. 4.

The behavior of surface roughness depicts that the roughness of the printed components is decreasing against increasing of carbon deposition in polymer material. Intermediate level of printing speed and maximum number of layer produces better roughness on the printed components. Carbon deposition is the primarily

**Table 4**
Means - Toughness analysis.

| Response Table for Means - Toughness | | | |
|---|---|---|---|
| Levels | Printing Speed | Carbon Deposition | No. of Layers |
| 1 | 1.991 | 2.06 | 2.053 |
| 2 | 2.313 | 2.207 | 2.188 |
| 3 | 2.083 | 2.121 | 2.147 |
| Delta | 0.322 | 0.147 | 0.134 |
| Rank | 1 | 2 | 3 |

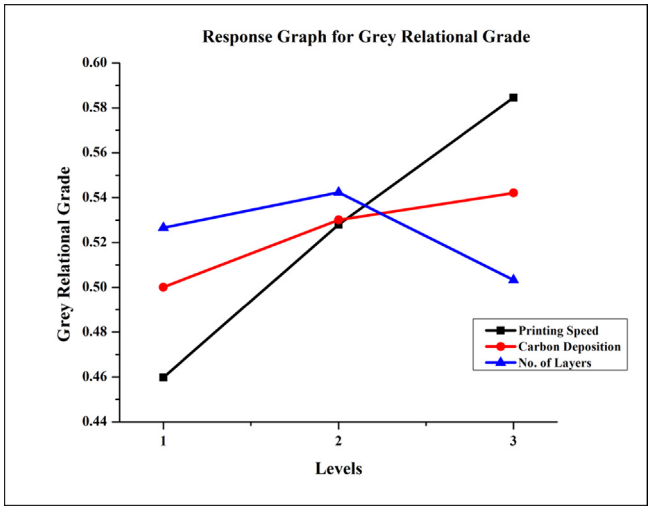



**Fig. 4.** Surface roughness against various levels.

**Table 5**
Means - Surface roughness Analysis.

| Response Table for Means – Surface Roughness | | | |
|---|---|---|---|
| Levels | Printing Speed | Carbon Deposition | No. of Layers |
| 1 | 0.005667 | 0.005778 | 0.005667 |
| 2 | 0.005556 | 0.005667 | 0.005778 |
| 3 | 0.005667 | 0.005444 | 0.005444 |
| Delta | 0.000111 | 0.000333 | 0.000333 |
| Rank | 3 | 1 | 2 |

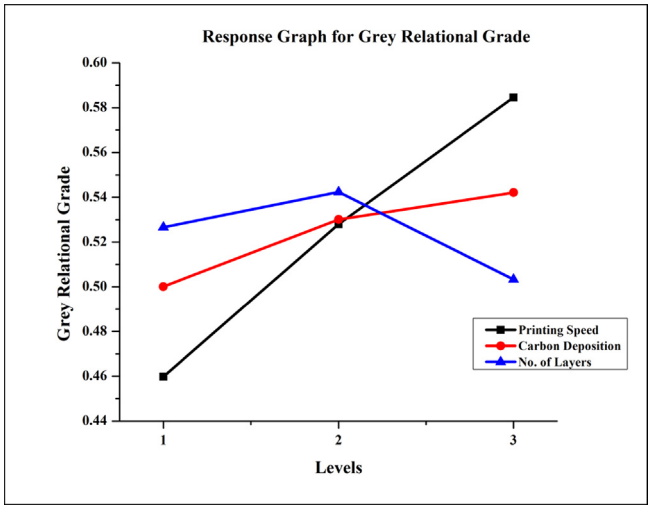



**Fig. 5.** Response graph for GRG.

**Table 6**
Means - Grey Relational Grade Analysis.

| Response Table for Means – Grey Relational Grade | | | |
|---|---|---|---|
| Levels | Printing Speed | Carbon Deposition | No. of Layers |
| 1 | 0.4598 | 0.5001 | 0.5266 |
| 2 | 0.5280 | 0.5301 | 0.5424 |
| 3 | 0.5846 | 0.5421 | 0.5033 |
| Delta | 0.1248 | 0.0420 | 0.0390 |
| Rank | 1 | 2 | 3 |

influencing process variable for roughness. Taguchi's response analysis has been done for determining the optimum process parameters and represented in Table 5. The behavioral characteristics and response values indicates the lower roughness is accomplished with the parameter combinations of A2B3C3.

## 7. Response analysis on GRG (multi performance index)

Grey relational analysis, a multi aspects optimization method has been adopted for determining multiple performance characteristics and multiple aspects are converted into single aspect. Then Taguchi's analysis has been adopted for ascertaining the best suitable parameter combination for attaining multiple performance. Fig. 5 shows the response analysis for multi performance index GRG.

Multiple performance of the selected FDM process is improved by increase in printing speed and carbon deposition. The illustration further depicts that the printing speed acts as the major process variable that influences the multiple performance of the selected process. Taguchi's response analysis has been done for determining the optimum process parameters and represented in Table 6. It is noted from the investigation that the multi-performance index of the selected process is improved with the parameter combinations of A3B3C2.

## 8. Interaction analysis of processing variables

Interaction is possible while the deviation in performance measures between various levels of considered factors and the analysis are performed. Interaction between the two factors are revealed by non-parallel line in the interaction graphs. Fig. 6 illustrates the interaction plot for tensile strength.

It is observed from the plot that the tensile strength of FDMed component has considerable interaction among all the process variables with different levels. Interaction plot for toughness of 3D printed component is shown in Fig. 7.

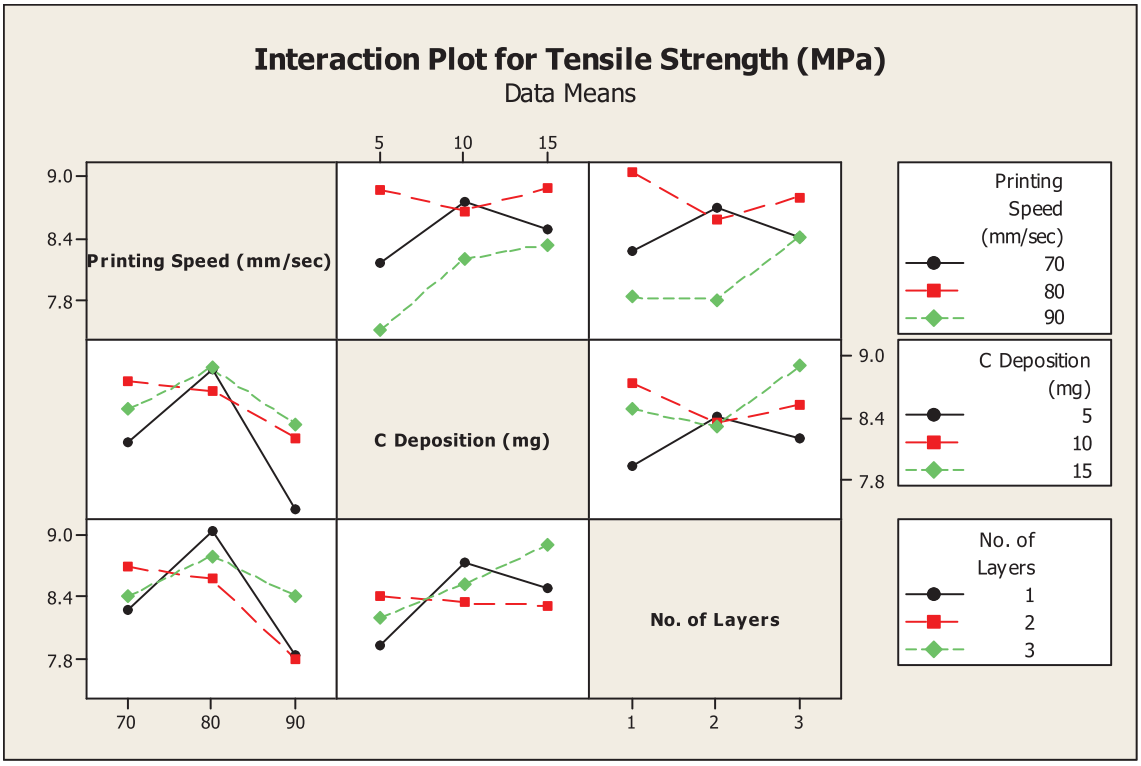



**Fig. 6.** Interaction plot for tensile strength.

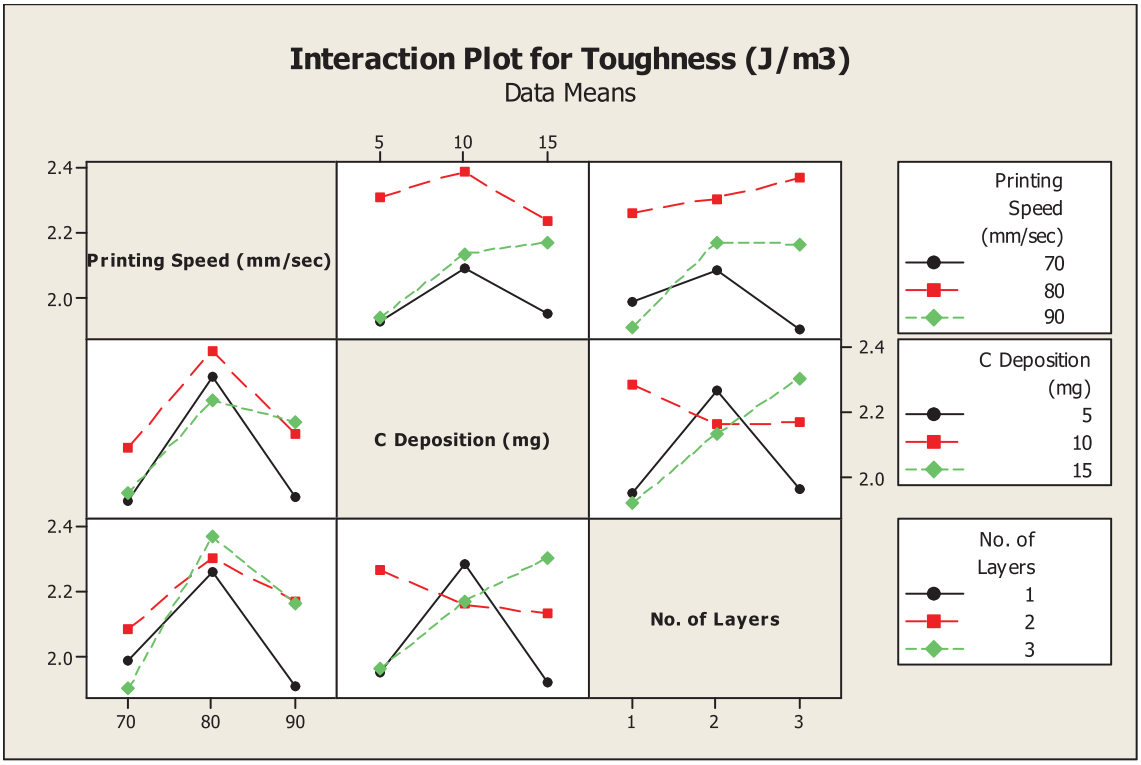



**Fig. 7.** Interaction plot for toughness.

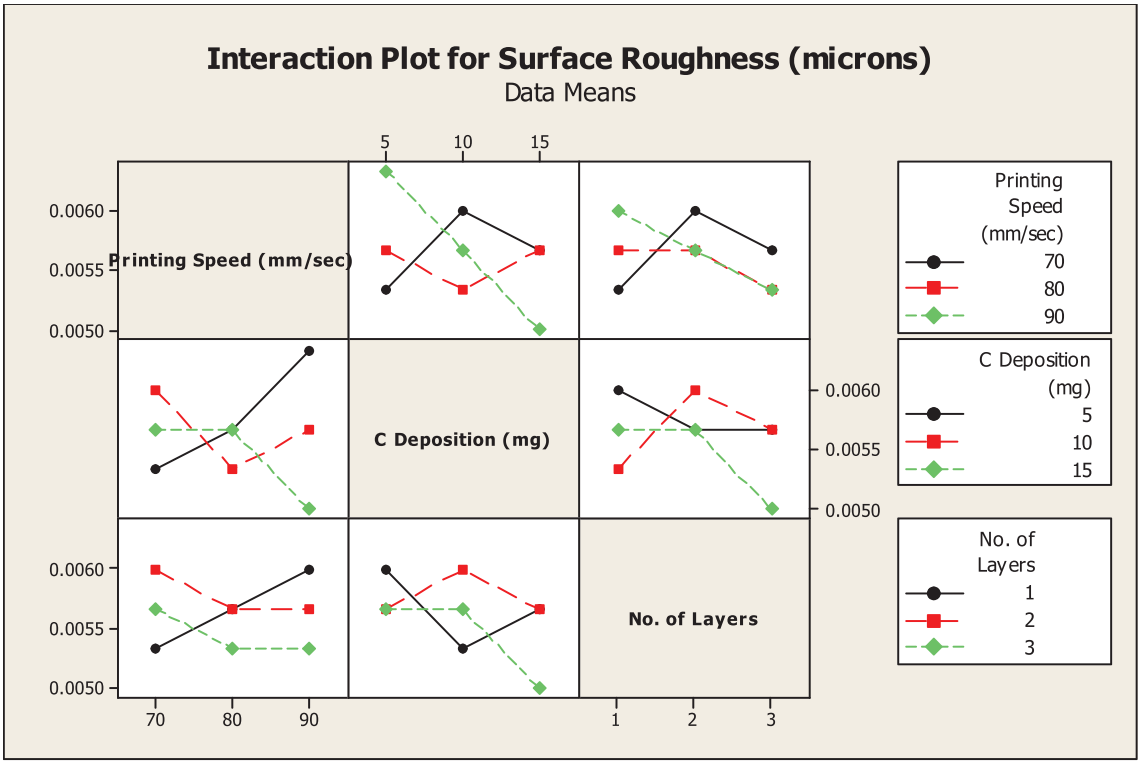



**Fig. 8.** Interaction plot for surface roughness.

The plot also depicts that there is ahuge interaction among the carbon deposition and number of layers at all levels. The interaction plot for the surface roughness is shown in Fig. 8. The surface roughness is combinatorial effect of all parameters with different levels.

## 9. Conclusions

- FDM process has been adopted for fabricating the in-situ PLA polymer with carbon particles. Single aspects optimization by Taguchi's analysis has been used for determining the best process parameters to attain optimal single performance.
- Tensile strength of the fabricated component is influence by printing speed and then it is followed by carbon deposition and number of layers.
- The toughness of fabricated samples is predominantly influenced by printing speed and then it is trailed by carbon deposition and number of layers.
- Carbon deposition is the major parameter that influences the surface roughness of the fabricated component and then it is trailed by number of layers and printing speed.
- GRA method has been engaged for ascertaining the multi performance index and the optimum process variables for attaining multiple performance are derived. According the GRA approach, printing speed is the significant process variable that influences the overall performance of the selected FDM process.
- The interaction effect of process variables are revealed by interaction analysis. Most the considered performance measures are having interaction effect with all parameters at different levels.

## Data availability

Data will be made available on request.

## 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.

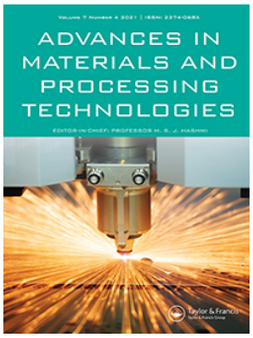
## References

[1] T.D. Ngo, A. Kashani, G. Imbalzano, K.T.Q. Nguyen, D. Hui, Additive manufacturing (3D printing): A review of materials, methods, applications and challenges, Compos. B Eng. 143 (2018) 172–196.
[2] J.-P. Kruth, G. Levy, F. Klocke, T.H.C. Childs, Consolidation phenomena in laser and powder-bed based layered manufacturing, CIRP Ann. 56 (2) (2007) 730–759.
[3] Novakova-Marcincinova, L., & Novak-Marcincin, J. (2013). Experimental testing of materials used in fused deposition modeling rapid prototyping technology. In *Advanced Materials Research* (Vol. 740, pp. 597-602). Trans Tech Publications Ltd.
[4] Ahn, Sung-Hoon, et al. "Anisotropic material properties of fused deposition modeling ABS." *Rapid prototyping journal* (2002).
[5] B.H. Lee, J. Abdullah, Z.A. Khan, Optimization of rapid prototyping parameters for production of flexible ABS object, J. Mater. Process. Technol. 169 (1) (2005) 54–61.
[6] A. Boschetto, L. Bottini, Design for manufacturing of surfaces to improve accuracy in fused deposition modeling, Rob. Comput. Integr. Manuf. 37 (2016) 103–114.
[7] R. Anitha, S. Arunachalam, P. Radhakrishnan, Critical parameters influencing the quality of prototypes in fused deposition modelling, J. Mater. Process. Technol. 118 (1-3) (2001) 385–388.
[8] B. Vasudevarao, D.P. Natarajan, M. Henderson, A. Razdan, in: August). Sensitivity of RP surface finish to process parameter variation, The University of Texas, Austin, TX, 2000, pp. 251–258.
[9] M.E.R. Bonifacio, A.E. Diniz, Correlating tool wear, tool life, surface roughness and tool vibration in finish turning with coated carbide tools, Wear 173 (1-2) (1994) 137–144.
[10] S.L. Campanelli, G. Cardano, R. Giannoccaro, A.D. Ludovico, E.L.J. Bohez, Statistical analysis of the stereolithographic process to improve the accuracy, Comput. Aided Des. 39 (1) (2007) 80–86.
[11] O.S. Es-Said, J. Foyos, R. Noorani, M. Mendelson, R. Marloth, B.A. Pregger, Effect of layer orientation on mechanical properties of rapid prototyped samples, Mater. Manuf. Processes 15 (1) (2000) 107–122.
[12] A.K. Sood, R.K. Ohdar, S.S. Mahapatra, Parametric appraisal of mechanical property of fused deposition modelling processed parts, Mater. Des. 31 (1) (2010) 287–295.
[13] F. Rayegani, G.C. Onwubolu, Fused deposition modelling (FDM) process parameter prediction and optimization using group method for data handling (GMDH) and differential evolution (DE), Int. J. Adv. Manuf. Technolo. 73 (1-4) (2014) 509–519.
[14] G.C. Onwubolu, F. Rayegani, Characterization and optimization of mechanical properties of ABS parts manufactured by the fused deposition modelling process, Int. J. Manuf. Eng. 2014 (2014) 1–13.
[15] A.K. Sood, R.K. Ohdar, S.S. Mahapatra, Improving dimensional accuracy of fused deposition modelling processed part using grey Taguchi method, Mater. Des. 30 (10) (2009) 4243–4252.
[16] M. Sarno, W.A. Abdalglil Mustafa, A. Senatore, D. Scarpa, One-step "green" synthesis of dispersable carbon quantum dots/poly (methyl methacrylate) nanocomposites for tribological applications, Tribol. Int. 148 (2020) 106311.
[17] A.A. Maxim, S.N. Sadyk, D. Aidarkhanov, C. Surya, A. Ng, Y.-H. Hwang, T.S. Atabaev, A.N. Jumabekov, PMMA thin film with embedded carbon quantum dots for post-fabrication improvement of light harvesting in perovskite solar cells, Nanomaterials 10 (2) (2020) 291.
[18] S.B. Aziz, O.G. Abdullah, M.A. Brza, A.K. Azawy, D.A. Tahir, Effect of carbon nano-dots (CNDs) on structural and optical properties of PMMA polymer composite, Results Phys. 15 (2019) 102776.
[19] C.S. Lee, S.G. Kim, H.J. Kim, S.H. Ahn, Measurement of anisotropic compressive strength of rapid prototyping parts, J. Mater. Process. Technol. 187-188 (2007) 627–630.
[20] T. Nancharaiah, D.R. Raju, V.R. Raju, An experimental investigation on surface quality and dimensional accuracy of FDM components, Int. J. Emerg. Technol. 1 (2) (2010) 106–111.
[21] M. Nouari, G. List, F. Girot, D. Coupard, Experimental analysis and optimisation of tool wear in dry machining of aluminium alloys, Wear 255 (7-12) (2003) 1359–1368.
[22] R. Raju, V. Sivalingam, J. Sun, M. Natarajan, Y. Zhao, Experimental and Taguchi-based grey approach of laser metal deposition technique on nickel-based superalloy, Trans. Indian Inst. Met. 72 (1) (2019) 205–214.
[23] Ramesh Raju, N Manikandan, D Palanisamy, D Arulkirubakaran, J S Binoj, P Thejasree, C Ahilan. A Review of Challenges and Opportunities in Additive Manufacturing. In: Palani, I.A., Sathiya, P., Palanisamy, D. (eds) Recent Advances in Materials and Modern Manufacturing. Lecture Notes in Mechanical Engineering. Springer, Singapore (2022). https://doi.org/10.1007/978-981-19-0244-4_3.
[24] Varma M M M K, Baghel P K, Raju R. Additive Manufacturing of Thermosetting Resins In-Situ Carbon Fibers: A Review. In: Palani, I.A., Sathiya, P., Palanisamy, D. (eds) Recent Advances in Materials and Modern Manufacturing. Lecture Notes in Mechanical Engineering. Springer, Singapore (2022). https://doi.org/10.1007/978-981-19-0244-4_11.
[25] K.G. Thirugnanasambantham, A. Francis, R. Ramesh, M. Aravind, Medagam Kesava Reddy. Investigation of erosion mechanisms on IN-718 based turbine blades under water jet conditions, Int. J. Interact. Design Manuf. (2022), https://doi.org/10.1007/s12008-022-00910-4.

Dr. RAMESH RAJU, completed Ph.D from NITT in the field of Laser Material Processing in the year 2015-16. Presently he is working as Associate Professor in Mechanical Engineering Department, Sree Vidyanikethan Engineering College (Autonomous), Tirupati 517102, A.P, India. His research interests include Advanced Machining, Additive Manufacturing and Surface Engineering. He published more than 50 research papers and awarded 10 patents. Ramesh is the primary author for the above subjected manuscript. He worked in the selection of materials and processing methods.

Mr. M. MURALI M K VARMA presently pursuing Ph.D from Department Of Mechanical School Of Engineering And Technology in NIU, Gautam Buddha Nagar, UP, India.and his field of Interest includes additive manufacturing of polymer materials for bio-medical engineering applications. Varma is the co- author for the above subjected manuscript. He contributed in the process parameter selection process

Dr. PUSHP KUMAR BAGHEL awarded M.Tech (Hon's) from Jamia Millia Islamia and Ph.D from Delhi Technological University in the year 2018 in the area of Welding Technology. He published more than 11 research papers in reputed national/International publications. Presently, he is working as Asso. Professor of Department Of Mechanical School Of Engineering And Technology in NIU, Gautam Buddha Nagar, UP, Uttar Pradesh, India. Babhel is the co- author for the above subjected manuscript. He contributed in the L27 process selection.



Advances in Materials and Processing Technologies

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tmpt20

# Parametric optimisation for detoxification of non-oil *Jatropha* residual material using Taguchi method

Shilpi Ahluwalia, Amrit Pal Toor, Pushpendra Singh & Jai Gopal Sharma

**To cite this article:** Shilpi Ahluwalia, Amrit Pal Toor, Pushpendra Singh & Jai Gopal Sharma (2022): Parametric optimisation for detoxification of non-oil *Jatropha* residual material using Taguchi method, Advances in Materials and Processing Technologies, DOI: 10.1080/2374068X.2022.2095756

**To link to this article:** https://doi.org/10.1080/2374068X.2022.2095756

Published online: 05 Jul 2022.

Submit your article to this journal

Article views: 5

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tmpt20

ADVANCES IN MATERIALS AND PROCESSING TECHNOLOGIES
https://doi.org/10.1080/2374068X.2022.2095756

Taylor & Francis
Taylor & Francis Group

Check for updates

# Parametric optimisation for detoxification of non-oil *Jatropha* residual material using Taguchi method

Shilpi Ahluwalia[a,b], Amrit Pal Toor[a], Pushpendra Singh[b] and Jai Gopal Sharma[b]

[a]Dr SSB University Institute of Engineering and Technology, Panjab University, Chandigarh, India; [b]Delhi Technological University, Delhi, India

**ABSTRACT**

Phorbol esters are the major toxic compounds found in *Jatropha* seed cake and oil. Removal of phorbol esters from *Jatropha* seed cake would allow the utilisation of detoxified seed cake as a protein-rich supplement for animal feed. Research was performed to determine the effects of submerged fermentation (SMF) on the phorbol ester (PE) degradation rate. In Taguchi design of experiments, the L18 orthogonal array was used for optimising the SMF process for detoxification of *Jatropha* seed cake. Process parameters selected for the study were pH, percentage of seed cake, time, temperature and rpm. The response parameter for detoxification was measured in terms of phorbol ester degradation. Analysis of variance (ANOVA) was performed for predicting the optimum process parameters. The percentage contribution of the process parameters with reference to phorbol ester degradation has been predicted. The hierarchy of significant process parameters is in the order of seed cake (1%), temperature (35 °C), pH (7.5) and time (12 h).

**ARTICLE HISTORY**
Accepted 24 June 2022

**KEYWORDS**
*Jatropha*; phorbol; detoxification; Taguchi; modelling; optimisation

## 1. Introduction

*Jatropha* is an oil-bearing plant belonging to the *Euphorbiaceae* family and has received significant interest in biodiesel production because of its high oil content [1]. *Jatropha* biodiesel industry is still in the developing stage because of its high cost, lack of utilisation of toxic by-products and no support from government in supply chain and infrastructure in some of the countries, such as India. Mechanical extraction of oil from the seed for biodiesel production results in a considerable amount of solid cake which is rich in proteins, minerals, amino acids, etc. [2]. Despite high nutritional quality, the *Jatropha* seed cake is non-edible because of several anti-nutritional factors like tannin, saponin, lectin, protease inhibitors, phytate and most importantly, the phorbol esters [2]. Phorbol esters are bioactive diterpene derivatives and have been recognised as a major toxic constituent in the *Jatropha* plant. It is present in varied amounts during different stages of maturity of the plant, ranging from 3.4–4.2 mg/g in *Jatropha* seed oil and 1.7–2.0 mg/g in *Jatropha* seed cake [3]. Co-carcinogenicity and tumour promotion are some of the biological effects of phorbol esters along with other biochemical and cellular effects [4]. At the same time, *Jatropha curcas* contains an alkaloid

**CONTACT** Shilpi Ahluwalia shilpi.ahluwalia@gmail.com Dr SSB University Institute of Chemical Engineering and Technology, Panjab University, Chandigarh 160014, India

© 2022 Informa UK Limited, trading as Taylor & Francis Group

in its latex which has anti-cancerous and antimicrobial properties which can be used externally for diseases like rheumatism, skin problems, etc. [5]. Several methods for removing phorbol esters from *Jatropha* seed cake have been reported, which would allow the use of detoxified meal as a food/feed supplement. Heat, radiations and different chemical treatments are being used to detoxify *Jatropha curcas* seed cake, but all these methods require a combination of one or other treatment which makes them expensive and time-consuming processes. Detoxification using *microorganisms* is still at the infancy stage, although few researchers have reported detoxification of *Jatropha* seed cake using fungi and bacteria by the solid state fermentation process. In a work performed by Anjali and Haresh [6], complete removal of phorbol esters from *Jatropha* seed cake resulted when solid state fermentation (SSF) was carried out with white rot fungi for 20 days. In another study conducted by Xing *et al.* [7], there was a 97% reduction in the phorbol ester content in 9 days when SSF of seed cake was performed with *Streptomyces fimicarius*. Chetna *et al.* [8] reduced the phorbol ester content to an undetectable level in 9 days when the seed cake was fermented (SSF) with *Pseudomonas aeruginosa* PseA strain at 30°C, pH 7 and 65% RH. In one such study, solid-state fermentation using *Enterobacter cloacae* Z11 was able to detoxify *Jatropha* seed cake by 51.6% under optimised conditions [9].

Detoxification by SSF has economic advantages, but at the same time, it is a time-consuming process. On the other hand, submerged fermentation is a rapid and best suited process for the microorganisms that require a high moisture content for their growth such as bacteria. In a work performed by Chang *et al.* [10], the *Bacillus* strains took 7 days under submerged fermentation for 100% degradation of phorbol esters. The present work aimed at detoxification of *Jatropha* seed cake by SMF using *Pseudomonas* sp. The fermentation process was optimised using the Taguchi technique which helps in studying a large number of factors and their effects on responses with a few experimental runs which makes this method superior to other similar statistical techniques including the Response Surface Methodology (RSM). The L18 orthogonal array was used to study the effect of fermentation conditions on the phorbol ester content of *Jatropha* seed cake. Multi-factorial optimisation of phorbol ester degradation by SMF using the Taguchi method has not been reported to date.

## 2. Materials and methods

### 2.1. Materials

Deoiled *Jatropha curcas* seed cake was obtained from the Biodiesel Centre, DTU Delhi, India. Culture media were procured from Hi Media Laboratories (Mumbai, India). Standard, i.e., Phorbol-12-myristate 13-acetate (Sigma Aldrich, ≥99 %), HPLC grade acetonitrile (Fisher Scientific) and HPLC grade methanol (SD fine) were used as received from the manufacturer.

### 2.2. Preparation of catalyst

Enzyme was prepared from *Pseudomonas aeruginosa* DS1 isolated in our laboratory by growing in enzyme production media containing (%) starch 2, $KH_2PO_4$ 0.5, $KNO_3$ 1.01, $NH_4Cl$ 0.535, $MgSO_4$ $7H_2O$ 0.001, $CaCl_2$ $2H_2O$ 0.01 and $Na_2HPO$ $12H_2O$ 0.8 at 1% (v/v) followed by incubation in a rotary shaker at 37°C for 24 h at 100 rpm. Enzyme was obtained by removing bacterial cells through centrifugation.

### 2.3. Submerged fermentation of deoiled Jatropha seed cake

Seed cake was autoclaved for 1 h and was added into supernatant (prepared above) for different combinations as per the L18 orthogonal array. The flasks were incubated in a rotary shaker set for each trial for different periods of time, and the phorbol esters were extracted. All the experimental runs were performed in triplicates.

### 2.4. Phorbol esters extraction from seed cake

Extraction of phorbol esters was performed with absolute methanol as described by Xing *et al.* [7]. Briefly, methanol was added to seed cake in a 7:1 ratio which was incubated in an ultrasonic bath for 15 min, and filtrate was collected. This extraction procedure was repeated twice, and all the methanol extract fractions were pooled. Further methanol was recovered using a rotary vacuum evaporator at 60 °C, and the extract was finally dried in a hot air oven at 40°C.

### 2.5. Quantification of phorbol esters

The untreated and treated seed cake samples were run on HPLC (Agilent-1260, USA) with a UV detector (Agilent G1315D MWD) and a reverse phase C18 column. The method adopted for quantification was in accordance with the study by Wink *et al.* [11], and solvent system used was water (A) and acetonitrile (B), programmed A:B as 60:40 (15 min), 25:75 (20 min) and 100: 0 (10 min) with a flow rate of 1.3 ml/min. All the HPLC experiments were performed at room temperature (25°C). PMA (phorbol 12-myristate 13-acetate) was used as an external standard. The characterisation was performed by Mass spectrometry analysis (LCMS, Waters-SYNAPT-G2) using the ES+ technique.

### 2.6. Optimisation using the Taguchi method

The process parameters which influence phorbol ester degradation in the most significant way were selected (through expert opinion, review of literature and pilot scale study) for investigating their effect by the submerged fermentation process. The process parameters considered were pH, percentage of the seed cake, time, temperature and frequency of rotation (RPM). The design parameters and their different levels considered for optimisation of the study by the Taguchi method are shown in Table 1. The effect of these variables was identified using the ANOVA based on the S/N ratio. The Taguchi L18 orthogonal array was followed for experimentation at each level, and the response was observed in terms of phorbol ester degradation.

### 2.7 Analysis of the experimental data

The observations of fermentation runs were analysed using the software MINITAB® and Design Expert. In the present work, phorbol ester degradation was considered as higher-the-better type because a lower value of phorbol ester degradation may result in

**Table 1.** Raw data and S/N ratios of the Taguchi L18 Orthogonal array.

| Process parameters | | | | | | Response | | | |
|---|---|---|---|---|---|---|---|---|---|
| S No. | pH | Seed cake (%) | Time (h) | Temperature (°C) | RPM | R1 | R2 | R3 | S/N ratio |
| 1 | 6.5 | 1 | 9 | 25 | 100 | 49.7 | 43.4 | 47.9 | 33.40 |
| 2 | 6.5 | 1 | 12 | 35 | 125 | 78.6 | 88.4 | 77.1 | 38.16 |
| 3 | 6.5 | 1 | 15 | 45 | 150 | 89.4 | 90.8 | 87.1 | 38.99 |
| 4 | 6.5 | 5 | 9 | 25 | 125 | 32.6 | 31.3 | 30.1 | 29.91 |
| 5 | 6.5 | 5 | 12 | 35 | 150 | 67.3 | 65.5 | 67.3 | 36.48 |
| 6 | 6.5 | 5 | 15 | 45 | 100 | 74.8 | 69.1 | 67.7 | 36.94 |
| 7 | 6.5 | 9 | 9 | 35 | 100 | 51.1 | 49.9 | 52.1 | 34.15 |
| 8 | 6.5 | 9 | 12 | 45 | 125 | 47.2 | 49.9 | 48.7 | 33.73 |
| 9 | 6.5 | 9 | 15 | 25 | 150 | 32.8 | 37.1 | 37.8 | 31.05 |
| 10 | 7.5 | 1 | 9 | 45 | 150 | 81.3 | 87 | 89.5 | 38.66 |
| 11 | 7.5 | 1 | 12 | 25 | 100 | 89.4 | 88.9 | 86.5 | 38.91 |
| 12 | 7.5 | 1 | 15 | 35 | 125 | 98.2 | 97.1 | 97.9 | 39.80 |
| 13 | 7.5 | 5 | 9 | 35 | 150 | 64.5 | 66.4 | 62.6 | 36.18 |
| 14 | 7.5 | 5 | 12 | 45 | 100 | 69.2 | 67.8 | 68.8 | 36.73 |
| 15 | 7.5 | 5 | 15 | 25 | 125 | 49.6 | 48.7 | 48.5 | 33.79 |
| 16 | 7.5 | 9 | 9 | 45 | 125 | 64.7 | 61.1 | 60.1 | 35.83 |
| 17 | 7.5 | 9 | 12 | 25 | 150 | 62.4 | 63.2 | 61.6 | 35.90 |
| 18 | 7.5 | 9 | 15 | 35 | 100 | 71.8 | 72.9 | 75.2 | 37.30 |
| **Total** | | | | | | 1174.6 | 1178.5 | 1166.5 | 645.9 |
| **Mean** | | | | | | | 65.18 | | |

R1, R2 and R3 refer to phorbol ester degradation values of each trial with three repetitions.

inadequate detoxification of the seed cake which is not desired if the seed cake is to be used for edible purposes. The S/N ratio for higher-the-better type of response was calculated [12] using the standard formula.

## 3. Results and discussion

### 3.1. Phorbol ester content in seed cake

HPLC chromatograms of *Jatropha* seed cake extract showed the presence of a major phorbol ester peak, and the total phorbol ester content was calculated by summing up the area of the PE peak which was found to be 1.5 mg/g. The precursor ion analysis of PMA and phorbol esters at m/z 311 represented the protonated phorbol nucleus ($[C_{20}H_{22}O_3^+ H]^+$) which gave the confirmation of the presence of phorbol esters in the raw *Jatropha* seed cake (Figure 2a). Enzyme catalysed degradation of phorbol esters resulted in 99.7% degradation in 15 h. The conditions for phorbol ester degradation were further optimised by the Taguchi model.

### 3.2. Optimisation of the process by Taguchi methodology

The average values of phorbol ester degradation, the raw data and S/N ratio for each parameter (three different levels) were calculated and plotted.

In Figure 1(a), the variation in phorbol ester degradation with respect to the pH of the production media can be observed. These results show that phorbol ester degradation at pH 6.5 is comparatively lesser but increases when the pH of the medium is 7.5. Most bacterial lipases work better under alkaline as well as acidic conditions [13]; however,

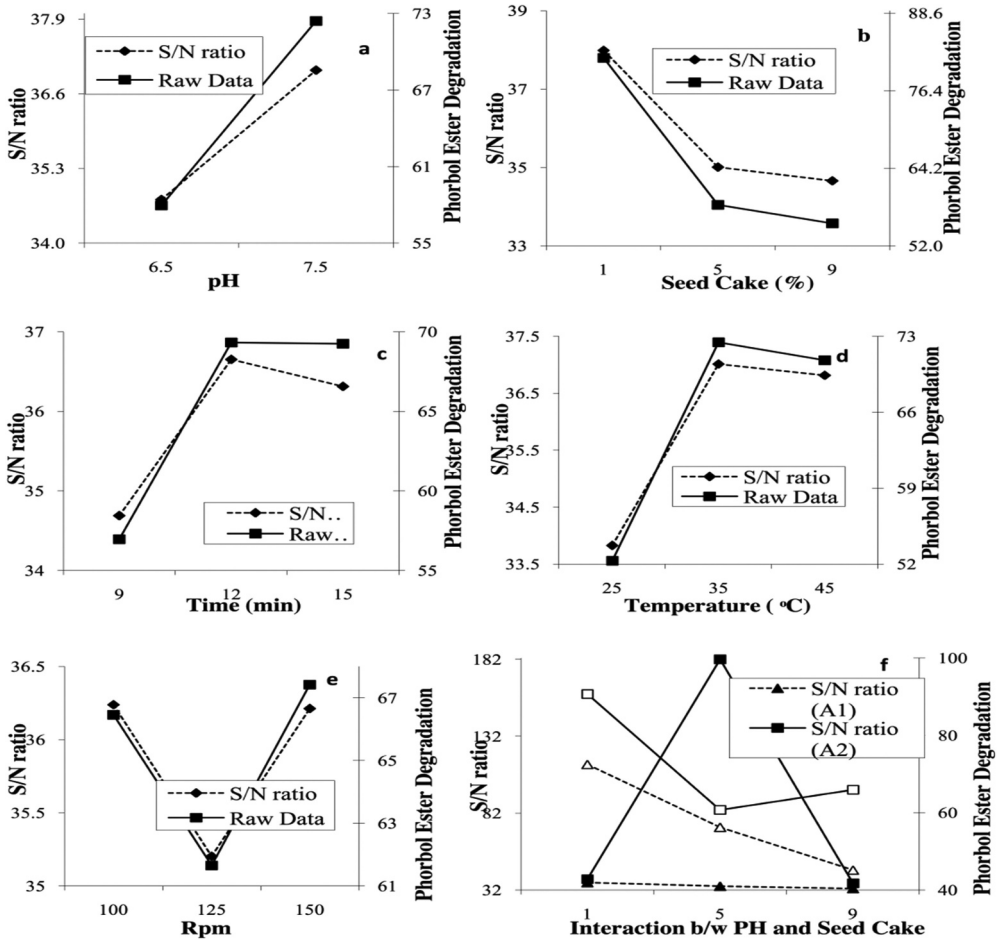



Figure 1. LC-MS chromatogram of the untreated sample (a) and detoxified seed cake (b).

*Pseudomonas* strain shows maximum lipase activity at slightly alkaline pH [14]. Higher degradation of phorbol esters at pH 7.5 can be attributed to optimum lipase activity by our strain *Pseudomonas aeruginosa* DS1 which indicates that the maximum the enzyme activity, the more will be the phorbol ester degradation. The higher S/N ratio value for pH 7.5 indicates that it may be considered as an optimal level of pH for attaining higher phorbol ester degradation.

From Figure 1(b), it can be observed that the phorbol ester degradation is more when the seed cake is 1 %. However, it is lower when the percentage seed cake is 5 % and lowest for 10 %. A higher S/N ratio value for the seed cake (%) when fixed at 1 shows that it could be regarded as the optimum percentage for obtaining higher phorbol ester degradation. As the enzyme units (inoculum size) are equal for all the concentrations of the seed cake, enzyme units will be the limiting factor for the degradation. Therefore, to enhance the degradation rate of higher concentrations of seed cake, greater volumes of biomass need to be added. Furthermore, if higher inoculum is used, the detoxification process may be shortened to a certain extent

(further studies required). Also, certain reports have discussed the effect of phorbol ester on the integrity of cellular proteins [15,16]. This could possibly be another reason that can be attributed to lower degradation of phorbol esters at higher concentrations of seed cake and requires further studies.

From Figure 1(c), it can be noted that phorbol esters degradation is more when the seed cake is incubated for 12 h and further, the rate of degradation slowed down. Although complete detoxification of the seed cake was achieved in 15 h, detoxification at 12 h can be regarded as the optimal time since the phorbol ester content in the seed cake was found to be within the permissible limit of 0.09 mg/g [17]. Figure 2(b) shows the LC-MS chromatogram of detoxified seed cake in which no phorbol esters peak appeared. Most of the researchers have worked with the biomass for detoxification, while our work mainly involved the enzymes in the broth after centrifugation to remove the cells. In a work performed by Azhar *et al.* [18], the phorbol ester concentration decreased with the incubation time because of some enzymes released by strain that degraded diterpene esters, i.e. phorbol esters. According to Chandra *et al.* [13], lipases and esterases are the enzymes responsible for hydrolysing these ester bonds. So, it can be stated that the incubation time has significant effects on the levels of phorbol esters in the seed cake in the presence of biomass or free enzymes. However, as seen in Figure 1(c), there is a slight decline in the phorbol ester degradation rate at 15 h; this may be due to the change in pH because of the secondary metabolites produced during fermentation which further affects the enzyme activity and the degradation of phorbol esters.

Figure 1(d) shows the variation in the phorbol ester degradation rate with respect to different temperatures of the process. Results show that the phorbol ester degradation is maximum at 35°C and starts to decline when the temperature is increased to 45°C. However, on reducing temperature from 35°C to 25°C, the degradation decreases. These results are in agreement with those in the study by Devappa *et al.* [19] who have observed that the increase in temperature to a certain point results in the reduction in the phorbol ester content in *Jatropha* oil and the rate starts decreasing with a further increase in temperature. The S/N ratio, however, is maximum at 35°C, comparable when the temperature is 45°C and decreases when the temperature is decreased to 25°C. Our findings indicate that the process is certainly catalysed by the enzymes that have temperature optima of 35 °C beyond which they slow down due to the change in the kinetic behaviour and also due to the structural violation (denaturation) of the enzymes at higher temperatures. At low temperatures, most of the enzymes show low activity and our enzyme also behaved in a similar manner.

The S/N ratios depict similar values at all the three levels of RPM, indicating that the response is not affected by variations in the speed. From Figure 1(e), it can be observed that the phorbol ester degradation is equal at 100 and 150 rpm. The RPM shows an insignificant effect on raw data and S/N ratio. Thus, any of the three levels of RPM can be selected for the study taking into consideration the user requirement and the cost.

As phorbol ester degradation comes under the higher-the-better type category, higher values of degradation were required. Each parameter at level, i.e. A-2, B-1, C-2 and D-2, may result in maximum degradation of phorbol esters [Figure 1(a-e)]. However, it can be observed from Figure 1 that the highest values of mean responses correspond to the highest values of S/N ratios.

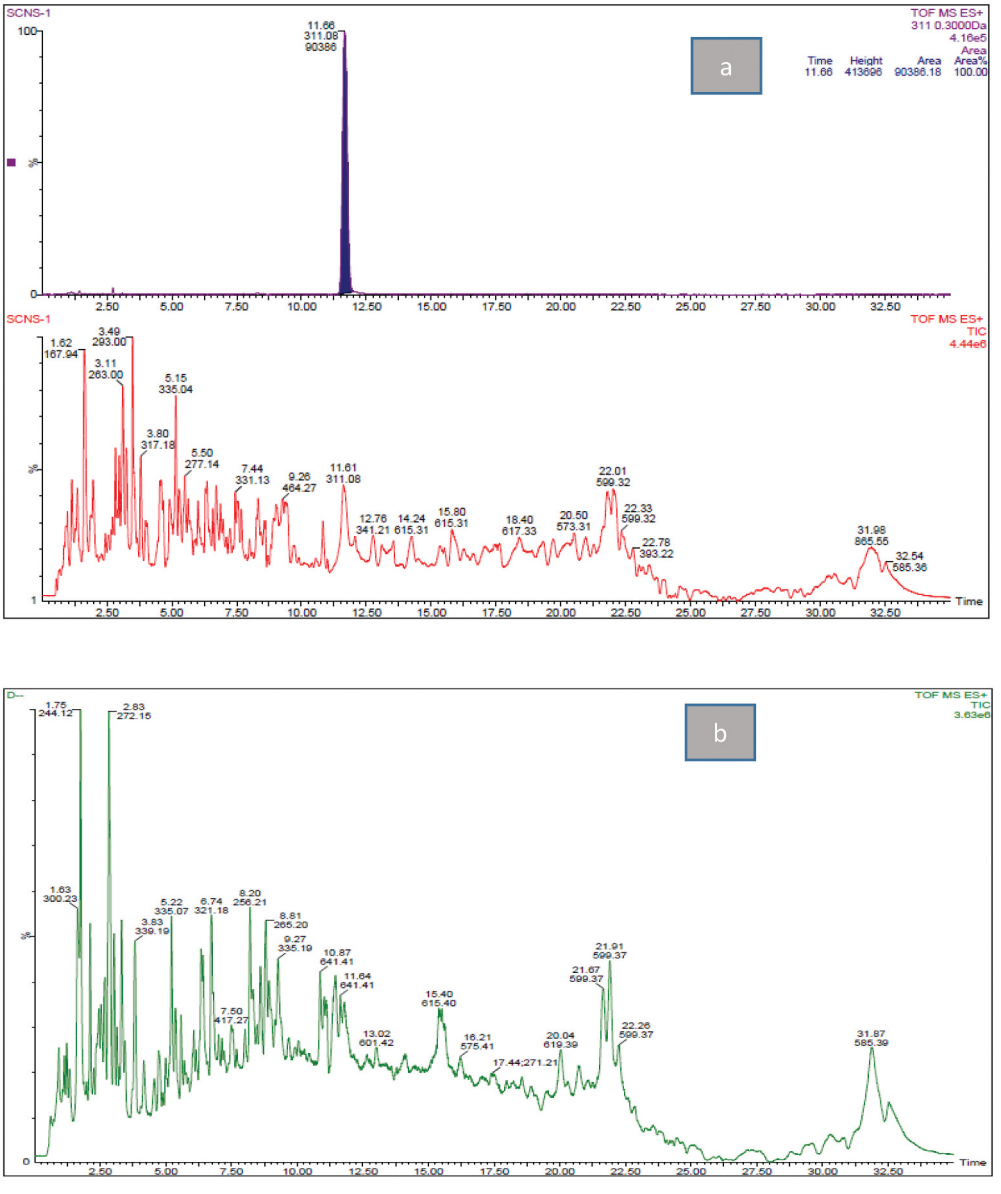



**Figure 2.** Variation of phorbol ester degradation (raw data and S/N ratio) with process parameters: (a) pH, (b) seed cake (%), (c) time (h), (d) temperature (°C), (e) rpm and (f) interaction between pH and seed cake.

ANOVA was performed for studying the significance of input parameters in the degradation of phorbol esters. The ANOVA of raw data as well as S/N data for phorbol ester degradation is shown in Table 2. From the table, it was observed that pH, seed cake (%), time and temperature had a significant effect on phorbol ester degradation values, while RPM had an insignificant effect. The percentage contribution of seed cake (%) was highest (39.39%), followed by temperature (24.02 %), pH (15.11%) and time (9.47%).

Table 2. Pooled ANOVA (raw data and S/N ratio)

| Source | Sum of Squares | | Degree of Freedom | Variance | | F-ratio | | P (%) Percentage contribution | |
|---|---|---|---|---|---|---|---|---|---|
| | Raw data | S/N data | | Raw data | S/N data | Raw data | S/N data | Raw data | S/N data |
| pH | 2819.56 | 22.87 | 1 | 2819.56 | 22.87 | 70.45* | 18.30* | 15.11 | 16.80 |
| Seed cake (%) | 7327.77 | 40.20 | 2 | 3663.89 | 20.10 | 91.54* | 16.08* | 39.39 | 29.23 |
| Time | 1822.91 | 13.21 | 2 | 911.46 | 6.60 | 70.45* | 5.28* | 9.47 | 8.32 |
| Temp | 4498.24 | 38.22 | 2 | 2249.12 | 19.11 | 56.19* | 15.29* | 24.02 | 27.76 |
| RPM | 168.17 | 4.19 | 2 | 84.08 | 2.09 | 2.10 | 1.67 | - | - |
| Error | 1761.10 | 9.99 | 8 | 40.02 | 1.25 | - | - | 12.01 | 17.82 |
| Total | 18,397.7 | 128.68 | 17 | - | - | - | - | 100 | 100 |

*Significant at the 95% confidence level.

### 3.3. Estimation of optimum levels

The significant parameters (A, B, C and D) at selected levels were used for predicting the optimum value ($A_2$, $B_1$, $C_2$ and $D_2$) of response (phorbol ester degradation), and the mean of the response factor was calculated as follows [20]:

Average value of phorbol ester degradation at **pH** ($\overline{A}_2$) = 72.40, **Seed cake** ($\overline{B}_1$) = 81.57, **Time** ($\overline{C}_2$) = 69.32 and **Temperature** ($\overline{D}_2$) = 72.44. The overall mean of phorbol ester degradation ($\overline{T}_{ped}$) = 65.18 (from Table 2).

The predicted optimum value of phorbol ester degradation ($\mu_{ped}$) has been calculated as

$$\mu_{ssr} = \overline{A}_2 + \overline{B}_1 + \overline{C}_2 + \overline{D}_2 - 3\overline{T}_{ped} = 100.19,$$

where $F_{0.05}$ (1, 46) = 4.05 (Tab F value) and $CI_{POP}$ = ±5.48, $CI_{CE}$ = ±9.17.

The 95% confidence intervals ($CI_{POP}$ and $CI_{CE}$) of the predicted optimum scatter of phorbol ester degradation are $CI_{POP}$: 94.71< $\mu_{SSR}$ <105.67 and $CI_{CE}$: 91.02< $\mu_{SSR}$ <109.36.

The optimum values of the process parameters for phorbol ester degradation are as follows:

pH (A, 2nd level) = 7.5, seed cake (B, 1st level) = 1%, time (C, 2nd level) n = 12 h and Temp (D, 2nd level) = 35 °C.

### 3.5 Conformation experiments

Three conformational experiments for the response characteristic, i.e. phorbol ester degradation, were performed at optimised levels of process parameters for the validation of results. The pH was set at the second level, % seed cake was set at the first level, time was kept at the second level, temperature was set at the second level and rpm was set at the first level. The average phorbol ester degradation was found to be 95.2% (0.08 mg/g) which was within the confidence interval of the predicted optima and also within the permissible limit.

## 4. Conclusion

Detoxifying *Jatropha* seed cake with complex toxic constituents is a great challenge for the industry. Various physical and chemical methods have been employed to detoxify the seed cake, but they contribute in making the process expensive, and disposal of toxins raises an environmental and health concern. The Taguchi method was used to identify the significant parameters in biological detoxification of phorbol esters present in non-oil *Jatropha* residual material. The various parameters studied and found to be significant based on the F-ratio and p-value of the design were seed cake (1%), temperature (35 °C), pH (7.5) and time (12 h). These results can further be extended to other environmental factors to make the detoxification process more efficient to be applied industrially and generate better feed supplements. Most of the earlier research studies involved the use of live microbial cells, while this study was conducted with enzyme preparations and thus have an upper hand in developing commercial processes with low cost of down-streaming. Although this is an attempt to improve the biotechnological application in detoxification of phorbol esters, the avenues are not limiting.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## References

[1] Seela CR, Shankar BR. Emulsified nano $Al_2O_3$ – jatropha methyl ester blends: application in variable compression ratio engine. World J Eng. 2020;17(5):733–737.
[2] Ahluwalia S, Bidlan R, Shrivastav AK, et al. Optimization of protein extraction from detoxified Jatropha seed cake using response surface methodology and amino acid analysis. Int J Environ Sci Technol. 2020;17:1087–1100.
[3] Mbako J, Clever K, Jerekias G. Quantification of Phorbol-12-myristate 13-acetate in Jatropha seed oil and cake at different stages of fruit maturity. Int J Environ Stud. 2022;79 (1):88–97.
[4] Wakandigara A, Nhamo LRM, Kugara J, et al. Mechanisms of phorbol ester toxicity, determined by molecular modelling. Int Res J Pure Appl Chem. 2020;*21*(18):10–24.
[5] Sharma AK, Gangwar M, Kumar D, et al. Phytochemical characterization, antimicrobial activity and reducing potential of seed oil, latex, machine oil and presscake of *Jatropha curcas*. Avicenna J Phytomedicine. 2016;6(4): 366–375
[6] Anjali B, Haresh K. Phorbol ester degradation in *Jatropha* seed cake using white rot fungi. Biotechnol. 2014;4:447–450.
[7] Xing HW, Lingcheng O, Liang-Liang F, et al. Detoxification of *Jatropha curcas* kernel cake by a novel *Streptomyces fimicarius* strain. J Hazard Mater. 2013;260:238–246.
[8] Chetna J, Priyanka M, Khare SK. Degradation of phorbol esters by *Pseudomonas aeruginosa* PseA during solid-state fermentation of deoiled *Jatropha curcas* seed cake. Bioresour Technol. 2011;102(7):4815–4819.
[9] Zhao YN, Wang HF, Liu JX. Detoxification of *Jatropha curcas* seed cake by a new soil-borne *Enterobacter cloacae* strain. Lett Appl Microbiol. 2018;67(2):197–204.
[10] Chang CF, Weng JH, Lin KY, et al. Phorbol esters degradation and enzyme production by *Bacillus* using *Jatropha* seed cake as substrate. Int J Environ Pol Remed. 2014;2(1): 32–36
[11] Wink M, Grimm C, Koschmieder C, et al. Sequestration of phorbol esters by the aposematically coloured bug *Pachycoris klugii* (Heteroptera: scutelleridae) feeding on *Jatropha* curcas (*Euphorbiaceae*). Chemoecol. 2000;10:179–184.

[12] Walia RS, Shan HS, and Kumar P, et al. Parametric optimization of centrifugal force-assisted abrasive flow machining (CFAAFM) by the Taguchi method. Mater Manuf Process. 2006;21(4):375–382.
[13] Chandra P, Enespa, Singh R, et al. Microbial lipases and their industrial applications: a comprehensive review. Microb Cell Factories. 2020;19:169.
[14] Priji P, Sajith S, Faisal PA, et al. *Pseudomonas* sp. BUP6 produces a thermotolerant alkaline lipase with *trans*-esterification efficiency in producing biodiesel. 3 Biotech. 2017;7(6):369.
[15] Ehsan O, Norhani A, Wan ZS, et al. Dose response effects of phorbol esters isolated from *Jatropha* meal on rumen microbial activities. Asian J Anim Vet Adv. 2014;9(1):37–46.
[16] Rath NC, Gupta A, and Liyanage R, et al. Phorbol 12-Myristate 13-Acetate-induced changes in chicken enterocytes. Proteomics Insights. 2019; 10: 1178641819840369.
[17] Gogoi R, Niyogi UK, Tyagi AK. Reduction of phorbol ester content in *Jatropha* cake using high energy gamma radiation. J Radiat Res Appl Sci. 2014;7:305–309.
[18] Azhar N, Norhani A, Wan Z, et al. Detoxification of toxic phorbol esters from Malaysian *Jatropha curcas* Linn. kernel by *Trichoderma* spp. and endophytic fungi. Int J Mol Sci. 2014;15:2274–2288.
[19] Devappa RK, Makkar HPS, Becker K. Optimization of conditions for the extraction of phorbol esters from *Jatropha* oil. Biomass Bioenerg. 2010b;34:1125–1133.
[20] Goyal T, Walia RS, Sidhu TS. Study of coating thickness of cold spray process using Taguchi method. Mater Manuf Process. 2012;27(2):185–192.

Wireless Personal Communications
https://doi.org/10.1007/s11277-022-09880-y

# Performance Analysis of CSS Over α–η–μ and α–κ–μ Fading Channel Using Clustering-Based Technique

**Sandeep Kumar[1] · Puspraj Singh Chauhan[2] · Rahul Bansal[3] · Manpreet Kaur[4] · Rajesh Kumar Yadav[4]**

Accepted: 29 May 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

**Abstract**
Cooperative spectrum sensing (CSS) is an efficient method, which is used to detect the presence/absence of primary user (PU) signal in the received spectrum at the secondary user (SU). Clustering is a vector quantization technique that classifies the observed signal into some fixed number of clusters depending upon the nearest mean. In this work, the *K*-mean clustering algorithm is used to classify the SUs received signal into the signal available class or signal unavailable class. The channel between the PU and SUs is assumed to be generalized (α–η–μ and α–κ–μ) fading channels. The performance of the proposed algorithm is compared with the conventional method of CSS with energy feature vector and found to be superior to that.

**Keywords** Spectrum sensing · Energy feature vector · Clustering · Generalized fading · Collaborative sensing · Vector quantization

## 1 Introduction

Due to the exponential proliferation in the usage of wireless traffic networks through the internet of things (IoT) and device-to-device communication applications, a severe shortfall of the radio spectrum is being faced worldwide. The main reason for the underlying problem is the fact that the available spectrum resources are not being used consummately. Cognitive radio (CR) network is a promising solution to this subsisting predicament [1]. There are two types of users in the CR network, the licensed user, which is the primary user (PU) and the un-licensed user, which is the secondary user (SU) [2]. SU can access the licensed spectrum when it is not being used by the corresponding PU, thus effectively abating the spectrum shortage issue. In the working of the CR network, SU has to be prudent in ensuring that zero interference is caused to the PU while accessing the spectrum of

✉ Manpreet Kaur
manpreetkaur_2k19phdco06@dtu.ac.in

[1] Central Research Laboratory, BEL, Ghaziabad, India

[2] Department of ECE, Pranveer Singh Institute of Technology, Kanpur, India

[3] ECE Department, AKGEC, Ghaziabad, India

[4] Department of CSE, Delhi Technological University, Delhi, India

the PU. The biggest challenge in the network is to sense the presence of available spectrum and hence spectrum sensing (SS) is the most important function in the CR network. Out of the various avenues used for SS, energy detection is the most widely studied in literature due to its ease of implementation and low complexity [3].

Most of the time, due to the unpredictable nature of the wireless channel, the detection of the void spectrum for an individual SU is a challenging task. To overcome this challenge, a more efficient technique called cooperative spectrum sensing (CSS) has been used to improve the SS performance of the system. In CSS, the multiple SUs are used to sense the same spectrum at the same time and their individual decisions are fused at the fusion center to give the final decision [4]. In [2], the CSS performance of the energy detector (ED) over $\alpha-\eta-\mu$ fading channel has been analyzed in terms of receiver operating characteristic (ROC) and area under the receiver operating characteristic curve (AUC). The work has shown that the performance of ED improved with the increase in $\alpha$, $\eta$, $\mu$ fading parameters. The performance of ED over $\alpha$–$\eta$–$\mu$/IG and $\alpha$–$\kappa$–$\mu$/IG has been evaluated in [3], while the detection performance over $\alpha$–$\eta$–$\mu$/gamma and $\alpha$–$\kappa$–$\mu$/gamma composite fading channels is studied in [5]. The work in [3] and [5] is extended for CSS and the performance improvement of the system using a cooperative approach has been demonstrated. Recently, the machine learning and deep learning techniques have been used to improve the performance of CSS [6, 7]. In [8], an approach using null space pursuit algorithm (NSP) and fuzzy *C*-means clustering algorithm (FCM) is proposed for CSS. To improve the CSS performance, a fuzzy c-means clustering algorithm is proposed in [9], and the improvement in the performance of the system has been demonstrated. The *K*-mean clustering algorithm for different features of the eigenvalues and the matrix trace is studied in [10]. The work in [10] has also shown that eigenvalues-based and the matrix trace-based features have better efficiency in comparison to the signal energy as a sole feature for SS. The analysis using multiple-antenna CSS based on the wavelet transform and Gaussian mixture model (MAWG) is proposed in [11] to improve the sensing performance in fading channel. The work also illustrates issues like path losses and shadowing, which are very frequently faced in single SU-based SS has also been provided. In the complex electromagnetic environment, the noise is very uncertain and thus, the SS becomes a big challenging task for these environmental conditions. A study for such a scenario using empirical mode decomposition algorithm (EMA) and information geometry is presented in [12]. The performance of SS at a low signal-to-noise ratio (SNR) is poor and in this regard, an approach to improve SS by modifying the support vector machine (SVM) algorithm for low SNR scenario is analyzed in [13]. Similarly, in [14], the empirical mode decomposition algorithm and *K*-mean clustering algorithm are proposed to improve SS under a low SNR scenario. A proposal for the allocation of the spectrum of multiple channels to the SU using the spectral handoff method is explained in [15] and an improvement in the assignment of a higher transmission average bandwidth for each secondary user is shown.

The $\alpha$–$\eta$–$\mu$ and $\alpha$–$\kappa$–$\mu$ are generalized fading channels that have inherent capability to subsume many other fading channel models [16]. The machine learning-based CSS performance analysis over $\alpha$–$\eta$–$\mu$ and $\alpha$–$\kappa$–$\mu$ generalized fading channel is not present in the open literature. Motivated by this, we have presented the analysis of ED over $\alpha$–$\eta$–$\mu$ and $\alpha$–$\kappa$–$\mu$ generalized fading channels using *K*-mean clustering-based CSS. The derived expressions can be considered as general expressions for other fading distributions by using the appropriate values of the fading parameters.

The organization for the paper is as follows. The system and channel model is described in Sect. 2. In Sect. 3, the performance analysis over fading channel is presented. In Sect. 4, the numerical results over different parameters are shown and discussed followed by the conclusion in Sect. 5.

## 2 System Model

### 2.1 Cognitive Radio and Spectrum Sensing

In the cognitive radio network, there are two types of users, PU and SU. PU being the licensed user, has the highest priority of using the allotted spectrum than the unlicensed SU [17]. SU can access the frequency band of PU when that is not being utilized by the PU. In case the SU, say *SUa* is using the frequency of band of some PU say *PUa*, and suddenly *PUa* wants to access his band, then *SUa* has to free the band of *PUa* and needs to switch to some other available bands. It means SU needs to sense the available spectrum continuously so that it can enjoy the unlicensed service provided by the network [6].

Usually, PU and SU are located at different locations and the signal propagation between them may be severely affected by small-scale and large-scale fadings [18]. Due to this, spectrum sensing becomes a difficult task for the SU. In CRN, one condition imposed on SU is that it can use the spectrum of PU without interfering in the operation of PU [19]. So, the SU cannot afford to make a wrong decision about the presence of a PU signal, as by doing this it will breach its restricted condition. To improve the sensing capability and the decision- ability about the availability of the PU signal, the cooperative philosophy is being used in literature [2]. The cognitive radio scenario is shown in Fig. 1, where the PU signal reaches the SUs after experiencing different propagation phenomenon. All SUs receive the samples of the signal corresponding to the band of PU.

### 2.2 Energy Feature Vector-Based CSS

The energy of the received signal is considered as a feature vector that helps the system to perform spectrum sensing. It is indeed appreciated that the energy vector is analogous to the feature vector as used in machine learning context. The normalized energy estimated at the $n$th SU is given [6].

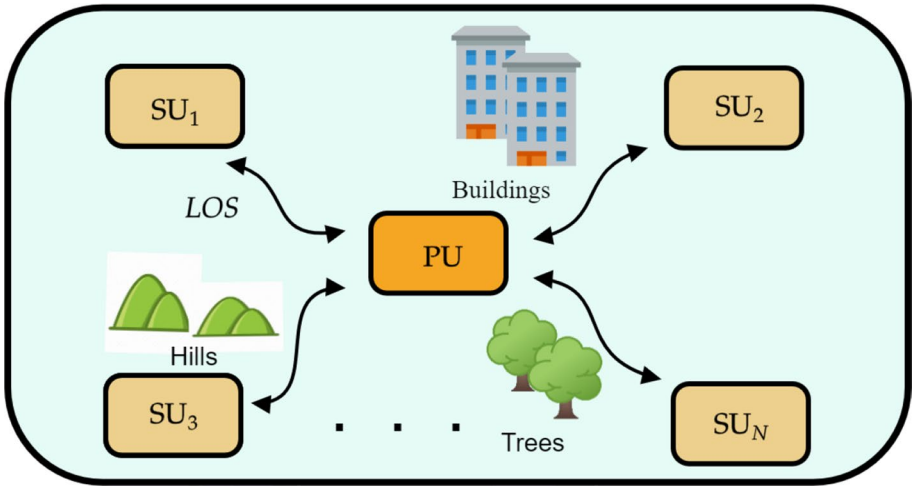



**Fig. 1** Cognitive radio system model

$$Y_n = \frac{1}{M}\sum_{i=1}^{M}\left|Z_n(i)\right|^2 \tag{1}$$

where $M$ is the number of samples and $Z_n(i)$ is the $i$th sample of the received signal at the $n$th SU.

$$Z(i) = \begin{matrix} w(i);\, H_0 \\ h * s(i) + w(i);\, H_1 \end{matrix} \tag{2}$$

here $w(i)$, $s(i)$ and $h$ are additive white Gaussian noise (AWGN), the sample of the transmitted symbol and channel gain respectively.

### 2.2.1 Traditional CSS

In traditional CSS, $Y(n)$ is compared with a predefined threshold ($\lambda$) and the individual decision about the presence/absence of the PU signal is made at the SU [5]. The sample of the received signal can be formulated in the form of a binary hypothesis problem [20]. Under $H_0$, $Y$ follows a central chi-square distribution with $\sigma^2 = 1$ and $2d$ degrees of freedom. Similarly, under $H_1$, $Y$ follows a non-central chi-square distribution with $\sigma^2 = 1$ and $u = 2\gamma$ [3]. Thus, PDF of $Y$ can be written by [5] as

$$f_Y(y) = \begin{cases} \dfrac{1}{2^n\Gamma(n)} y^{n-1} \exp\left(\dfrac{-y}{2}\right);\, H_0 \\ \dfrac{1}{2}\left(\dfrac{y}{2\gamma}\right)^{n/2-1/2} \exp\left(\dfrac{-y-2\gamma}{2}\right) I_{n-1}\left(\sqrt{2y\gamma}\right);\, H_1 \end{cases} \tag{3}$$

where, $n = TW$ is the time-bandwidth product which is nothing but the number of time samples and it must be integer, $T$ is the time interval of the integrator and $W$ is the bandwidth of the system. $I_v(.)$ is the modified Bessel's function and $\Gamma(.)$ is gamma function. $\gamma = |h|^2 E_S^2/N_0$ is the received SNR with $E_S$ be the signal energy and $N_0$ is the one-sided power spectral density (PSD). In energy detection based CSS, each one of the $N$ SU nodes sends their autonomous one-bit decision to the fusion centre (FC) which gives a rule-based combined decision about PU signal. Figure 2 shows, how the individual decisions contribute in the cooperative decision making.

## 2.3 K-Mean Clustering-Based CSS

As discussed in the previous section, the classifier is the core component and the decision-maker in clustering-based CSS algorithm. Prior to the online decision-making, the classifier needs to be trained on similar feature vectors. Once the training of the classifier is complete it is ready for online decision making. The classifier decides the presence/absence of PU signal as shown in Fig. 3. The steps for training and testing in the classifier are given in Table 1.

$$A^* = \begin{cases} -1, \text{ channel is unavailable} \\ +1, \text{ channel is available} \end{cases} \tag{4}$$

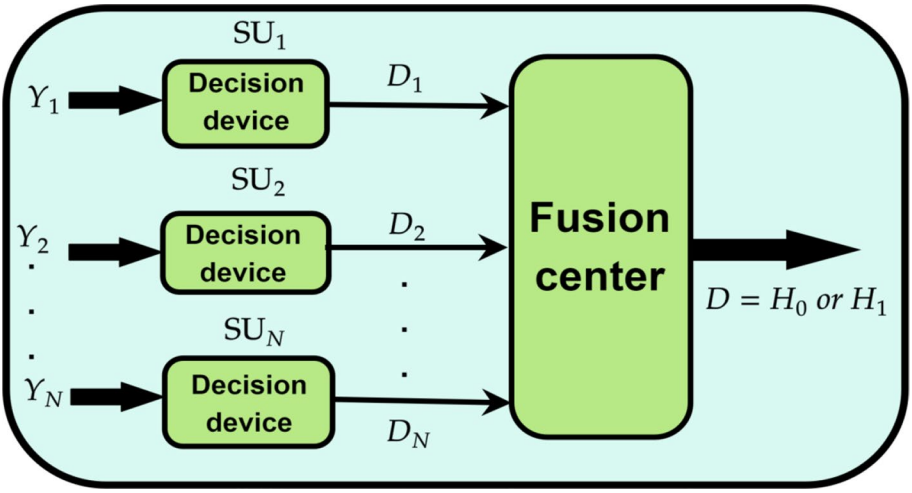



**Fig. 2** Traditional CSS model

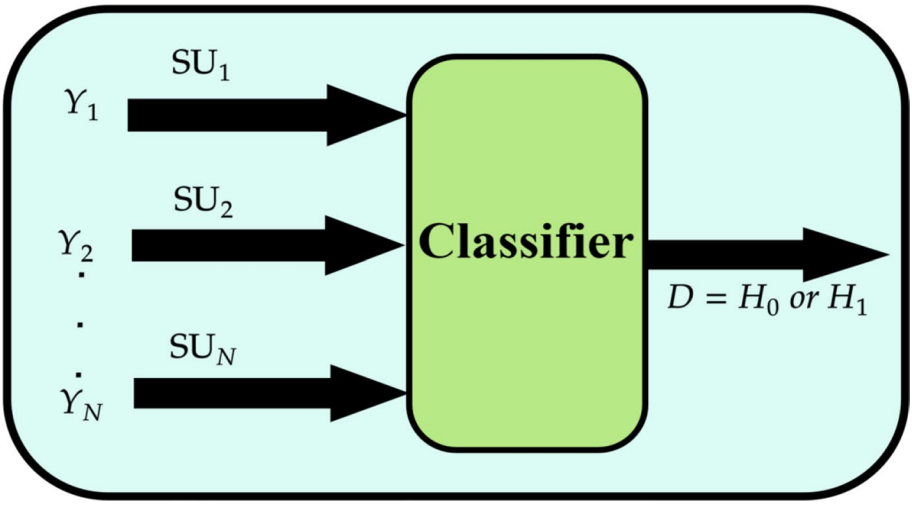



**Fig. 3** Clustering-based CSS model

$$\frac{\left\|\tilde{Y} - C_1\right\|}{\min_{k=2,3,\ldots K}\left\|\tilde{Y} - C_k\right\|} \geq \xi \tag{5}$$

## 3 Channel Model and Performance Analysis

### 3.1 Channel Model

The α–η–μ and α–κ–μ are generalized multipath fading distributions which are used to model practical signal propagation scenarios of NLOS and LOS, respectively [3]. Most of

**Table 1** K-mean clustering: Training and testing of energy vectors

| |
|---|
| Steps for training |
| 1. Take a sufficiently large number of the non-biased energy vector data set of length, $L$ i.e. $Y = [Y_1, Y_2, ...Y_L]$ |
| 2. Group the training data set from all $N$ SUs into the $K$ non-overlapping clusters |
| 3. Find the centroid of each clusters say $C_k$, where $k = 1, 2, ...K$ |
| 4. In this scenario, cluster 1 belongs to $H_0$, i.e. only noise is present in the received signal |
| Steps for testing |
| 5. In the testing, the energy vectors are classified into two classes as defined in (3) |
| 6. The metric parameter ($\xi$) based criterion for deciding the channel unavailable class for an energy vector is given in Eq. (5) |

the other multipath fading models such as Rayleigh, Ricean, Nakagami, Weibull, α–μ, η–μ and κ–μ are special cases of these generalized distributions. The probability density function (PDF) of received SNR, $\gamma$ over α–η–μ fading channel is given by [2] as

$$f_\gamma(\gamma) = \frac{\sqrt{\pi}\alpha\mu^{(\mu+1/2)}h^\mu\gamma^{(\alpha/2)(\mu+1/2)-1}}{\Gamma(\mu)H^{(\mu-1/2)}\overline{\gamma}^{(\alpha/2)(\mu+1/2)}}\exp\left(-\frac{2\mu h\gamma^{(\alpha/2)}}{\overline{\gamma}^{(\alpha/2)}}\right)I_{(\mu-1/2)}\left(\frac{2\mu H\gamma^{(\alpha/2)}}{\overline{\gamma}^{(\alpha/2)}}\right), \gamma \geq 0 \tag{6}$$

where $\alpha$ and $\mu$ are used to characterize the non-linearity and non-homogeneity of the propagation channel. $\eta$ is the ratio of scattered-wave power of the in-phase and quadrature components. The parameters $h$ and $H$ are defined in terms of $\eta$ in two formats as defined in [2] and $\overline{\gamma}$ is the average received SNR. Similarly, the PDF of α–κ–μ fading channel is defined by [21] as

$$f_\gamma(\gamma) = \frac{\mu\alpha k^{\left(\frac{1-\mu}{2}\right)}(1+k)^{\left(\frac{1+\mu}{2}\right)}\left(\sqrt{\frac{\gamma}{\overline{\gamma}}}\right)^{\left(\frac{\alpha}{2}+\frac{\alpha\mu}{2}-1\right)}}{2\sqrt{\gamma\overline{\gamma}}\exp\left[\mu\left(k+(1+k)\left(\sqrt{\frac{\gamma}{\overline{\gamma}}}\right)^\alpha\right)\right]}I_{(\mu-1)}\left(2\mu\sqrt{k(k+1)\left(\sqrt{\frac{\gamma}{\overline{\gamma}}}\right)^\alpha}\right) \tag{7}$$

where $k$ represents the ratio of the total power of the dominant component and the scattered wave components.

### 3.2 Performance Evaluation

The probability of detection ($P_d$) and probability of false alarm ($P_{fa}$) are the performance parameters that are used to validate the performance of a spectrum sensing detector [22]. In this section, we will see the details of these parameters for both traditional systems and clustering-based systems.

#### 3.2.1 CSS Using the Traditional Approach

In the traditional approach of CSS, the average probability of detection, $\overline{P_d}(\lambda)$ and probability of false alarm,$P_{fa}$ for a fading channel are expressed by [2] as

$$\overline{P_d}(\lambda) = E[P_d(\gamma, \lambda)] = \int_0^\infty P_d(\gamma, \lambda) f_\gamma(\gamma) d\gamma \tag{8}$$

$$P_{fa}(\lambda) = \frac{\Gamma(n,\ \lambda/2)}{\Gamma(n)} \tag{9}$$

where $n$ is the number of samples of the received signal and $\lambda$ is the threshold of detection. $P_d(\gamma,\ \lambda)$ is the probability of detection for AWGN. The mathematical expression of $\overline{P_d}(\lambda)$ is evaluated by putting the PDF of the channel model in Eq. (8).

### 3.2.2 CSS Using K-Mean Clustering (Pd and Pfa)

As explained in Sect. 2, the output of the classifier will the either of the two classes as defined in Eq. (4). The performance parameters are defined by [23] as

$$\begin{aligned} P_{fa} &= P\left[A^* = -1 | A = +1\right] \\ P_d &= P\left[A^* = -1 | A = -1\right] \end{aligned} \tag{10}$$

where $A$ is the actual state and $A*$ is the predicted state of the channel.

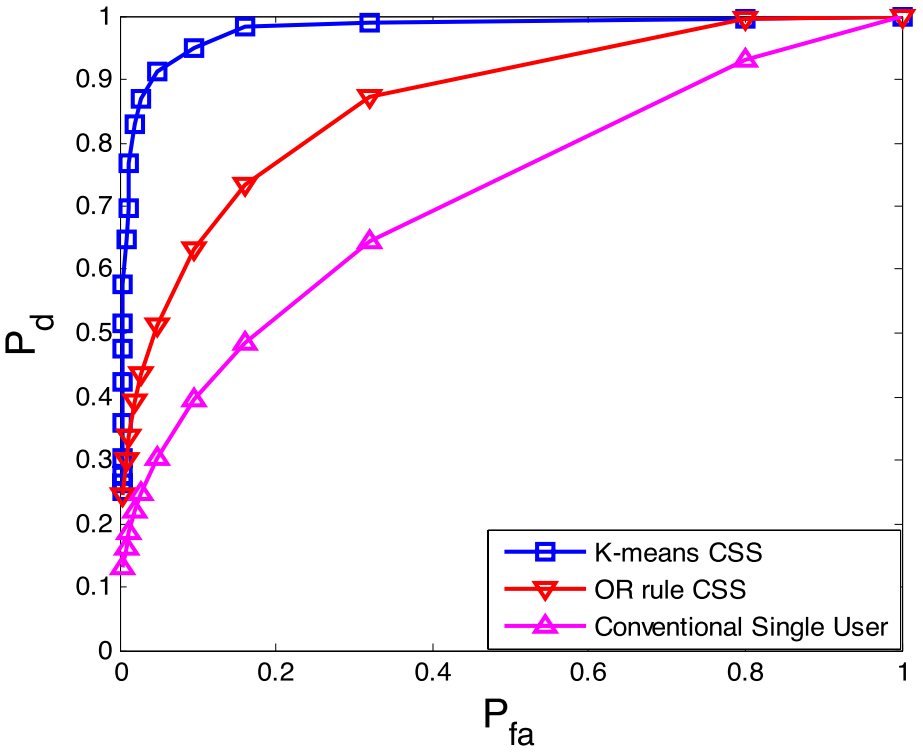



**Fig. 4** A performance comparison plots (Pd vs. Pfa) of K-mean with the conventional method over α–η–μ fading channel with [α, η, μ] = [1.5, 0.7, 1.5] and average SNR = -12 dB

## 4 Numerical Results and Discussion

To verify the accuracy of the proposed methodology derived in the preceding sections, we have plotted the simulation results in this section. The performance evaluation of K-mean clustering-based CSS in different channel conditions has been carried out. The performance of the proposed algorithm is compared with the conventional method of CSS with the energy feature vector. The α–η–μ and α–κ–μ fading models has now been widely accepted, as it leads to many known wireless channel models by the appropriate choice of α, η, κ, and μ parameters and elegantly represent useful channel models based on channel measurement data [24]. The numerical values of the fading parameters used in the simulations are taken from the field measurement data as presented in [24, 25]. The effect of the variation of different fading parameters, the average SNR, number of SUs, and the number of samples used to construct each energy vector on the detection performance of the fading models is demonstrated. The results for the K-mean clustering-based CSS have been generated through simulation. The transmitted signal from the PU is assumed to be BPSK (low-pass equivalent) modulated and the probability that a PU is in the active state is 0.5. A total of 500 energy vectors have been used for the purpose of the training and the testing has been performed on another 500 energy vectors.

The proposed algorithms are implemented by using MATLAB (R2014) in a 64-bit computer with a core i10 processor (clock speed of 2.8 GHz) and 8 GB RAM. The detection

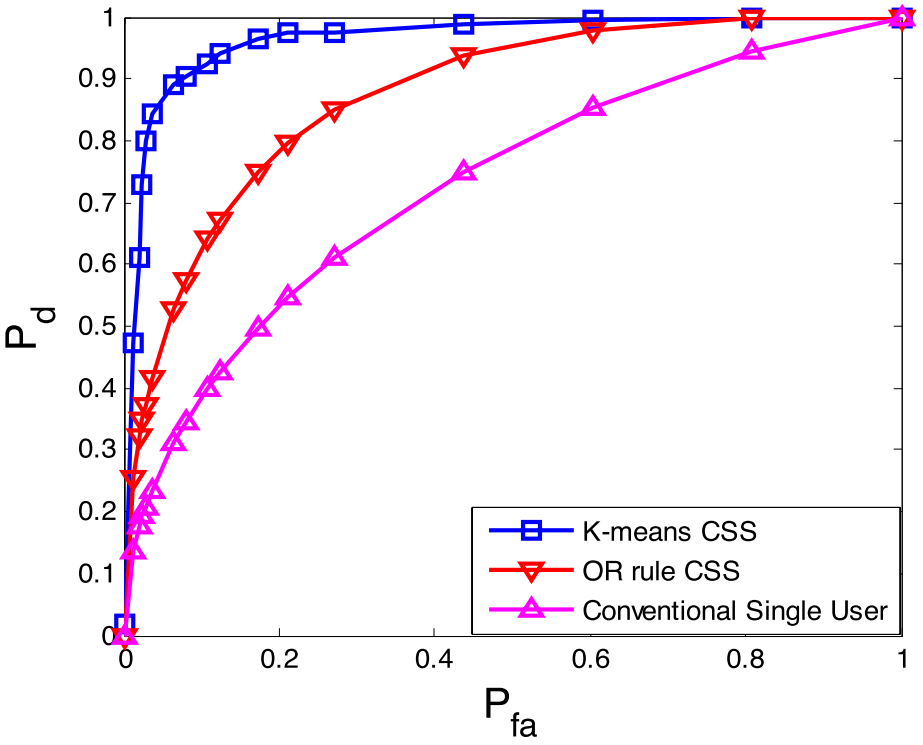



**Fig. 5** A performance comparison plots (Pd vs. Pfa) of K-mean with the conventional method over α–κ–μ fading channel with [α, κ, μ] = [2.5, 1.5, 1.5] and average SNR = -12 dB

performance of the proposed clustering scheme for 2 SUs with the conventional single-user and the OR rule CSS with 2 SUs over α–η–μ and α–κ–μ fading channels has been depicted in Figs. 4 and 5, respectively. The fixed parameters considered in Fig. 4 are 5 are [α, η, μ] = [1.5, 0.7, 1.5] and [α, κ, μ] = [2.5, 1.5, 1.5] with average SNR = -12 dB. The significant performance improvement is seen in the proposed scheme for both the generalized fading channels. More specifically, for the considered scenarios in Figs. 4 and 5, for a given $P_{fa}$ = 0.2, the $P_d$ of 2 cooperative users *K*-mean clustering is 85% larger than the single user conventional SS and 27% larger than the ORing scheme (where the logical OR operation is applied on the individual SUs decision, to get the cooperative decision at the FC) of 2 cooperative users for α– η–μ fading channel while the corresponding performance improvement in $P_d$ is 76% and 20% for the α–κ–μ fading channel.

The impact of the fading parameters on the ROC has been plotted in Figs. 6 and 7 over α–η–μ and α–κ–μ fading channels. It is noted from the plots that the system performance improves with an increase in α, the nonlinearity parameter. Furthermore, there is a notable improvement in system performance as μ and/or κ are incremented. This makes sense since μ represents the multipath clustering and an increase in κ implies higher dominance of the LOS component. Similarly, the system performance improves with the increase in η value. The comparison of ROC for different numbers of SUs over

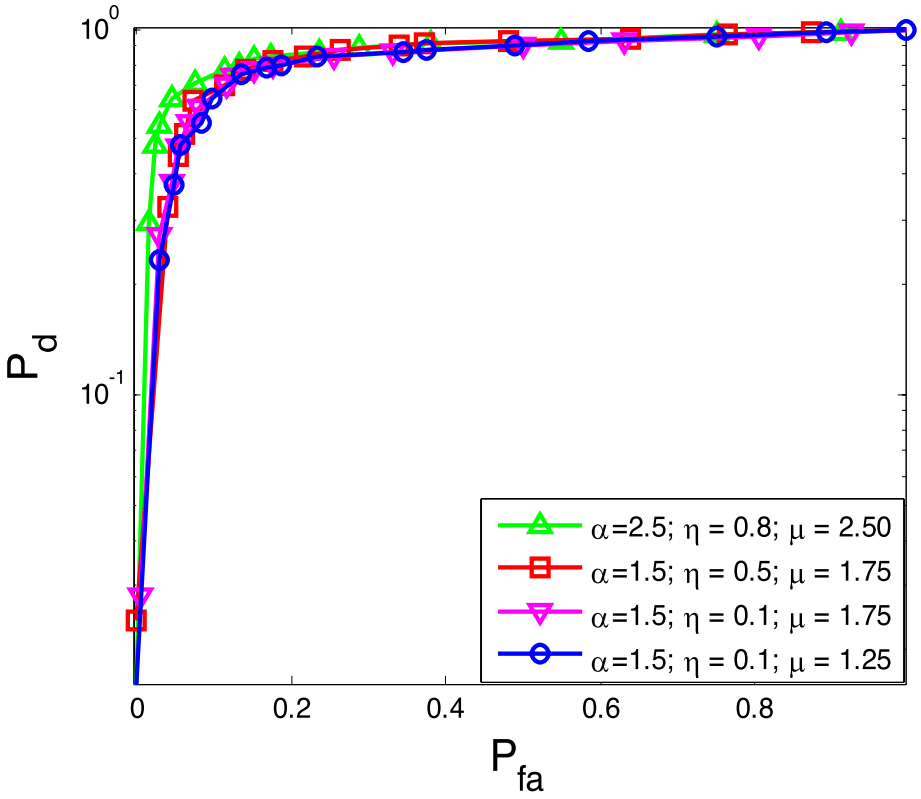



**Fig. 6** The impact of fading parameters on the ROC plots over α–η–μ fading channel for a fixed value of average SNR = -12 dB and *N* = 2

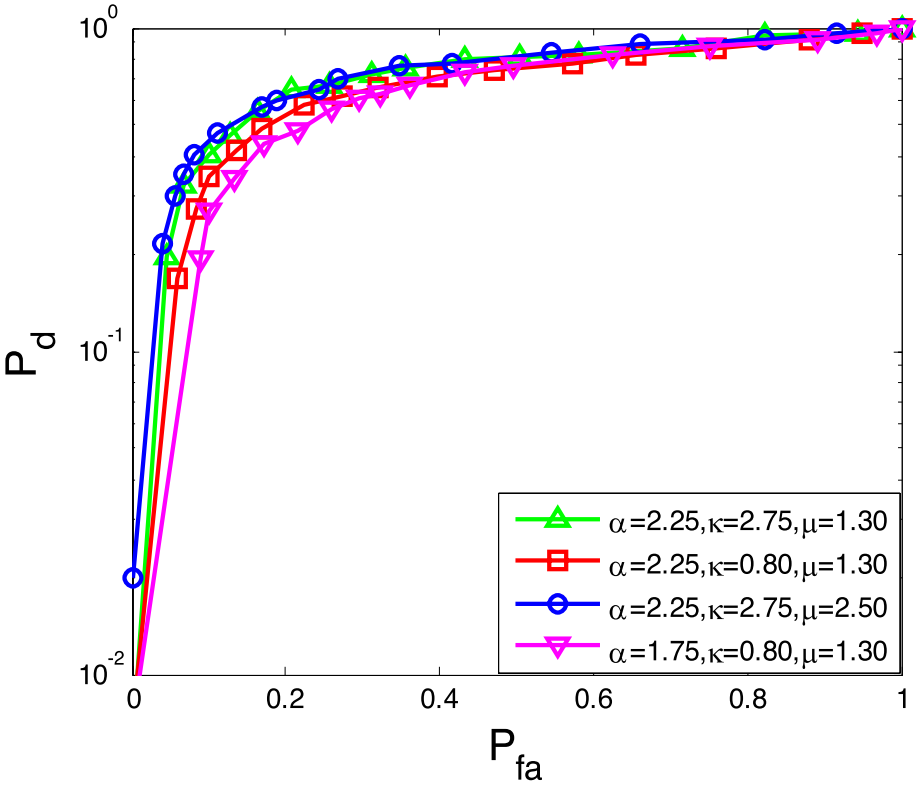



**Fig. 7** The impact of fading parameters on the ROC plots over α–κ–μ fading channel for a fixed value of average SNR = -12 dB and $N = 2$

α–κ–μ fading channel is shown in Fig. 8. It is very clear from the Figure that a significant performance gain can be achieved by choosing more number of SUs in the collaborative sensing. For a target probability of false alarm, $P_{fa} = 0.1$, the $P_d$ for $N = 3$ is 10% larger and for $N = 4$ is 29% larger than the corresponding value for $N = 2$.

The classification and clustering of energy vectors for 3 numbers of SUs are presented in Fig. 9. The two clusters (one each for channel available class and channel unavailable class) with their corresponding centroid $C_1$ and $C_2$ are shown in the Figure. The variation of ROC plots for different values of average SNR for α–η–μ fading channel have been presented in Fig. 10. The result confirms that the detection performance of the system increases with an increase in average SNR. Particularly, for $P_{fa} = 0.2$, the probability of detection, for $\overline{\gamma} = -$ 12 dB is 110% larger and for $\overline{\gamma} = -$ 14 dB is 77% larger than the corresponding value for $\overline{\gamma} = -$ 18 dB. The effect of different numbers of samples on the performance of clustering-based CSS has been demonstrated in Fig. 11.

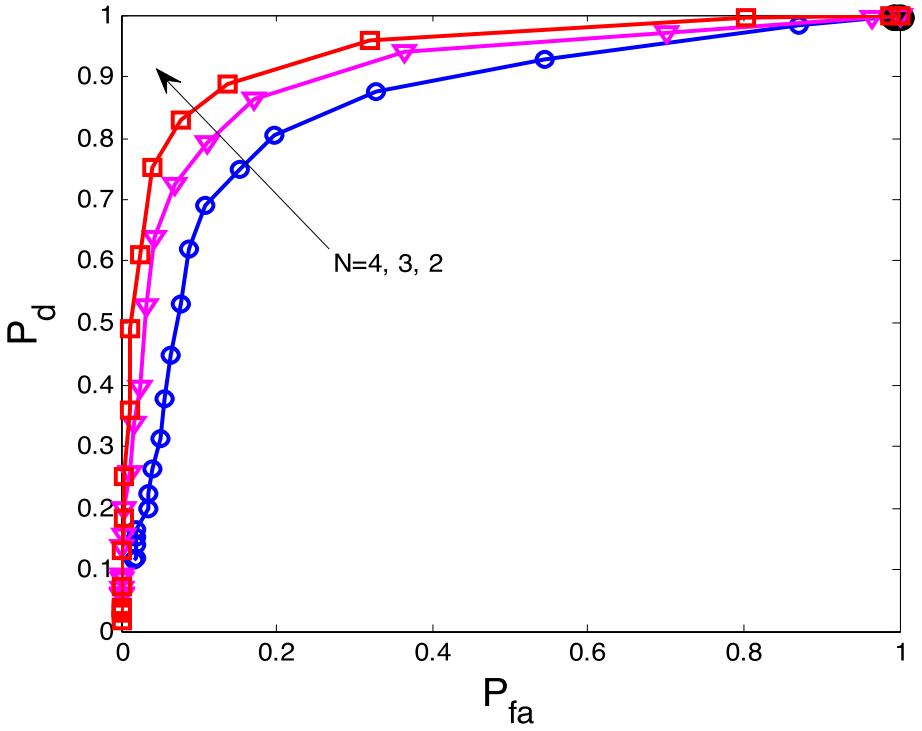



**Fig. 8** Pd as a function of Pfa for different number of SUs over α–κ–μ fading channel for a fixed value of average SNR = -14 dB with [α, κ, μ] = [2.0, 0.1, 3.0]

The result is in agreement with the theoretical concept that the detection performance of the system improves with the increase in the number of samples used to construct each feature vector. More specifically, for $P_{fa} = 0.1$, the probability of detection for $M = 200$ samples is 56% larger and for $M = 500$ samples is 150% larger than for $M = 100$ samples. It is also observed that the performance improvement due to the increase in the number of samples diminishes with an increase in the probability of false alarm.

## 5 Conclusion

Taking into account the advantage of machine learning techniques to solve the problem of SS, we have evaluated the performance of CSS over the generalized α-η-μ and α–κ–μ fading models using K-mean clustering algorithm. Clustering is a vector quantization

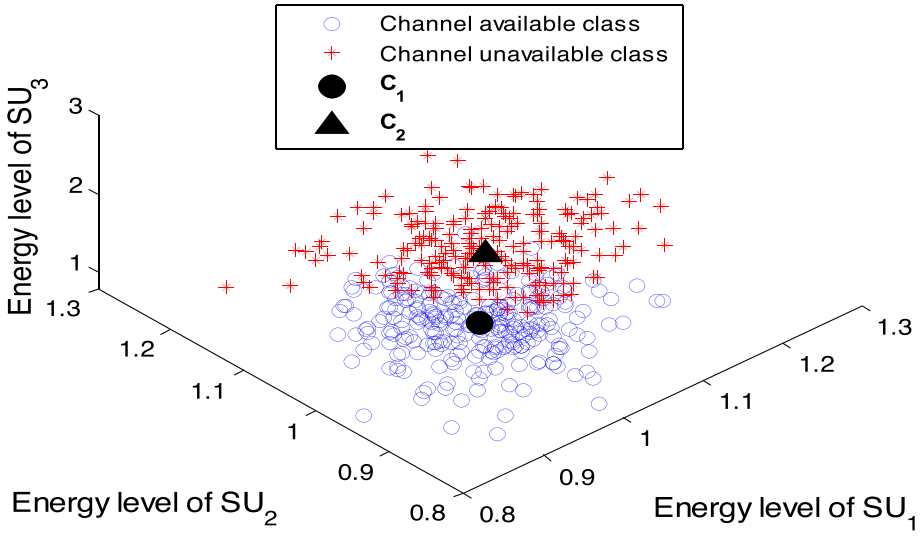



**Fig. 9** Classification and clustering of energy vectors for 3 number of SUs over α–κ–μ fading channel for a fixed value of average SNR = -14 dB with [α, κ, μ] = [2.0, 0.1, 3.0]

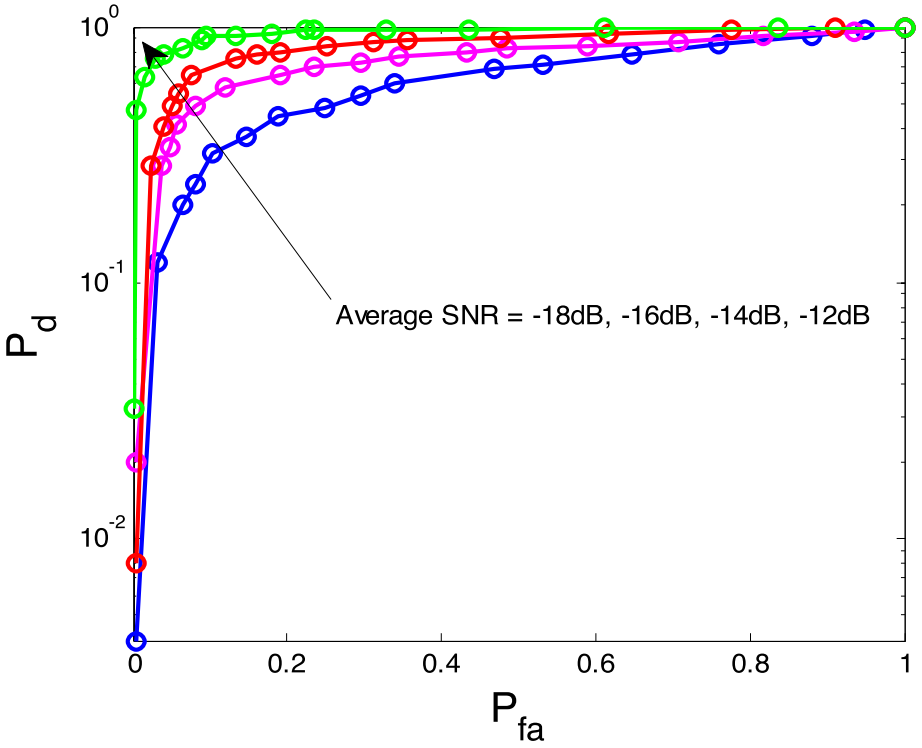



**Fig. 10** The effect of the variation of average SNR on ROC for α–η–μ fading channel with [α, η, μ] = [3.0, 0.7, 3.0] and $N = 2$

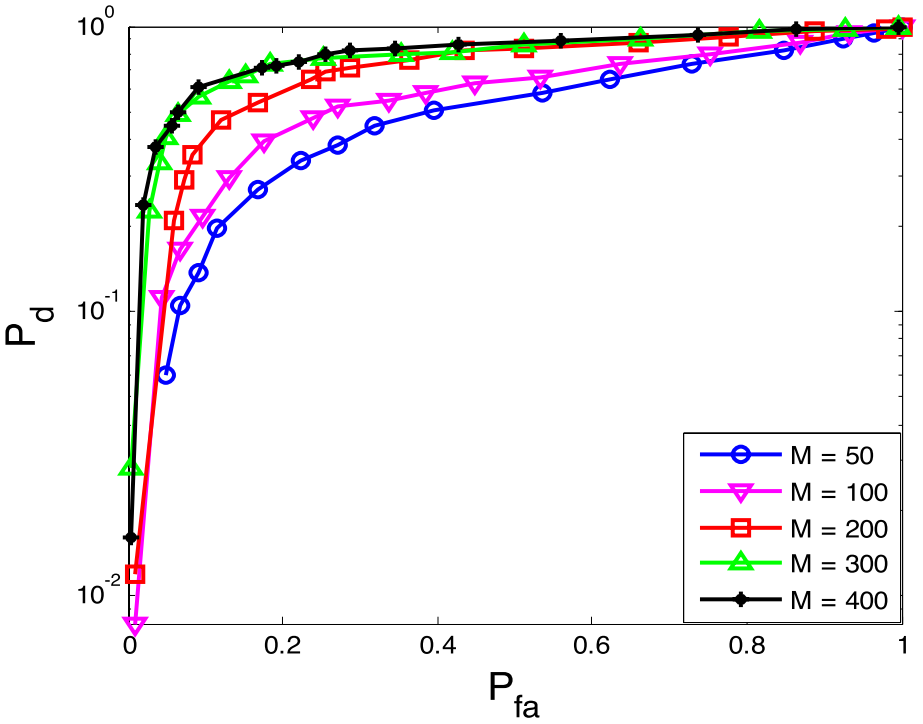



**Fig. 11** Effect of the number of samples on ROC over α–η–μ fading channel for a fixed value of average SNR = -14 dB with [α, η, μ] = [3.0, 0.7, 3.0]

technique that classifies the observed signal into some fixed number of clusters depending upon the nearest mean. *K*-mean clustering algorithm is used to classify the SUs received signal into the signal available class or signal unavailable class. The effect of the variation of different fading parameters, the average SNR, number of SUs, and the number of samples used to construct each energy vector on the detection performance for both the fading models is demonstrated. It has been concluded that the performance of the proposed algorithm is superior to the traditional methods of CSS with an energy feature vector.

**Acknowledgements** The authors would like to thank Lt. Achint Gupta and the anonymous reviewers for their constructive comments that have improved this manuscript.

**Authors' Contributions** All the authors have equally contributed in this manuscript.

**Funding** No funding was received for this work.

**Data Availability** N/A.

**Code Availability** N/A.

## Declarations

**Conflict of interest** Authors declare that there is no conflict of interest.

## References


1. Rasethuntsa, T. R., & Kumar, S. (2019). An integrated performance evaluation of ED-based spectrum sensing over $\alpha - \kappa - \mu$ and $\alpha - \kappa - \mu$-extreme fading channels. *Transactions on Emerging Telecommunications Technologies, 30*(5), e3569.
2. Kumar, S. (2018). Performance of ED based spectrum sensing over $\alpha - \eta - \mu$ fading channel. *Wireless Personal Communications, 100*, 1845–1857.
3. Kumar, S., Chauhan, P. S., Raghuwanshi, P., Kaur, M., & Singh, K. (2018). ED performance over α–η–μ/IG and α–κ–μ/IG generalized fading channels with diversity reception and cooperative sensing: A unified approach. *International Journal of Electronics and Communications, 97*(2018), 273–279.
4. Wang, Y., Zhang, Y., Wan, P., & Zha, S. (2018). A cooperative spectrum sensing method based on a feature and clustering algorithm. *Wireless Communications and Mobile Computing, 2018*, 1029–1033.
5. S. Kumar, M. Kaur, N. K. Singh, K. Singh, & Chauhan, P. S. Energy detection based spectrum sensing for gamma shadowed α–η–μ and α–κ–μ fading channels. *International Journal of Electronics and Communications (AEU)*, pp. 26–31.
6. Janu, D., Singh, K., & Kumar, S. (2022). Machine learning for cooperative spectrum sensing and sharing: A survey. *Transactions on Emerging Telecommunications Technologies, 33*(1), e4352.
7. Agarwal, A., Jain, H., Gangopadhyay, R. & Debnat, S. (2017). Hardware implementation of k-means clustering based spectrum sensing using usrp in a cognitive radio system. In *International Conference on Advances in Computing, Communications and Informatics*, pp. 1772–1777.
8. Zhang, Y., Wang, Y., Wan, P., Zhang, S. & Li, N. (2018). A spectrum sensing method based on null space pursuit algorithm and FCM clustering algorithm. *Springer Nature Switzerland* pp. 231–242.
9. Zhang S. et al. (2019) A cooperative spectrum sensing method based on information geometry and fuzzy c-means clustering algorithm. *EURASIP Journal on Wireless Communications and Networking* (pp. 1–12).
10. Yongwei, Z., Pin, W., Shunchao, Z., Yonghua, W. & Nan, L. (2017) A spectrum sensing method based on signal feature and clustering algorithm in cognitive wireless multimedia sensor networks. *Advances in Multimedia* (pp. 1–10).
11. Zhang, S., Wang, Y., Yuan, H., Wan, P., & Zhang, Y. (2019). Multiple-antenna cooperative spectrum sensing based on the wavelet transform and gaussian mixture model. *Sensors, 6*, 1–18.
12. Wang, Y., et al. (2019). *A cooperative spectrum sensing method based on empirical mode decomposition and information geometry in complex electromagnetic environment* (pp. 1–14). Wiley.
13. Ghazizadeh, E., Abbasi-moghadam, D., & Nezamabadi-pour, H. (2018). *An enhanced two-phase SVM algorithm for cooperative spectrum sensing in cognitive radio networks* (pp. 1–13). Wiley.
14. Wang, Y., Zhang, Y., Wan, P., Zhang, S. & Yang, J. (2018). A spectrum sensing method based on empirical mode decomposition and K-means clustering algorithm. *Wireless Communications and Mobile Computing* (pp. 1–11).
15. Marquez, H., Salgado, C. & Hernández, C. (2017). Multichannel assignment using K-Means in cognitive radio networks. *Tecnura* (pp. 68–78).
16. Salahat, E. & Hakam, A. (2014). Performance analysis of α-η-μ and α-κ-μ generalized mobile fading channels. In *20th European Wireless Conference* (pp. 1–6).
17. Chauhan, P. S., et al. (2021). Performance analysis of ED over air-to-ground and ground-to-ground fading channels: A unified and exact solution. *AEU - International Journal of Electronics and Communication, 138*, 153839.
18. Chauhan, P. S., Tiwari, D., Soni, S. K., & Kumar, S. (2019). Energy detector performance over lognormal fading channel with diversity reception. *Journal of Electromagnetic Waves and Applications, 33*(17), 2242–2256.
19. Verma, P. K., Soni, S. K., & Jain, P. (2018). Performance evolution of ED-based spectrum sensing in CR over Nakagami-m/shadowed fading channel with MRC reception. *AEU - International Journal of Electronics and Communications, 83*, 512–518.
20. Digham, F. F., Alouini, M. S., & Simon, M. K. (2007). On the energy detection of unknown signals over fading channels. *IEEE Transactions on Communications, 55*(1), 21–24.
21. Souza, R. A. A., Ribeiro, A. M. O., & Guimarães, D. A. (2015). On the efficient generation of $\alpha - \kappa - \mu$ and $\alpha - \eta - \mu$ white samples with applications. *International Journal of Antennas and Propagation, 2015*, 1–13.

22. Kumar, S., Soni, S., & Jain, P. (2018). Performance of MRC receiver over Hoyt-lognormal composite fading channel. *International Journal of Electronics, 105*(9), 1433–1450.
23. Thilina, K. M., Choi, K. W., Saquib, N., & Hossain, E. (2013). Machine learning techniques for cooperative spectrum sensing in cognitive radio networks. *IEEE Journal on Selected Areas in Communications, 31*(11), 2209–2221.
24. Fraidenraich, G. & Yacoub, M. D. (2006). The α–η–μ and α–κ–μ fading distributions. In *IEEE ninth international symposium on spread spectrum techniques and applications* (pp. 16–20).
25. Papazafeiropoulos, A. K., & Kotsopoulos, S. A. (2011). The α−λ−μ and α–η–μ small-scale general fading distributions: A unified approach. *Wireless Personal Communications, 57*, 735–751.

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

**Sandeep Kumar** received his B. Tech. in electronics and communication from Kurukshetra University, India in 2004 and Master of Engineering in Electronics and Communication from Thapar University, Patiala, India in 2007. He received his Ph.D. from Delhi Technological University, Delhi, India in 2018. He is currently working as Member (Senior Research Staff) at Central Research Laboratory, Bharat Electronics Limited Ghaziabad, India. His research interests include the study of wireless channels, cognitive radio networks and 5G and beyond technologies etc. He is also serving as a reviewer for IEEE, Elsevier and Springer journals.

**Puspraj Singh Chauhan** was born in UP, India. He received his Bachelor of Technology in Electronics and Communication Engineering from Uttar Pradesh Technical University, India in 2009, the M.Tech degree in Digital Signal Processing and Ph.D. in Wireless Communication from Govind Ballabh Pant Engineering College, Pauri, Uttarakhand, India in 2011 and 2021, respectively. Currently he is working as assistant professor in the department of Electronics and Communication Engineering at Pranveer Singh Institute of Technology, Kanpur, Uttar-Pradesh, India. His research interests include Wireless Communication and Propagation Channel Modeling.

**Rahul Bansal** received Ph.D. degree from Delhi Technological University (Formerly Delhi College of Engineering), New Delhi, India, in 2020. He was the recipient of Junior Research Fellowship and Senior Research Fellowship from Council for Scientific and Industrial Research (CSIR), India. He received Research Excellence Award from Delhi Technological University, Delhi, India in March 2020. He is a reviewer of several international journals of Springer, IEEE, Elsevier, Taylor & Francis, Emerald. He is currently working as an Assistant Professor with the Department of Electronics and Communication Engineering, Ajay Kumar Garg Engineering College, Ghaziabad, India. His research interests include nonlinear filtering, state and parameter estimation of the systems, stochastic filtering.

**Manpreet Kaur** was born in Punjab, India in 1983. She received her B. Tech. in Electronics and Communication Engineering from Punjab Technical University, Punjab, India in 2005 and Master of Engineering in Electronics and Communication from Thapar University, Patiala, India in 2007. Presently, she is pursuing Ph.D. from Delhi Technological University, Delhi. She is currently working as Member (Senior Research Staff) at Central Research Laboratory, Ghaziabad, Bharat Electronics Limited. Her research interest includes the wireless communication and cognitive radios networks.

**Rajesh Kumar Yadav** is presently working as an Assistant Professor in the Department of Computer Science and Engineering, Delhi Technological University, New Delhi, India. His areas of research interest are Mobile ad-hoc Networks and Wireless Sensor Networks.

*2022 3rd International Conference for Emerging Technology (INCET)*
*Belgaum, India. May 27-29, 2022*

# Power Flow Management and DC bus Stabilization for SPV based Battery Energy Storage System

Anupam
*Electrical Engineering Department*
*Delhi Technological University*
Delhi, India
anupamsharma478@gmail.com

Mayank Kumar, *SM IEEE*
*Electrical Engineering Department*
*Delhi Technological University*
Delhi, India
mayankkumar@dtu.acin

Suman Bhowmick
*Electrical Engineering Department*
*Delhi Technological University*
Delhi, India
suman.bhowmick@dce.ac.in

2022 3rd International Conference for Emerging Technology (INCET) | 978-1-6654-9499-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/INCET54531.2022.9824508

***Abstract* — Solar photovoltaic (SPV) energy has grown in popularity and attention as a result of its ability to generate power by converting solar irradiance into electricity using SPV cells. The SPV energy generation has gained a lot of commercial and research attention these days. The major goal of this work is the management of the SPV generated power and maintaining the DC bus voltages under varying irradiance and load conditions. The maximum power point tracking (MPPT) algorithm is used for optimal power generation under variable irradiance and load conditions. In this work, the bidirectional converter is used to control the voltage of the DC buses and to balance the power generated by the SPV array and the battery energy storage system (BESS) during day / night time. The complete solar PV system with BESS and variable load is simulated in the MATLAB Simulink platform. The results validate the proposed technique.**

***Keywords- Battery energy storage system (BESS), boost converter, maximum power point tracking (MPPT), perturb and observe (P&O), solar photovoltaic (SPV) array.***

## I. INTRODUCTION

The concern for building an emission-free environment has grown due to the depletion of fossil fuels and conventional source reserves. In this respect, renewable sources of energy are being increasingly considered in both grid-connected and standalone modes of operation. The energy storage elements are interconnected via dc-dc converters for high reliability of power supply to the loads. To improve the efficiency and stability of renewable energy system, power flow management control is the primary condition [1]. Furthermore, the enhancement of power electronic interfaces in a DC microgrid facilitates the control of SPV system and charging/discharging process of BESS. It also balances the control of power flow in the system under varying irradiance and load conditions.

The DC bus voltage control is required major attention to improve SPV system stability [2], [3]. The use of energy storage devices along with a bidirectional converter connected to the DC bus is required to achieve power management. Power management and coordination control ensure that the reliable operation of DC microgrid operates under different operating conditions. It allows more flexibility in control scheme implementation and hence ensure that the DC voltages remain below acceptable limits along with maintenance of system power balance in both islanded as well as grid connected modes [4], [5]. The buck-boost converter is used as bidirectional DC-DC converter in this work. Furthermore, the load considered in this model are dc household loads.

The state of charge (SOC) of an energy storage device is used to obtain safe charge/discharge limits [6]. The SOC slope measures the battery's charge/discharge rate. During the daytime, the SPV provides the electricity to loads as well as battery charging [7], [8]. The bidirectional DC-DC converter charges the battery in buck mode during day-time, and at night, it enables to discharge the battery to the DC load in boost mode [9], [10]. In the proposed model, controllers are designed to achieve primary objective of maintaining constant DC bus voltage at 48V. The power flow management is also presented under variable irradiance condition [11], [12].

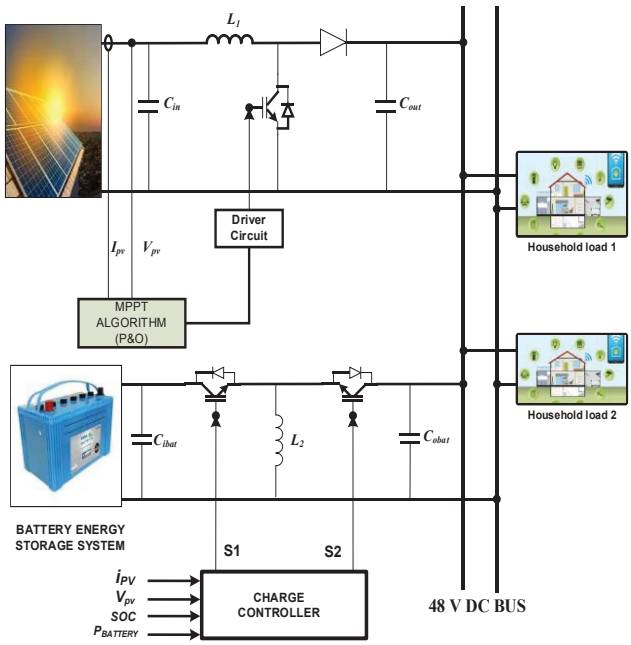



Fig.1 SPV system integrated with BESS and household loads.

## II. SYSTEM MODEL DISCRIPTION

The solar cell's output produces fluctuating DC voltage. The atmospheric conditions are also responsible for the changing DC voltage. A Soltech 1STH-215-P solar panel is used in the simulation. The SPV panel parameters are shown in Table 1. The above SPV system is simulated at various irradiances conditions such as 1000 W/m$^2$, 600 W/m$^2$, 400 W/m$^2$, 800 W/m$^2$ and at 50°C. The SPV system is integrated with DC-DC boost converter. The boost converter circuit consists of IGBT switches with anti-parallel diode, inductor, power diode, driver circuit and other components. The MPPT controller receives the signal from the PV panels, which includes SPV voltage and current. The boost converter gate



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:24:53 UTC from IEEE Xplore. Restrictions apply.

pulses are controlled using perturb and observe (P&O) algorithm for maximum power harvesting from SPV panel. A buck-boost bidirectional converter is used for power flow from dc microgrid to BESS and BESS to dc microgrid. During the operation of buck mode, the batteries are charging, while in the boost operation mode, the batteries are discharging. The complete charging/discharging methodology is discussed in section IV.

## III. Modelling and Design of System Components

### A. Modeling of PV module

A photovoltaic module is composed of a group of photovoltaic cells connected in a defined manner (series or parallel). The circuit equivalent of an ideal PV cell, which has a current source connected in series with a diode is shown in Fig. 2. The non-ideal conditions are considered for the modelling of SPV cell. Therefore, shunt and series resistances are included in the model, where $R_{sh}$ is the high-value equivalent shunt resistance; and $R_s$ is the small-value equivalent series resistance. From Fig. 2 following mathematical relations can be developed.

$$I_{ph} = I_d + I_{R_{sh}} + I_{pv} \tag{1}$$

$$I_{pv} = I_{ph} - I_d - I_{R_{sh}} \tag{2}$$

$$I_{pv} = I_{ph} - I_o e[\frac{q(V_{pv} + I_{pv}R_s)}{nKT}] - \frac{V_{pv} + I_{pv}R_s}{R_{sh}} \tag{3}$$

where, $I_{pv}$ is the solar cell current; $I_o$ is the reverse saturation current; $I_{ph}$ is the cell photo current; $V_{pv}$ is the solar cell voltage; $K$ is the Boltzmann's constant; $T$ is the temperature in Kelvin, $q$ is the unit charge; $R_S$ is series resistance; $R_{sh}$ is parallel shunt resistance

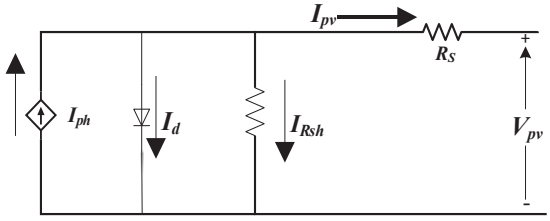



Fig.2 Equivalent circuit of single SPV cell.

The induced current of solar cell is determined by the irradiance and the temperature applied to the cell, and it is stated as follows:

$$I_{ph} = I_{sc} + K_i(T_c - T_r)I_r / 1000 \tag{4}$$

where, $I_{sc}$ = short-circuit current of cell; $K_i$ = temperature coefficient; $I_r$ = irradiance of cell; $T_c$ = cell temperature; $T_r$ = reference temperature.

TABLE I. Electrical Characteristics of PV Panel (Soltech 1STH-215-P) at STC (1000 W/M$^2$ and 25°C)

| Parameters | Value |
|---|---|
| $N_P$ (PV parallel strings) | 2 |
| $N_S$ (PV series strings) | 1 |
| $V_{OC}$ (Open circuit voltage) | 36.3V |
| $I_{SC}$ (Short circuit current) | 7.84A |
| $V_{MP}$ (Voltage at MPP) | 30V |
| $I_{MP}$ (Current at MPP) | 7.35A |
| $P_{MP}$ (Power at MPP)) | 213.15W |

### B. Designed Parameters for Bidirectional Converter

Fig. 3 shows the DC-DC bidirectional converter, the working of converter and its operating modes depend on state of charge (SOC) of the BESS. The bidirectional converter has two modes of operation, i.e. charging mode and discharging mode of operation. In case of discharging mode, switch $S_1$ is on, which discharges the battery and power flows from battery to dc bus. In case of charging mode, the switch $S_2$ is on, which charges the battery and power flows from dc bus to battery.

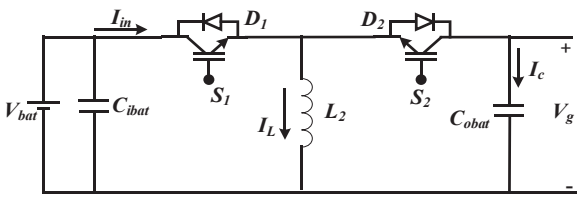



Fig.3 Circuit diagram of bidirectional converter

*Bidirectional Converter in Charging Mode:*

In this mode $S_2$ and $D_1$ will be in on-state whereas $S_1$ and $D_2$ will be in off-state. This mode of operation can further be classified into two intervals:

a. *Interval 1 ($S_2$-ON, $D_2$-OFF; $S_1$-OFF, $D_1$-OFF):* When switch $S_2$ is on, the inductor $L_2$ charges. The steady-state analysis of interval-1 can be written as follows:

$$V_g = V_L = L\frac{di}{dt} \tag{5}$$

$$\frac{di}{dt} = \frac{\Delta i_L}{\Delta t} = \frac{V_g}{L} \tag{6}$$

$$\Delta i_L\big|_{S_2=on} = \left(\frac{V_g}{L}\right)DT \tag{7}$$

b. *Interval 2 ($S_2$-OFF, $D_2$-OFF; $S_1$-OFF, $D_1$-ON):* In this interval, switches $S_1$ and $S_2$ both are in off state. The current flowing in the inductor cannot be changed instantly so polarity of voltage across $L_2$ get reversed and will be in series with input voltage, so diode $D_1$ will be in forward bias and inductor current will charge the capacitor $C_{ibat}$. The steady-state analysis of interval-2 can be written as follows:

$$V_{bat} = V_L \tag{8}$$

$$V_L = L\frac{di}{dt} = L\frac{\Delta i_L}{\Delta t} = V_{bat} \tag{9}$$

$$\Delta t = (1-D)T \tag{10}$$

$$\Delta i_L\big|_{D_1=on} = \frac{V_{bat}}{L}(1-D)T \tag{11}$$

$$\text{in steady state;} \quad \Delta i_L\big|_{S_2=on} + \Delta i_L\big|_{D_1=on} = 0 \tag{12}$$



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:24:53 UTC from IEEE Xplore. Restrictions apply.

$$\frac{V_{bat}}{L}(1-D)T+\frac{V_g}{L}(DT)=0 \quad (13)$$

$$\frac{V_{bat}}{V_g}=\frac{-D}{(1-D)} \quad (14)$$

$$\Delta I_L=\frac{V_g D}{f_S L} \quad (15)$$

$$\Delta V_{C_{ibat}}=\frac{I_{bat}D}{fC} \quad (16)$$

where, $I_L$ = inductor current; $V_c$ = capacitor voltage; $I_{bat}$ = average value of battery current; and $f_s$ = switching frequency.

*Bidirectional Converter in Discharging Mode:*

In this mode $S_1$ and $D_2$ will be in on-state whereas $S_2$ and $D_1$ will be in off-state. This mode of operation can also further be classified into two intervals. In interval 1, $S_1$-ON, $D_2$-OFF, $S_2$-OFF, $D_1$-OFF; whereas in interval 2, $S_1$-OFF, $D_2$-ON, $S_2$-OFF, $D_1$-OFF. The operation of power flow under discharging mode will be similar to charging mode.

### C. Control topology of Bidirectional Converter

In this section, control algorithm (closed loop PID control technique) of bidirectional converter is discussed and it is shown in Fig. 4. The output voltage ($V_g$) is sensed and compared with the mentioned reference voltage ($V_{ref}$) to obtain the error and further this error is applied to PID controller. The output of PID controller is compared with carrier frequency signal ($V_{carrier}$) for constant switching frequency operation and resulting gate pulses are generated. The voltage control provided here is to maintain the value of 48V at the output DC bus. So, the output voltage is modified from this control loop and returns to its normal position. The design parameter of bidirectional converter is mentioned in Table II.

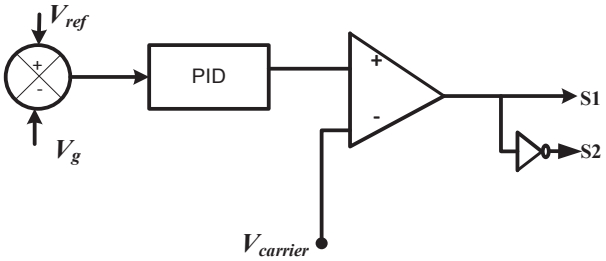



Fig. 4 Control algorithm of buck-boost converter

TABLE II. Parameters of Bidirectional Converters

| Sl. No. | Parameters | Value |
|---|---|---|
| 1 | DC link capacitor (source side) | 1000μF |
| 2 | DC link capacitor (Load side) | 220μF |
| 3 | Energy transfer inductance | 1mH |
| 4 | Switching frequency | 5kHz |

### D. Battery Energy Storage

The energy storage device considered for the simulation of proposed work is lead-acid battery. The discharging characteristics of lead-acid battery is shown in Fig. 5 (a). The battery is charged by battery charge controller with 3-stage charging strategy that includes MPPT bulk charge, constant voltage absorption charge and float charge stages as shown in Fig. 6. The charging /discharging of battery depend on the SOC and depth of discharge (DOD). If SOC is 100% then battery will be fully charged and DOD will be 0%, if SOC is 75% then battery charging status will be 75% and DOD will be 25%. The energy calculation of battery is done as per the energy required (i.e. working of various appliances, quantity and its operating hours) and autonomy (in days) of the system, which is shown in Table III.

$$\%\eta=\frac{\text{Discharged } Ah \text{ after fully charged}}{\text{Delivered } Ah \text{ to battery during charge}}\times 100 \quad (17)$$

Equation (17) shows the charging efficiency of lead acid battery which depends upon the SOC, temperatures and charging rates.

TABLE III. Considered DC Load for Household Applications

| Sl. No. | Appliances | Quantity | Power Required (W) | No. of Hours |
|---|---|---|---|---|
| 1. | Desktop (100W) | 1 | 100 | 6 |
| 2. | LED Bulb (20W) | 3 | 60 | 5 |
| 3. | Fan (60W) | 1 | 60 | 5 |
| 4. | LCD-TV (150W) | 1 | 150 | 4 |
| 5. | Refrigerator (80W) | 1 | 80 | 1 |
| **Total** | | | **450W** | |

Total energy consumption from the load appliances:

$$\begin{aligned} Wh_{load} &= (100\times1\times6)+(20\times3\times5) \\ &\quad +(60\times1\times5)+(150\times1\times4)+(80\times1\times1) \\ &=(600+300+300+600+80)=1880\text{Wh} \end{aligned}$$

The nominal voltage of dc bus voltage is used in the proposed manuscript is $V_{bus}=V_g=48V$

No. of autonomy days ($N$) = 1 day (considered)

Battery efficiency $\eta=85\%$

Depth of discharge of battery (DOD) = 50%

$$Battery\ capacity=\frac{Wh_{load}\times(N+1)}{\eta\times DOD\times V_g}=\frac{1880\times2}{0.85\times0.5\times48}=184.3Ah$$

For safe criteria battery capacity is to be taken as 200 Ah.

$$\begin{aligned} \text{Peak load demand} &= (100+60+60+150+80) \\ &= 450W \end{aligned}$$

$$\text{Maximum load current}=\frac{450}{48}=9.375A$$

Nominal discharge current is to be taken as 10A for transferring excess PV power to battery without disturbing bus voltage during high insolation period and with low power demand, nominal discharge current to be taken as 10A.

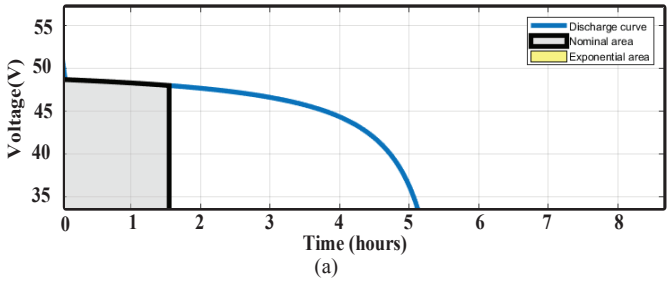


Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:24:53 UTC from IEEE Xplore. Restrictions apply.

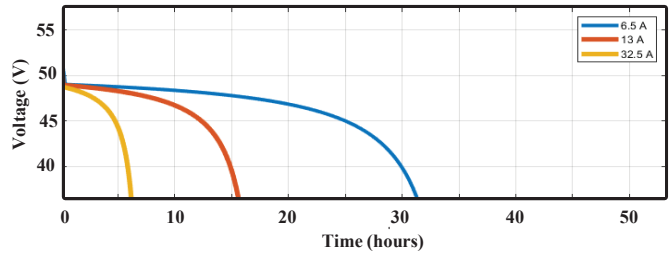
Now, C-rating of battery $\left(\frac{200\,Ah}{10}\right) = C/20$

Hence, battery can provide 10A of current continuously for 20 hours in absence of PV array and with 200 Ah capacity of 48V system can be used for 1 day of autonomy.

The discharging characteristics of lead-acid battery for 200 Ah capacity, 48V and 50% of SOC is shown in Fig.5 (a). It shows total charge extracted from lead acid battery before the discharging stops. It is observed that once the voltage of battery reaches to its full charge capacity, it is controlled by constant voltage charging method. In this case, battery voltage reaches to constant value at 48V, and then battery current starts decreasing gradually with time. Fig.5 (b) represents the time taken by the lead acid battery to discharge at different C-rates. The highest discharge current of 32.5A that takes 6 hrs to complete the discharging process, it will take 15.5hrs to discharge the 13A of current. The lowest discharging current is of 6.5A which takes 32hrs to complete the discharging process of the battery.



(a)



(b)

Fig.5 Lead acid battery characteristics for (a).discharging at single C-rate with constant voltage; (b) discharging at different C-rates.

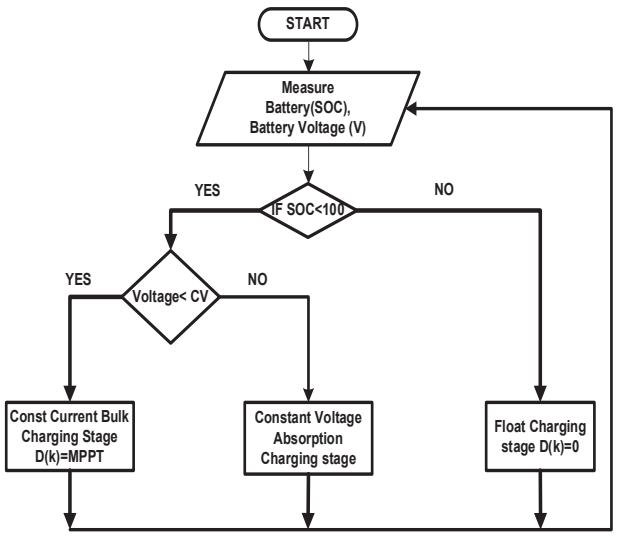



Fig.6. Flowchart of lead acid battery controller.

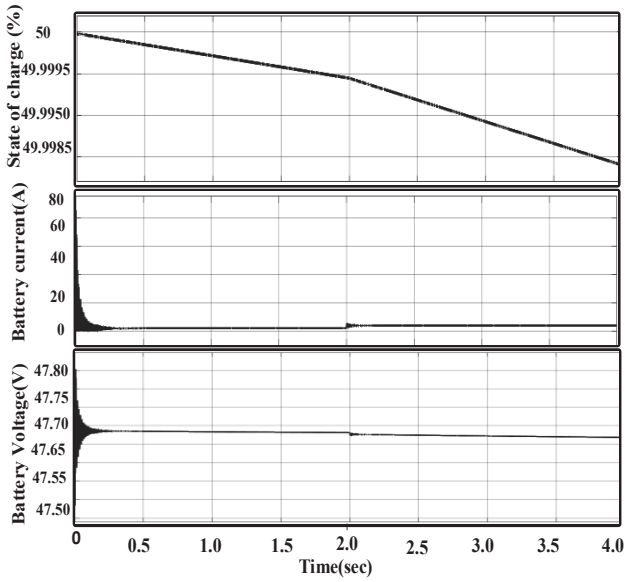



Fig.7. Charge controller performance of lead acid battery

Fig. 7 shows that the irrespective of constant current and voltage, the SOC decreases after reaching at constant voltage of 48V. The charge controller will no longer charge the battery after this stage and the duty cycle will vary between MPPT and zero for maintaining the voltage of 48V at DC bus.

## IV. Charging Algorithm

### A. Solar power charge controller

The flow chart indicating the performance of solar charge controller which is demonstrated in Fig. 8. This charge controller contains three modes of charging techniques. These three modes include constant current (CC), constant voltage (CV) and float charging. The first mode is constant current charging, which is also called as bulk charging mode, in which the charging current is charged on its rated capacity. The second mode is constant voltage (CV) charging which is known as absorption charging mode, when the battery is to be charge at constant voltage (CV), at this mode the MPPT get disabled. The third mode is float charging mode which maintains SOC of battery at its full capacity (i.e. 100%). It also prevents the battery from overheating due to excessive charge at uncontrolled level, otherwise it enters into float charging stage, at which duty cycle is zero. In the second case, Voltage level of battery determines the MPPT bulk or CV absorption charging level. If the voltage of battery comes below the (CV) point, charger will switch to MPPT at constant current (i.e. bulk charging stage), otherwise it will disable MPPT and switch at constant voltage absorption charging mode by rapid switching between MPPT and zero. This phenomenon occurs to maintain the battery voltage at the constant voltage set-point till the float charging stage is reached.

### B. Maximum power point tracking

This paper uses P&O technique for MPPT which is commonly used for tracking solar output signals in effective manner. This algorithm makes the solar output signals to reach at maximum level as per the duty cycle ΔD. In the above technique, introduction of minor perturbation will cause the variation in the PV power module. If voltage



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:24:53 UTC from IEEE Xplore. Restrictions apply.

increases, the power increases, then ΔD increases. On other hand voltages decreases, power decreases hence ΔD decreases, steps of the above will continue till maximum power point reaches. The above corresponding voltage at which maximum power point reaches is called reference point ($V_{ref}$).This technique is used to improve the efficiency of the solar panel MPPT. This technique of MPPT is also called as impedance matching technique.

### C. Constant Voltage charging

The lead acid battery uses CV charging method. Using this method, the battery capacity is increased by some percentage (i.e. 20%), in this process voltage remains constant while charging the battery. When battery works in discharge condition, current gradually decreases, which picks up the charge and increases the back emf. Some of the advantages comes under constant voltage charging are that it allows different capacities of cell at the different degree of discharge to be charged. In the beginning of charging, the large charging current is of small duration which will not deteriorate the cell.

It limits the initial charge current which sets the charge voltage according to specified charge and temperature. But in the end of the voltage charging, above charging current decreases to zero, because the battery voltage becomes equal to supply voltage. CV method of charging allows the battery to absorb large amount of charge in short interval of time without damaging the battery.

## V. Power Management of DC Microgrid

The aforesaid proposed techniques and charging algorithms depict the power flow balance between PV power generation, DC bus, energy storage device, and load power demand which is explained in Fig.8. Actual conceptualization of DC Bus stabilization and microgrid operation has originated to increase the efficiency and sustainability of grid. The power management of this microgrid with energy storage device operates in three modes.

***(i) Battery charging mode*:** The charging mode of Battery depends upon SOC condition.

$$SOC = \frac{Usable\ capacity(Ah)}{Maximum\ capacity(Ah)} \times 100 \qquad (18)$$

If the SOC of battery is smaller than its upper limit (SOC < $SOC_H$ ) and SPV array voltage is larger than load power demand, We can operate solar charge controller in its maximum power mode, the dc source which was isolated from the system will remain in idle state.

***(ii) Battery Discharging Mode:*** In this mode, if SOC of battery gets higher than its minimum limit (SOC > $SOC_L$) and if SPV output power is lesser than load power demand, then the battery starts discharging. The charge controller will start operating into the MPPT mode, and the deficit power which is being provided by the battery and the dc source will switch into idle state.

***(iii) Ideal Battery Mode:*** Battery mode in ideal conditions is classified into two different modes, namely,

***Sub-Mode 1:*** Whenever SOC of BESS get more than its maximum limit (SOC > $SOC_H$), and the SPV power generated exceeds load demand, then charge controller of SPV array can function according to load power demand i.e. MPPT mode gets disabled and the dc power source will remain in idle state.

***Sub-Mode 2*:** When the energy storage device's SOC is lesser than its minimum limit (SOC < $SOC_L$) and the generated power from SPV array is lesser than the load power demand, then additional load power is compensated by isolated dc source and meanwhile the charge controller will operates in MPPT mode, which is also called as isolated dc source mode. The above explained operating mode of battery can be used to regulate electricity for charging/discharging of BESS which has used in bidirectional converter for DC microgrid application.

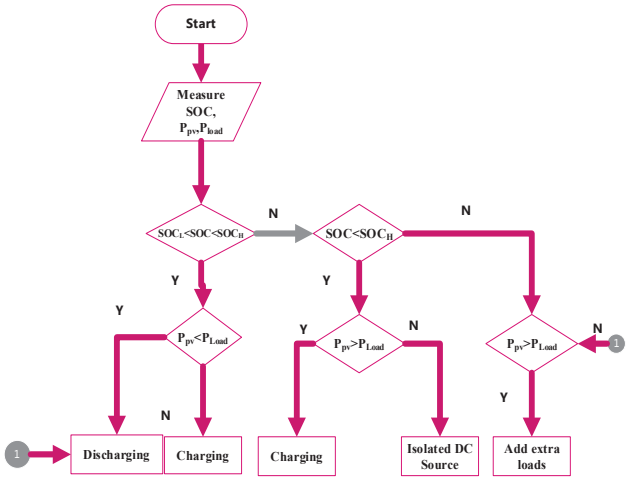



Fig. 8. Flowchart depicting power flow control algorithm.

## VI. Simulation Results

The BESS integrated SPV array with DC microgrid is implemented using the MATLAB Simulink platform. The parameters of the system are given in Tables I-III. The solar PV irradiance level is set to 400, 600, 800 and 1000 W/m$^2$ to represent changing solar irradiance. DC loads have been considered at three different voltage levels – 48 V, 24 V (representing low voltage residential DC appliances) and 310 V (representing high voltage residential DC appliances). The 310 V DC bus is obtained from the 48 V DC bus voltage using a step-up converter. Finally, the 24 V DC bus is obtained from the 48V DC bus using a step-down converter. Moreover, at each voltage level, two different load resistance values have been considered, which are changed (at time t = 1sec.) to represent a dynamically changing load. For the 48V DC bus, initially, the load is represented by a resistance of 25 Ω, which is changed to 40 Ω at time t=1sec. Similarly, for the 310 V DC bus, initially the load is represented by a resistance of 58 Ω, which is changed to 29 Ω at time t=1sec. Finally, for the 24 V DC bus, initially, the load is represented by a resistance of 20 Ω which is changed to 40 Ω at time t=1 sec.

Figs. 9-11 shows the variation of the 48 V, 310 V and 24 V DC bus voltage with time vis-à-vis simultaneous variation of solar irradiance and load. From Figs. 9-11, it is observed that the DC bus voltages at all levels (48 V, 310 V and 24 V) are maintained at constant values by the proposed technique, independent of the amount of solar irradiance and the load variation. This validates the effectiveness of the proposed technique.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:24:53 UTC from IEEE Xplore. Restrictions apply.

## VII. Conclusions

This paper proposes the concept of integration of solar PV array, DC microgrid and energy storage devices for various load applications. Power exchange with the energy storage devices is done using a bidirectional converter which is accustomed to supply power to DC bus from a solar PV array. For MPPT, P & O technique has been used to extract maximum power from solar panels. The DC bus voltages at all three levels are maintained constant using battery charging/discharging techniques vis-à-vis varying solar irradiance levels and varying load. This validates the proposed technique.

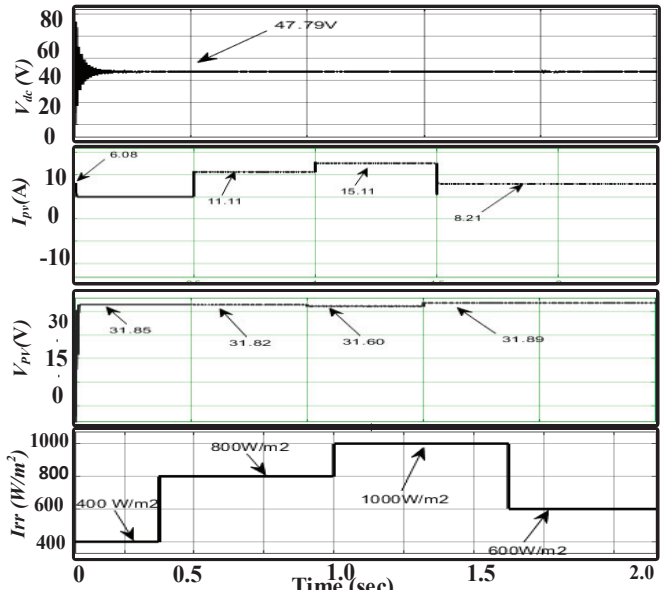



Fig.9. Variation of 48 V DC Bus voltage vs. time for varying solar irradiance

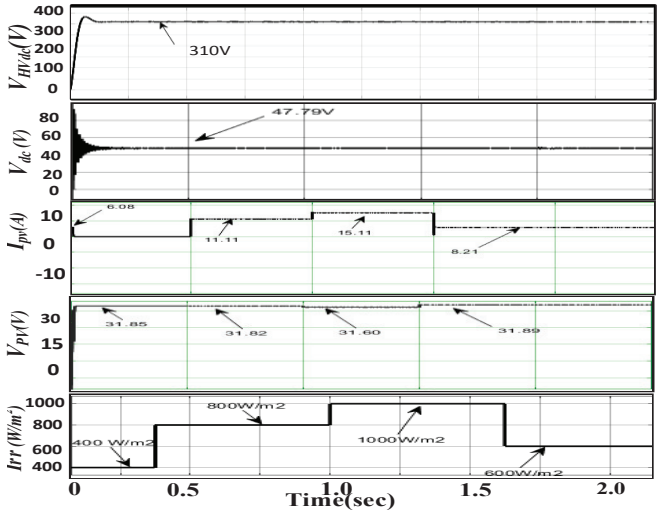



Fig.10. Variation of 310 V DC Bus voltage vs. time for varying solar irradiance

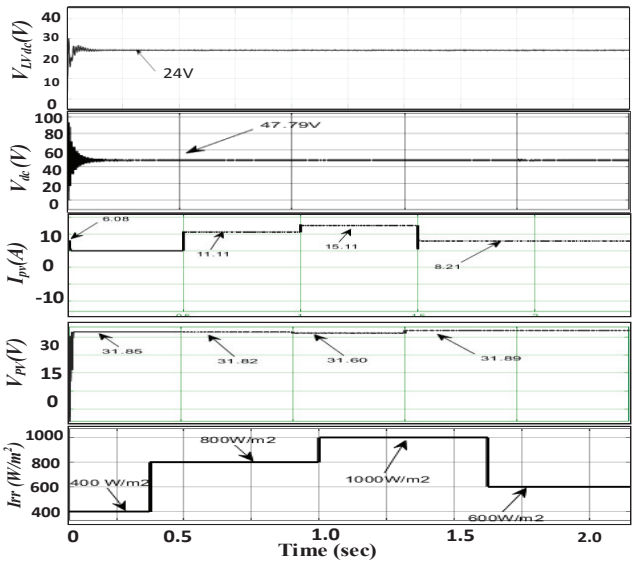



Fig.11. Variation of 24 V DC Bus voltage vs. time for varying solar irradiance

## References


[1] H. K. Mahore and V. Dake, "Modelling and analysis of grid-integrated solar system with power management scheme using battery," *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, 2019, pp. 1620-1623.

[2] A. Khamis, A. Mohamed, H. Shareef, A. Ayob and M. S. M. Aras, "Modelling and Simulation of a Single Phase Grid Connected Using Photovoltaic and Battery Based Power Generation," *2013 European Modelling Symposium*, 2013, pp. 391-395.

[3] M. Kumar, "Solar PV Based DC Microgrid under Partial Shading Condition with Battery-Part 2: Energy Management System," *2018 8th IEEE India International Conference on Power Electronics (IICPE)*, 2018, pp. 1-6.

[4] D. Singh, A. Agrawal and R. Gupta, "Power Management In Solar PV Fed Microgrid System With Battery Support," *2017 14th IEEE India Council International Conference (INDICON)*, 2017, pp. 1-6

[5] A. S. Anees, "Grid integration of renewable energy sources: Challenges, issues and possible solutions," *2012 IEEE 5th India International Conference on Power Electronics (IICPE)*, 2012, pp. 1-6.

[6] S. Park *et al*., "Hardware implementation of optimization technique based sensorless MPPT method for grid-connected PV generation system," *2009 International Conference on Electrical Machines and Systems*, 2009, pp. 1-6 .

[7] P. Gayathri, B. Ashok Kumar and S. Senthilrani, "Control of DC Link Voltage of Single Phase Grid Connected Solar PV System," *2018 National Power Engineering Conference (NPEC)*, 2018, pp. 1-7.

[8] A. Fethi, B. Douadi, M. Laour and B. El-madjid, "Power control of three phase single stage grid connected photovoltaic system," *2016 International Renewable and Sustainable Energy Conference (IRSEC)*, 2016, pp. 1136-1141.

[9] A. Merabet and R. K. Dhar, "Solar Photovoltaic Microgrid Simulation Platform for Energy Management Testing," *2019 Algerian Large Electrical Network Conference (CAGRE)*, 2019, pp. 1-5

[10] N. C. Sahoo, S. Mohapatro and M. K. Senapati, "A SoC based voltage control strategy for DC microgrid," *2015 IEEE Electrical Power and Energy Conference (EPEC)*, 2015, pp. 185-190,

[11] M. Abu-Zaher, Y. Atia, F. K. Abo-Elyousr and E. El-Zohri, "Experimental Realization for P&O Maximum Power Point Tracking Applied for Single-Stage Three-Phase Grid-Connected Photovoltaic System," *2019 21st International Middle East Power Systems Conference (MEPCON)*, 2019, pp. 704-709

[12] L. M. Satapathy, A. Harshita, M. Saif, P. K. Dalai and S. Jena, "Comparative analysis of boost and buck-boost converter in photovoltaic power system under varying irradiance using MPPT," *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 2018, pp. 1828-1833.




Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:24:53 UTC from IEEE Xplore. Restrictions apply.

*2022 3rd International Conference for Emerging Technology (INCET)*
*Belgaum, India. May 27-29, 2022*

# Real-time Vehicle Detection, Classification & Tracing Using State-Of-The-Art Dataset

Aadhar Dutta
*Department Information Technology*
*Delhi Technological University*
Delhi, India-110042
aadhardutta_2k18it001@dtu.ac.in

Aman Jha
*Department Information Technology*
*Delhi Technological University*
Delhi, India-110042
amanjha_2k18it018@dtu.ac.in

Rajat Rao
*Department Computer Science & Engineering*
*Delhi Technological University*
Delhi, India-110042
rajat_2k18co279@dtu.ac.in

Dinesh Kumar Vishwakarma
*Department Information Technology*
*Delhi Technological University*
Delhi, India-110042
dinesh@dtu.ac.in

2022 3rd International Conference for Emerging Technology (INCET) | 978-1-6654-9499-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/INCET54531.2022.9824516

***Abstract*— In Today's era, real-time Detection of Vehicles has become highly challenging because, mainly every coming year, various new Vehicles of different models, different skeleton structures, and varied colours are produced. These vast differences make Real-time detections of the Vehicles Approach invalid and less accurate, thus causing an imbalance in the requirement. Furthermore, the real-time Detection of Vehicle Approach is not well trained on Day-Night images, further increasing a loss of response in the frame of time needed for Real-time Detection. The previously developed approach does not process or enhance the Day-Night Frames, which bring down the response time, and responding accuracy has been seen to go drastically down. We solely aim to improve the automobile classification accuracy at Daytime-Nighttime & Night-Time and provide experimental results. We work on classifying the Vehicle type better to give a further matured accuracy after working on Vast Classified Vehicle Types. Detection of the object under Night-time conditions such as soft illumination, an excessive amount of Glare, non-visible contrast makes the process of classifying object during Night-Time a challenging task. The development is distributed into several segments/steps, respectively casing various prerequisites and stipulations for the submission's efficacious accomplishment and disposition. We conducted a unique experiment on our Self-Generated Dataset (VEHICLE_BRL_DTU). This unique and novel experimentation helped us find that the accuracy of our approach is very high if the Real-time Videography per frame rate is run through our enhancing application, thus further pushed to our application.**

***Keywords— Real-time Vehicle Detection, YOLOV5, ResNeXt50, CCTV, VGG199, L-STM, R-CNN, Efficient Net.***

## I. INTRODUCTION

Many traffic video surveillance approaches have been introduced in recent years, providing detailed traffic information, such as traffic flow and automobile speed[1][2][3]. Therefore, the acquisition of vehicles from these street view images is an integral part of an intelligent navigation system monitoring system.

From a computer perspective, automotive acquisition aims to find an object of interest, feature output and recognition, i.e., a car. Therefore, orthodox techniques use various image processing ways to discover motors vehicle[4], in-depth learning, many forms constructed proceeding the CNN to find a car. However, deep learning methods require many handwritten footnotes to train CNN[5]. Moreover, many different images need to be handwritten[6]. Moreover, artificial labelling is dynamic, consuming time, making it essential to take current annotated data to help new annotated data.

At hand remain at present numerous handwritten evidences of fresh metropolitan reconnaissance cameras in the light up day pictures of the competent prototypical for C.N.N. Concurrently, here is diminutive or none unbiased handwritten actuality in the evening pictures[7]. This paper focuses on research to make the most of day-time photographs (Location Source) to facilitate vehicles' detection in unregistered night images. Our attempt, directly using the CNN archetypal accomplished at Foundation Realm to locate Aim Realm vehicles, show deficient performance[8]. Owing to modifications in the field provision of Root Demesnes and Goals. The night-time pictures are dissimilar from the day-time imageries, the tough to depict the environment, the changing lamp posts, the wavy vision, etc. [9].

We suggest using CNN via Domain Adaptation (DA) to get a car overnight to lessen the realm scattering capriciousness for Realms and Cause Identification. The DA is a representative means of transmitting learning. In general, where the allocation of cause realm data and aim realm is dissimilar. The DA can better use the two domains' combined data to improve the aim realm's presentation[10]. The CNN prototypical used in the projected attainment automobile is More rapidly R-CNN [11] due to its better-quality meticulousness and rapidity in procurement the purpose. System is to transfer style between day-time and night-time pictures by Generative Adversarial Networks (GAN). The uninterrupted rendition system Phase GAN [12]is used in this style transfer.

We've renamed object detection as one retrieving problem, directly from the image's pixels to the linking box and classroom opportunities[13]. Our model is consumed; you will only look once within the picture towards forecasting whatever articles. YOLO is pleasingly modest. Unique concurrent resolution system forecasts more necessary squares and erudition opportunities in those squares. YOLO Pullman's pictures effusively and unswervingly improve encounter enactment. Incorporated ideal has numerous recompenses moreover to outmoded



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:25:50 UTC from IEEE Xplore. Restrictions apply.

approaches of entity procurement. Majorly YOLO has a rapid performance.

The main contribution of the work is as under:

- We generated a new diverse Dataset of Vehicle detection (VEHICLE BRL DTU) from CCTV footages which were broken down to frames thus then manually annotated to produce diverse and clear properties of different vehicles.
- Our Novel research approach on accuracy for the State-of-the-Art Models on Real time vehicle detections is one our most cherished contribution.
- We have brought about a technique to further amplify Realtime Detection by Comparing Night-time and Day-time conditions on a single run.
- Our analysis on how Day-time data can support night-time data for classification, and how Object detection models behave on Greyscale dataset has brought over significant upgradation to the earlier State-of-the-art models.
- Unique Analysis on how models trained in different environment (Colored / Black & White / Night / Day) can inter-work.

## II. Related Work

Apparition-established automobile calculating is an exhilarating supercomputer inspecting problematic incrustation dissimilar approaches. Weighing up on the tax implemented in[14], the scheming technique is broadly divided into five main categories: frame-based calculation][15], acquisition calculation[16], motion-based calculation, population-based calculation [17]and in-depth study based calculations. While demographics follow a controlled approach, they do poorly on highly visual videos and contain large vehicles. Methods supported by human reasoning also are limited in their range of acquisition and inability to follow an object. Finally, all told of those computational methods. In-depth learning-based computational methods have made great strides in recent years. Advances within the construction of their existing buildings significantly improved the performance of the Realtime automobile counting system. During this study, we focus mainly on the study of arithmetic methods supported in-depth construction. Awang et al. in [18]planned an in-depth learning process that supports automobile number tables supported a skipping layer within the convolutional neural network (CNN). Before doing the calculations, their method divides the car into different categories consistent with their other characteristics.

Dai et al. in[5] AV constructed vehicle computing procedure are prompted by means of three-step object acknowledgement, sprawling and mark out examination. Their technique uses a trail control algorithm that truthfully analyzes automobiles' numeral in their groupings and trails automobile routes to get traffic information. This method incorporates the remaining reading along the primary panache deposits to sum total the Pullmans in single guise. More recently, one existed exposed that 1 to 1 optical performances devour the probable to appear at mutually rapidity and accuracy [4] valuable necessities for article and settlement acknowledgement. This, in turn, may convince be beneficial in processing image frames at breakneck speeds that may accurately predict car counting in real-time situations.

From earlier state-of-the-art models, it has been seen that most of the approaches suffer from first and foremost exhaustive research dataset which causes unfortunate training model complications. Furthermore as discussed the Realtime Detection models have significantly faced major errors due to less training data on varied vehicle types, such issues arise due to the fact that in context of trucks there are different sub-categories which are not accounted for in the above state-of-the-art models. In context of trucks itself there are sub-categories based on Axel type and whether it is an HMV or HCV. Real-time detection of vehicle approaches that have been stated above have not been able to process in Day-Night Time data as on the on-set of Sunset the Color code of the frames changes thus causing a requirement of training on certain timeline data to get better results.

## III. Dataset

Partaking and building a Convention Data-set of 20 Automobile Modules comprising Pictures from in cooperation Day_Time and Night_Time pictures.

Complete Day_Time Pictures: 34,567

Complete Night_Time Pictures: 26,780

Examples of research data on the 20 Indian Automobile classes with their respective Day-Time Count & Night-Time Count can be seen below in Table I:

TABLE I. Represents the total Number of Annotations obtained

| S.No | Automobile Class Name | Day Time Count | Night Time Count |
|---|---|---|---|
| 1.* | Cycle | 4855 | 1.997 |
| 2. | Two-Wheeler | 74,016 | 26,799 |
| 3.* | Cycle-Rickshaw | 283 | 209 |
| 4.* | E-Rickshaw | 710 | 778 |
| 5.* | Handcart | 350 | 62 |
| 6.* | Animal Cart | 1 | 5 |
| 7. | Passenger Auto | 44,492 | 14,614 |
| 8.* | Commercial Auto | 4,024 | 496 |
| 9. | Four-Wheeler | 75,213 | 48,758 |
| 10.* | Mini Light Commercial Vehicle | 8,463 | 2,805 |
| 11.* | Light Commercial Vehicle | 8,010 | 3,120 |
| 12.* | Minibus | 895 | 313 |
| 13. | Bus | 3,311 | 1,915 |
| 14.* | Two Axle Trucks | 3,252 | 1,713 |
| 15.* | Three Axle Trucks | 1,655 | 888 |
| 16.* | Four to Six Axle Trucks | 2,112 | 1,034 |
| 17.* | More than 6 Axle Trucks | 17 | 7 |
| 18.* | Heavy Class Machinery | 2,326 | 9 |
| 19.* | Tractor without Trailer | 149 | 7 |
| 20.* | Tractor with Trailer | 209 | 0 |



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:25:50 UTC from IEEE Xplore. Restrictions apply.

The data derived is sourced to the Client's origin and hence would not be termed public without their prior agreement. The Dataset was further augmented into multiple subset datasets.

Below is a *flowchart* depicting Data Augmentation & Division flowchart, Thus Fig 1. Flowchart rightfully explains the throughout division with the counts on image thus helping understand in-depth Augmentation & Division process.

Fig. 1. *Data Augmentation & Division flowchart*

## IV. DIGITAL_IMAGE_PROCESSING

### A. *Procedure I: Divergence Augmentation*

Our model used OpenCV Library in Python to :

- Smear Histogram equalization for contrast improvement.
- Application of the (CLAHE) the histogram estimisation to overflood the amplification contrast. Fig 2, 3 & 4 presented below to explain Technique I:

Fig. 2, 3 & 4 below consists of two sub images where the (a) part of the Fig. displays the Unique carbon copy and (b) part of the Fig. shows the improved picture processed after the application of the DIP Technique I:

(a) (b)

Fig. 2. Application of Digital Image Processing technique I of night time: a Displays the original image. b Displays the enhanced image.

(a) (b)

Fig. 3. Application of Digital Image Processing technique I of day time

(a) (b)

Fig. 4. Application of Digital Image Processing technique I of night time

### B. *Procedure II: Scowl Decline*

- Smear Gaussian Filtering (Gaussian Blur) to remove Gaussian Noise from the Image and Smoothen it.
- Edge the image above 200_Intensity_value.
- Differentiate entity borders by means of 'measure_label' from 'Scikit _Learn' frame_work.
- Camouflaging snowy areas, enhacing the dissimilarity and the vividness adjacent the areas of high-intensity pixels.

In conclusion, denoising the picture for a second time although regulating contrast using 'fast_N1_Means_Denoising'.

Fig. 5 & 6 below consists of two sub images where the (a) part of the Fig. shows the Unique picture and (b) part of the Fig. shows the improved picture processed after the application of the DIP Technique II:



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:25:50 UTC from IEEE Xplore. Restrictions apply.

(a) (b)

Fig. 5. Application of Digital Image Processing technique II of night time

(a) (b)

Fig. 6. Application of Digital Image Processing technique II of day time

## V. YOLO V5 TRAINING CONCLUSIONS

The following Table II depicts all TrainingExperiments and their testing results on Day_Time and Night_Time Test-Set,

To further explain how each experiment after running through the methodologies on configuration models in Table II, we depict all the standards obtained on each experiment number for Day-time and Night-time data in Table III, thus a brief understanding can be depicted using Table III:

The above table under the row Experiment number 8 shows values 31.262 & 24.719 have been marked as bold as they show us the best weighted Standard Deviation for the Daytime greyscale & Nighttime greyscale mix in context to Model L.

The above table under the row Experiment number 12 shows values 41.904 & 40.809 have been marked as bold as they show us the best weighted Standard Deviation for the Nighttime standard & Nighttime greyscale mix in context to Model S.

From the above table under the row Experiment number 16, We can see values 41.252 & 41.798 have been marked as bold as they show us the best weighted Standard Deviation for the Daytime Standard & Daytime greyscale mix in context to Model S. Thus, from the testing shown on '*Table III'* results we can infer the above **Conclusions**:

TABLE II. REPRESENTING VARIOUS TRAINING EXP DETAILS.

| Exp No. | Data-Set | Config. | Methodology |
|---|---|---|---|
| 1. | Day Best | L | Trained from Scratch on Yolov5l Coco-dataset weights. |
| 2. | Day Best | L | Froze the Backbone and trained for 100 epochs using Exp 1 weights. |
| 3. | Day Grey Best | L | Trained from Scratch on Yolov5l Coco-dataset weights. |
| 4.* | Night | S | Trained from Scratch on Yolov5s Coco-dataset weights. |
| 5.* | Night Grey | S | Trained from Scratch on Yolov5s Coco-dataset weights. |
| 6. | Night Best | L | Trained from Scratch on Yolov5l Coco-dataset weights. |
| 7. | Night Best T.F. | L | Froze the Backbone and trained for 100 epochs using Exp 1 weights. |
| 8. | Grey Mix Best | L | Trained from Scratch on Yolov5l Coco-dataset weights. |
| 9. | Grey Mix Big Best | L | Trained from Scratch on Yolov5l Coco-dataset weights. |
| 10. | Grey Mix Big Best | M | Trained from Scratch on Yolov5m Coco-dataset weights. |
| 11. | Grey Mix Big Best | S | Trained from Scratch on Yolov5s Coco-dataset weights. |
| 12.* | Grey Mix Best | S | Trained from Scratch on Yolov5s Coco-dataset weights. |
| 13.* | Grey Mix Big | S | Trained from Scratch on Yolov5s Coco-dataset weights. |
| 14. | Grey Mix Big | S | Trained from Scratch on Yolov5s Coco-dataset weights. |
| 15.* | Day Night Mix | S | Trained from Scratch on Yolov5s Coco-dataset weights. |
| 16.* | Day Night Grey Mix | S | Trained from Scratch on Yolov5s Coco-dataset weights. |
| 17* | Day | S | Trained from Scratch on Yolov5s Coco-dataset weights. |



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:25:50 UTC from IEEE Xplore. Restrictions apply.

TABLE III. DEPICTING THE STANDARDS OBTAINED ON DAY-TIME AND NIGHT-TIME DATA BY ALL EXPERIMENTS APPLIED.

| Exp. No. | Day Time Standard | Day Time Greyscale Standard | Night Time Standard | Night Time Greyscale Standard |
|---|---|---|---|---|
| 1. | 36.597 | 53.491 | 68.785 | 71.404 |
| 2. | 37.131 | 53.749 | 67.723 | 70.517 |
| 3. | 68.262 | 71.320 | 46.468 | 48.814 |
| 4.* | 73.508 | 74.804 | 48.800 | 50.292 |
| 5.* | 77.290 | 77.089 | 55.018 | 54.904 |
| 6. | 67.406 | 70.303 | 42.409 | 43.678 |
| 7. | 71.605 | 76.999 | 66.566 | 69.735 |
| 8. | 31.818 | **31.262** | 25.140 | **24.719** |
| 9. | 44.735 | 44.227 | 35.899 | 35.672 |
| 10. | 44.330 | 43.737 | 48.072 | 47.091 |
| 11. | 49.864 | 48.835 | 54.449 | 53.6052 |
| 12.* | 45.252 | 44.236 | **41.904** | **40.809** |
| 13.* | 47.933 | 47.549 | 52.263 | 52.101 |
| 14. | 49.315 | 49.076 | 52.778 | 52.254 |
| 15.* | 45.048 | 54.255 | 50.817 | 53.185 |
| 16.* | **41.252** | **41.798** | 48.133 | 48.184 |
| 17.* | 42.005 | 55.67 | 66.608 | 68.900 |

1. We also notice that, when models are trained using '*Day-Time Greyscale Data'*, the accuracy achieved for both Coloured and Greyscale Images are nearly equal.
2. Comparing results obtained from '*Experiment 4*' and '*Experiment 5*' we can infer that **Augmenting' *Night-Time data'* into Greyscale may provide decreased accuracy**.
3. Taking inference from '*Experiment 8*' and '*Experiment 9'* we conclude that a dataset containing a **mix of Grey Scale Day-Time and Night-Time images performs appreciably good**. (20: 80 Ratio respectively). As we increase the ratio, the performance of the model decreases. The same observation is also observed between '*Experiment 12'* and '*Experiment 13'*.
4. Comparing results obtained from '*Experiments 16'* and '*Experiment 17', we can concur that creating a mix of 'Day-Time Data', 'Night-Time Data', and* **a set of all Coloured Images converted to Greyscale may enhance the accuracy**.
5. Comparing results obtained from '*Experiments 9'*, '*Experiment 10'* and '*Experiment 11'* we infer that, as the **model size decrease**, the **accuracy over the same dataset may decrease**.
6. Concluding from the results obtained from '*Experiment 7*', The approach of Transfer learning from '*Day-Time Data'* to '*Night-Time data'* is ineffective.

## VI. YOLOV5 TRAINING RESULT

### *A. Best Results Obtained for Model L*

Aimed at Model L of Yolov5 the preeminent scores are acquired on '*Grey Mix Best'* Dataset. Model has received good-quality greyscale data with contains uniform distribution of classes. Also, the ratio of '*Day-Time Data'* and '*Night-Time Data'* per class is such that they support each other detection and classification. (Refer to '***Experiment 8'***).

Thus to further explain the results obtained for Model L we have present in Fig 10 for explanation purposes, Fig 10 will help explain in concrete about the distribution & count of day time and night time mix to greyscale thus explaining through an accurate bar chart formation.

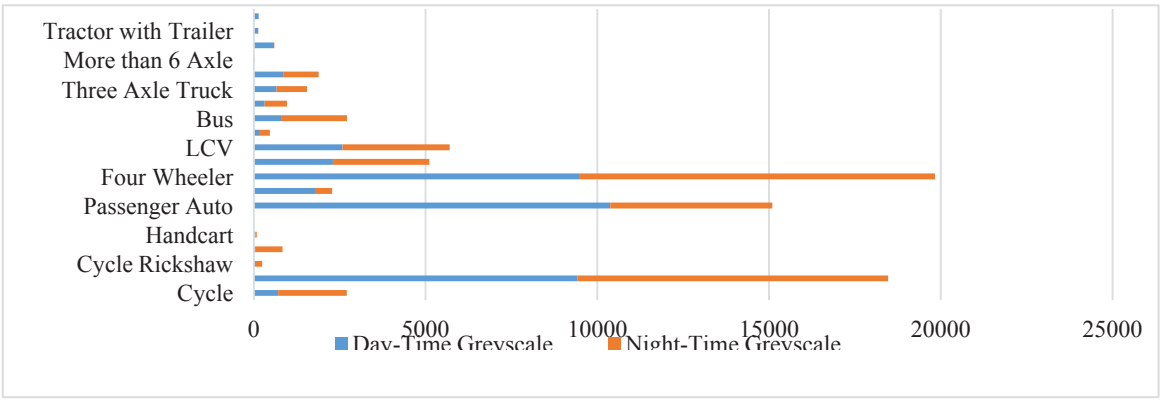



Fig. 7. Depicting the distribution and count of 'Grey Mix Best' Dataset



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:25:50 UTC from IEEE Xplore. Restrictions apply.

The trained model can be successfully run by converting by converting all visual data into Grey-Scale (Refer '*Table III*').

Accuracy acquired by the proficient Model on **I.O.U.** of **0.8**, and Confidence of **0.8**:

1. Day_Time_Greyscale_Weighted_Accuracy: **90.803%**
2. Night_Time_Greyscale_Weighted_Accuracy: **97.50%**

To further explain the Results from the Bar Chart 1 regarding the Day-time greyscale and Night-Time greyscale we indulge into creating the confusion matrix for Model L for both Day-Time and Night-Time Greyscale respectively which are present below names Fig. 11 for Day-time Greyscale & Fig. 12 for Night-timeGreyscale:

The results of the Daytime greyscale test set on Model L confusion matrix helps our cause exceptionally by explaining important aspects that are listed below:

1. High Accuracy can be obtained on Day-Time Indian Vehicles.

2. On Classes with High Annotations, the Model trained on a 'Grey-scale Day-Time + Grey-scale Night-Time" Dataset can provide a weighted accuracy above **80%**.

3. The issue of under sampling can also be addressed by increasing the annotations of Less annotated vehicles.

The results of this confusion matrix help a lot our cause by explaining important aspects that are listed below:

- High accuracy can be obtained on Night-Time Indian Vehicles if the model is trained on the dataset with "Day-Time Vehicular Objects in Greyscale"
- On Classes with High Annotations, the Model trained on a 'Grey-scale Day-Time + Grey-scale Night-Time" Dataset can provide a weighted accuracy above 87%.

## VII. Best Results Obtained of Yolov5s

Since time is an important measure of performance, we have selected Yolov5 for faster Image processing. (Refer '*Table IV*': **5.39% faster** than model L on average).

We have trained the Yolov5 Model on a Dataset Combined using:

We have created a **Day-night classifier** using OpenCV module, where we resize the image into 3x3. The image is black-and-white if all the pixels of the diagonal have the same value for the three RGB channels. The distribution of the dataset among the 20 classes is shown in '*Bar Chart II*'.

Table IV has been created below to better explain the time taken to process at different Traffic conditions for Model S and Model L respectively:

TABLE IV. Represents the processing Time taken (s) by Model S and Model L on the same traffic conditions. All videos are 60 sec's length.

| Traffic Conditions | Model S | | Model L |
|---|---|---|---|
| Day-Time High Traffic | 193.541s | | 199.871s |
| Night-Time High Traffic | 137.462s | | 149.501s |
| Day-Time Mild Traffic | 161.799s | | 165.423s |
| Night-Time Mild Traffic | 105.831s | | 115.248s |

Thus to further explain the results obtained for Model S we have present Fig 13 for explanation purposes, Bar Chart 2 will help explain in concrete about the distribution & count of day time and night time mix to greyscale and moreover it will all depict Day Night Grey Mix Final which showcases the final case

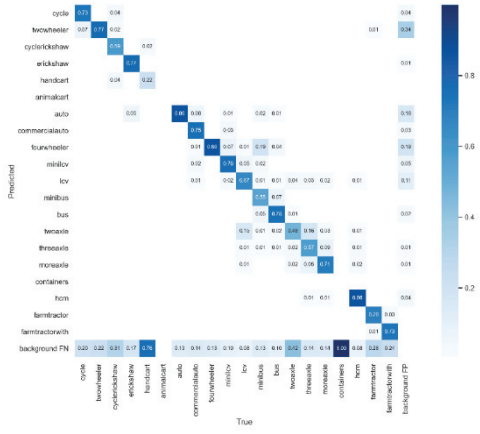



Fig. 11 denotes Confusion_matrix obtained on the Daytime Greyscale test-set

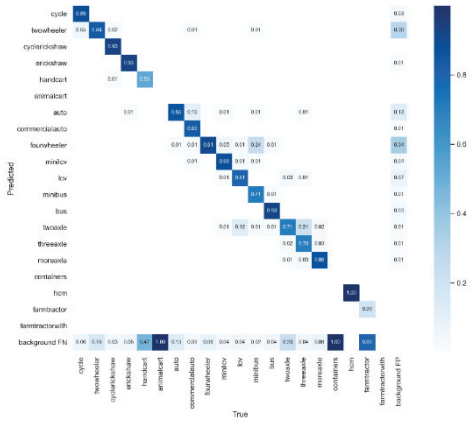



Fig. 12 denotes Confusion_matrix obtained on the nighttime Greyscale test-set



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:25:50 UTC from IEEE Xplore. Restrictions apply.

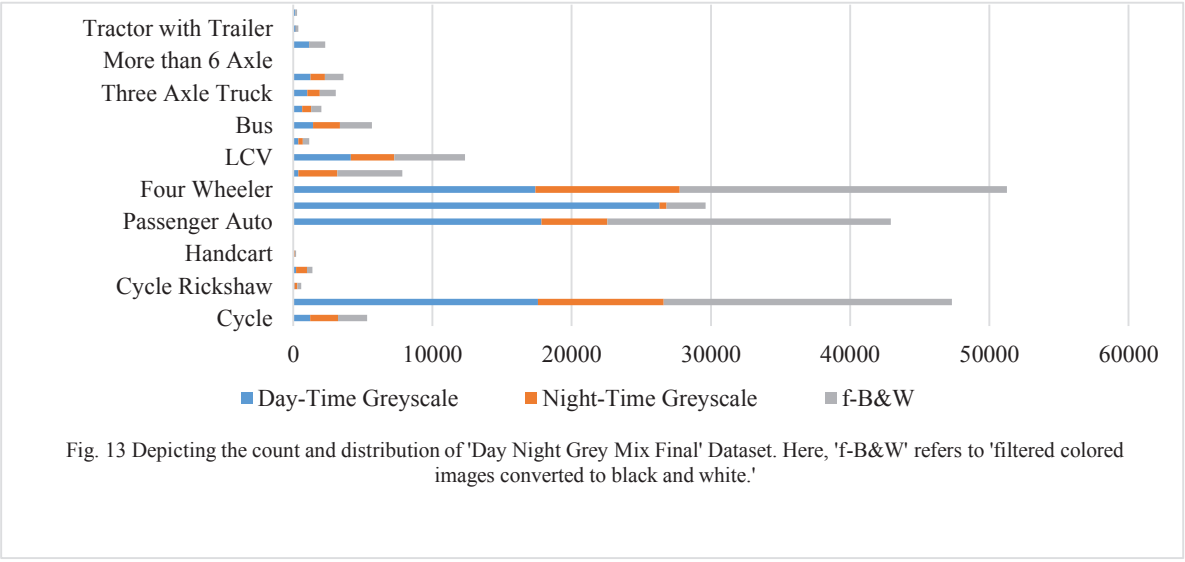



Fig. 13 Depicting the count and distribution of 'Day Night Grey Mix Final' Dataset. Here, 'f-B&W' refers to 'filtered colored images converted to black and white.'

The standard achieved thru proficient modelling, I.O.U. of **0.95** & Confidence of **0.95**:

- Day_Time_Standard: **40.910**
- Day_Time_Grey_Standard: 41.778
- Night_Time_Standard: **45.201**
- Night_Time_Grey_Standard: 45.755

Thus, the model can be successfully run in Day-Time and Night-Time environments as it is. The accuracy acquired thru the proficient Model, I.O.U. of **0.83** & Confidence of **0.83**:

- Day_Time_Weighted_Accuracy: **89.887%**
- Night_Time_Weighted_Accuracy: **90.463%**

Thus, we can **reach an accuracy of 90%** for both Night-Time and Daytime environments. To further explain the Results from the Bar Chart 2 regarding the Day-time greyscale and Night-Time greyscale we indulge into creating the confusion matrix for Model S for both Day-Time and Night-Time Greyscale respectively which are present below names Fig. 14 for Day-time Greyscale & Fig. 15 for Night-time Greyscale:

The results of the Daytime greyscale test set on Model S confusion matrix helps our cause exceptionally by explaining important aspects that are listed below:

- High accuracy can be obtained on Day-Time Indian Vehicles.
- On Classes with High Annotations, the Model trained on a 'Grey-scale Day-Time + Grey-scale Night-Time" Dataset, is able to provide a weighted accuracy above **79%**.
- The issue of under sampling can also be addressed by increasing the annotations of Less annotated vehicles.

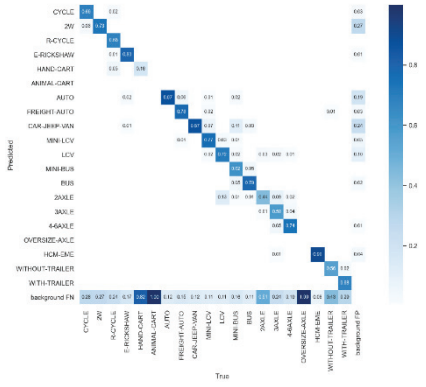


Fig. 14 represents the Confusion matrix obtained on Day-Time Greyscale test-set

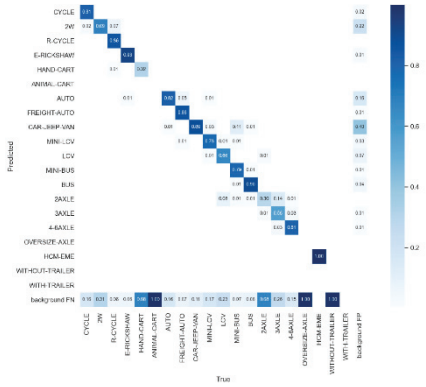


Fig. 15 represents the Confusion matrix obtained on the Night-Time Greyscale test-set.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:25:50 UTC from IEEE Xplore. Restrictions apply.

The results of this confusion matrix for Night Time Grey Scale for Model S helps our cause by explaining important aspects that are listed below:

- By training on the dataset with "Day-Time Vehicular Objects in Greyscale" a comparative High Accuracy can be obtained on Night-Time Indian Vehicles if the model.
- 'Grey-scale Day-Time + Grey-scale Night-Time" Dataset, is able to provide a weighted accuracy above 80% on Classes with High Annotations, that is on Model S.

## REFERENCES


[1] Janua, pp. 315–319, 2018, doi: 10.1109/ICCSNT.2017.8343709.

[2] H. Kuang doi: 10.1109/TITS.2017.2702665.

[3] K. Huang, L. Wang, T. Tan, and S. Maybank, "A real-time object detecting and tracking system for outdoor night surveillance," *Pattern Recognit.*, vol. 41, no. 1, pp. 432–444, 2008, doi: 10.1016/j.patcog.2007.05.017.

[4] Z. Dai *et al.*, "Video-based vehicle counting framework," *IEEE Access*, vol. 7, pp. 64460–64470, 2019, doi: 10.1109/ACCESS.2019.2914254.

[6] J. Ji, Z. Xu, H. Yu, L. Fu, and X. Zhou, "Domain Adaptation For Vehicle Detection In Traffic Surveillance Images From Daytime To Nighttime," ... *Res. Board 99th ...*, no. March, 2020, [Online]. Available: https://scholarworks.utrgv.edu/cs_fac/21/.

[7] J. Tripathi, K. Chaudhary, and A. Joshi, "International Journal of Innovative and Emerging Research in Engineering Automatic Vehicle Counting and Classification," *Int. J. Innov. Emerg. Res. Eng.*, vol. 2, no. 4, pp. 32–36, 2015, [Online]. Available: www.ijiere.com.

[8] A. Crouzil, L. Khoudour, P. Valiere, and D. N. Truong Cong, "Automatic vehicle counting system for traffic monitoring," *J. Electron. Imaging*, vol. 25, no. 5, p. 051207, 2016, doi: 10.1117/1.jei.25.5.051207.

[9] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv*, 2020.

[10] V. Mandal and Y. Adu-Gyamfi, "Object detection and tracking algorithms for vehicle counting: A comparative analysis," *arXiv*, no. March 2021, 2020, doi: 10.1007/s42421-020-00025-w.

[11] "No Title," doi: 10.1371/journal.pone.0120644.

[12] A. Ramanathan and M. Chen, "Spatiotemporal Vehicle Tracking, Counting and Classification," *Proc. - 2017 IEEE 3rd Int. Conf. Multimed. Big Data, BigMM 2017*, pp. 246–247, 2017, doi: 10.1109/BigMM.2017.85.

[13] S. Awang and N. M. A. N. Azmi, "Vehicle counting system based on vehicle type classification using deep learning method," *Lect. Notes Electr. Eng.*, vol. 449, pp. 52–59, 2017, doi: 10.1007/978-981-10-6451-7_7.

[14] D. Bell, W. Xiao, and P. James, "Accurate Vehicle Speed Estimation from Monocular Camera Footage," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 5, no. 2, pp. 419–426, 2020, doi: 10.5194/isprs-annals-V-2-2020-419-2020.




Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:25:50 UTC from IEEE Xplore. Restrictions apply.

**How to Cite**:
Meena, S., Vats, K., & Kumar, D. . . (2022). Review of factors affecting facial recognition algorithms performance. *International Journal of Health Sciences*, *6*(S3), 9528–9541. https://doi.org/10.53730/ijhs.v6nS3.8253

# Review of factors affecting facial recognition algorithms performance

**Shweta Meena**
Delhi Technological University
Email: shwetameena@dtu.ac.in

**Kanishk Vats**
Delhi Technological University
Email: 210600.kv@gmail.com

**Deepak Kumar**
Delhi Technological University
*Corresponding author email: deepak9013216128@gmail.com

***Abstract*---**The face is a significant part of the human body, recognizing people in large groups of individuals. Thus, because of its uniqueness and universality, it has turned into the most generally utilized and acknowledged biometric technique. Many algorithms have been used by various researchers for face detection and recognition. Research, innovation progression, and applications consolidating face recognition in the last twenty years have grown massively. In this paper, some essential existing approaches which are adjusted with dealing with the issues of face recognition have been presented close by their Face recognition accuracy and the variables capable of debasing the performance of the review. In the first section, various factors that decrease facial detection and recognition accuracy have been researched like posture variety, illumination, aging, facial expressions, etc. While in the second section of the paper, various methods have been examined that attempt to relieve the impact of discussed factors. Various algorithms give various exhibitions in various conditions like enlightenment, noise, posture, and mask change. All the previously mentioned methods are represented briefly to give an overall idea. The motive of the paper is to carry all the various methods to a similar spot and simplify it to review the paper.

***Keywords*---**ANN, facial recognition, facial detection and recognition, gabor wavelet, LDA, PCA.

## Introduction

Facial Detection and Recognition (FDR) is one of the huge check structures reliant upon biometrics which can be used for the confirmation cycle just as reconnaissance. With the speedy extension in frauds step by step, Facial Recognition (FR) is transforming into a basic framework for us. Different uses of a capable FR framework are criminology, identification of hoodlums, extortion prohibition, surveillance, and so forth. A great deal of exploration has been done at every level, yet even after all that, no system is developed which can work in real-time under an unconstrained environment. FR has consistently been an exceptionally complicated task. Its genuine test lies in showing a computerized structure that matches the human capacity to see faces. Thus, a customized programmed electronic system with somewhat greater recognition accuracy, and quick handling is required. FDR process has steps as shown in Fig. 1:

- Face detection: To perceive the appearances in a picture or video using important features on the face like eyes, nose, mouth, and so on
- Feature extraction: Standardization, arrangement of the faces for stronger recognition exactitude.
- Face recognition: To recognize a specific individual in a picture or video by comparing the data set.

This paper is divided into five major sections. The first section introduced us to FDR. In second section, various factors that reduce FR accuracy have been discussed. The third section briefly describes some FDR methods and their importance. Section four gives a brief review of some algorithms used by researchers in recent years. Finally, the last section contains the conclusions we derived from this.

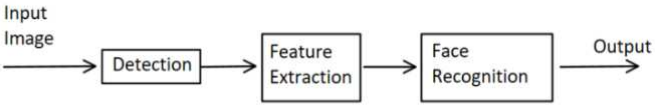



Fig 1. Steps of FDR

## Factor degrading the FR performance

FR from pictures and recordings is a formidable task. Researchers have done a great deal of work trying to accomplish 100% precision, but it is not feasible due to several factors. It has been found that the components that spoil the precision of the FR systems are: noise, occlusion, illumination, pose variation, low resolution, aging, expressions, plastic surgery, and database. These elements can be requested into two classes: natural and outward factors. Characteristic parts incorporate the normal condition of the human face like the look, maturing, plastic surgery, and so on, impacting the framework, while extraneous elements include those which affect the face from the outside like noise, occlusion, illumination, low resolution, and pose variation. However, databases are also a major factor in deciding face recognition accuracy.

**Noise**

Computerized pictures are prone to various types of noise. This noise prompts low accuracy of FDR. Noise is present in pictures in various ways that are subject to picture creation. Dealing with these before proceeding further is a basic step in these common FDR frameworks [3]. Noise is visualized in Fig. 2.

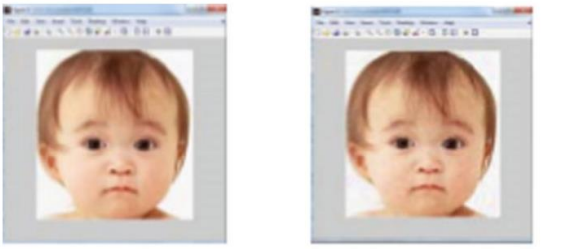



Fig. 2. Example of image having noise [3]

Fig. 3. Simulation of partial occlusion [23]

Fig. 4. Effect of illumination in face image [9]

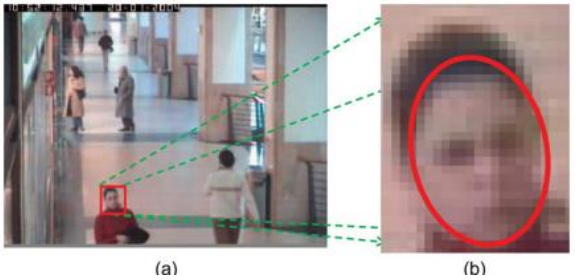



Fig. 5. A-frame from the CCTV video. (a) frame, (b) captured face [2]

**Occlusion**

One of the major challenges of FR is an incomplete face as it would be hard for even humans to recognize a person if a distinctive part of their face is hidden. For example, long hair and sunglasses can hide ears and eyes, masks or bushy beards can hide the lower half of the face, and so on. Partial occlusion is shown in Fig. 3. The accuracy of the FDR system is degraded by these factors. Various methodologies have been researched over the years to overcome these issues [1].

**Illumination**

The variations in illumination interfere with the identification of important parts of the face, thereby reducing the probability of FR. The main cause behind these varieties could be unbalanced light, shadow, contrast, darkness, etc. as shown in Fig. 4. Various methodologies related to illumination are being discussed in [4-9].

**Low Resolution**

The surveillance cameras are strict to look out for suspicious activities, hence they are not high quality enough to record high-quality videos leading to low-resolution faces as shown in Fig. 5. Comparing the low-resolution query picture with the high-resolution database picture is a difficult task. This degrades the performance of the FR system. Distinctive approaches are being researched to tackle this problem [1,2].

**Pose Variation**

Distinctive posture conveyance is another important issue of the FR system. Front-facing face recreation is needed to compare the query face with the database face. This is needed as the database contains the front-facing face. Comparing it with that different pose will yield unusable results. Various approaches were researched by the specialists to change the non-front-facing face over to the front-facing face to increase FR accuracy [10]. How different poses depreciate the performance of the FDR algorithm is being discussed in the proposed approaches [6, 11]. Pose variations are shown in Fig. 6.

**Aging**

The face is skin tissues and facial muscles. On Aging, these muscles contract or expand, leading to a change in facial highlights of an individual, as shown in Fig. 7. Although these are basic changes, they are still a factor that lowers the performance of FR systems [8, 12].

**Expressions**

We use facial expressions to express our feelings but it changes the shape of the human face due to the contraction of facial muscles. A small change like that can easily create ambiguity for the FR framework. Continuous research is being done for FR by considering facial expressions [4, 8]. Facial Expressions are shown in Fig. 8.

**Plastic Surgery**

One of the major concerning factors that affect FR accuracy. A small change will lead to tightening of facial muscles, while a major change may lead to an entirely new face, of which the database has no record. Many criminals or undercover agents go for plastic surgery to bury their identities. An example is shown in Fig. 9.

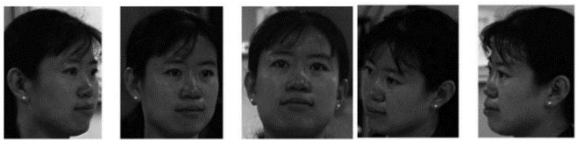


Fig. 6. Pose variations [24]

Fig. 7. Aging variations [12]

Fig. 8. Example of face having different expressions

Fig. 9. Example of Surgery effects

**Databases**

It is also a major factor in deciding FR accuracy. Databases used by researchers have different image quantities containing different combinations of the above factors, which leads to different accuracy results as their parameters are adjusted to them. Researchers only compare their results with those researches done on the same database, hence different results, since not all researchers take care of all the above factors, as it's not possible with current means. Some of the datasets used by researchers are CASIA WebFace, Colour FERET Database, LFWcrop Database, SCface - Surveillance Cameras Face Database, The Yale Face Database, and PIE Database, CMU.

**A Brief Explanation of FR Methods**

It is not possible to recreate an experiment since parameters as well as algorithm

details are not made available. That is why researchers compare the database used and the results they obtained with the previous results. A typical and well-designed protocol can quickly benchmark various methods. The FR methods used by researchers can be divided into three groups [26]:

- Appearance-based Methods
- Feature-based Matching Methods
- Hybrid Methods

**Appearance-Based Methods**

These methods are encompassing techniques where the complete face is compared with the display face. The benefits of this methodology are:

1. They focus on only restricted regions or central focal points without annihilating any of the data in pictures.
2. Information on faces and geometry isn't required.
3. Recognition is straightforward as compared to other methodologies.
4. Quick and simple.

Yet, they are very inclined to facial varieties like illumination, posture, and articulations. Likewise, acknowledgment accuracy is low.

**Feature-Based Matching Methods**

These methods compare the test face with the gallery face by utilizing neighborhood highlights of the face like eyes, mouth, nose, etc. The area highlights have been taken care of into the essential classifier to perceive the face. In this method, recognition accuracy is better than previously discussed approaches. In any case, these techniques have exceptionally high memory usage and a lot of marked and clean datasets for preparing are required.

**Hybrid Methods**

It's a mixture of both methods. Both methods are merged to improve the recognition rate. However, hybrid methods are complicated and hard to implement. Research reviews of the above methods are listed in Table I

Table 1
Recent progress made on FDR

| | Authors | Published in year | Dataset used | Algorithm used | Recognition accuracy | Review |
|---|---|---|---|---|---|---|
| Appearance-based methods | Mhadgut [31] | 2021 | Custom Dataset | YOLOv3 | 95.83% | (1) model detects the masked and non-masked face with a mean average precision of 95% |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | (2) considered occlusion and pose variation<br>(3) different YOLO models considered with the best results obtained in YOLOv3 |
| | Ejaz et al. [28] | 2019 | ORL combined with a custom dataset | PCA, Viola-Jones | 72-95% | (1) poor recognition rate on the masked face as the number of features obtained is less.<br>(2) face without a mask gives a better recognition rate |
| | Khan et al. [19] | 2018 | NRC-IIT facial database and real-time video stream | PCA | 69–86% | (1) it does not work well in an unconstrained environment<br>(2) results are not good for low-resolution videos and pose variation |
| | Abdullah et al. [5] | 2017 | Real-time video stream | PCA | 80% | (1) not useful for video data sets<br>(2) good for front view faces<br>(3) other factors are not considered<br>(4) Performance evaluation of the algorithm is not done using an actual dataset<br>(5) recognition accuracy is 80% |
| | Gao and Jong [11] | 2015 | FERET | Lucas Kanade, SIFT, two-phase alignment error | 99.5%, ±15°<br>26.0%, ±60° | (1) this approach can deal with up to ±60° pose variation<br>(2) beyond 40°, recognition accuracy drops drastically<br>(3) time complexity is very high compared to other algorithms |

TABLE 1 (continued)

| | Authors | Published in year | Dataset used | Algorithm used | Recognition accuracy | Review |
|---|---|---|---|---|---|---|
| Feature-based Methods | Gupta et al. [32] | 2021 | ORL | SIFT, SURF, Decision Tree, Random Forest | 95.5% | (1) different combinations of SIFT and SURF are used with 64 and 128 feature vector dimensions<br>(2) test conducted on 5 datasets each containing different combinations of factors affecting FR<br>(3) both classifiers were used to validate the performance of the proposed system.<br>(4) system was computationally efficient |
| | | | FACE94 | | 87.3% | |
| | | | M2VTS | | 98.8% | |
| | | | FERET | | 76.8% | |
| | | | YALE2B | | 99.7% | |
| | Fredj et al. [29] | 2020 | LFW, YTF, Training with CASIA WebFace | CNN based on GoogLeNet inception-v1 | 99.2% | (1) considered occlusion, illumination, pose variation, and noisy images<br>(2) test performed on two datasets |
| | | | | | 96.63% | |
| | Luo et al. [4] | 2018 | Training with CASIA WebFace | Unsupervised transfer learning, maximum mean discrepancy | 98.70–98.93% | (1) Doesn't work well if the alignment of images compared is different<br>(2) only image datasets are considered<br>(3) It's Focus on variable illumination, facial expression, and age but does not consider poses variation and occlusion |
| | | | GBU, FERET | | | |
| | Fu et al. [1] | 2017 | CASIA WebFace | Guided CNN, loss function | 91.9–97.1% | (1) A lot of training is required as the system is trained for every low resolution of each image<br>(2) only partial occlusion and different resolution is considered |
| | | | LFW | | | |
| | Huang et al. [20] | 2015 | COX face | Point-to-set correlation | 50.96–53.26% | (1) Aging, occlusion, and plastic surgery are not considered in the dataset |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | learning (PSCL) | | (2) No effective results were obtained |

| | Authors | Published in year | Dataset used | Algorithm used | Recognition accuracy | Review |
|---|---|---|---|---|---|---|
| Hybrid Methods | Taskiran et al. [33] | 2021 | UvA-NEMO | VGGFace, VGGFace2, ArcFace | 72.26 – 100% | (1) considered illumination, pose variation, noise, low-resolution, expressions (smile only)<br>(2) different hybrid methods are considered, having promising results for each of the factors on both datasets<br>(3) intensity of factors is varied to obtain a comprehensive result to find to which extent the said model can give acceptable accuracy |
| | | | FEEDTUM | | 71.76 – 100% | |
| | Aung et al. [30] | 2021 | FDDB | YOLOv2, VGG16 network model | 93% | (1) considered pose variation, illumination, and skin color.<br>(2) train on the high-resolution image.<br>(3) has high computational complexity. |
| | Banerjee et al. [10] | 2018 | PaSC videos and CW images, CMU multi-PIE dataset | Supervised learning, generic 3D model, and Viola-Jones | 88.45–97.28% | (1) Major failure beyond 40° pose variation<br>(2) focused on pose and illumination only<br>(3) Supervised Learning needs cleaned and labeled data |
| | Fathima et al. [21] | 2015 | AT&T, MIT-India, and Faces94 datasets | Linear discriminant analysis and Gabor wavelet | 88–94.02% | (1) Not considered pose variation |
| | Lei et al. [23] | 2009 | CMU-MIT face | Modest AdaBoost, | 95.20% | (1) a very huge number of images are needed to train |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | database | improved independent component analysis, Hausdorff distance | | the system to be capable of recognition |

**A Brief Explanation of FR Algorithms**

There are many techniques used by researchers over the years for FDR. Some of them are machine learning-based approaches, such as Principal Component Analysis and Artificial Neural Network, while there are also feature-based approaches like Gabor Wavelet. This section briefly describes some of these techniques.

**Principal Component Analysis (PCA)**

PCA was a technique given by Turk and Pentland in 1991 [13]. From the given dataset of pictures, PCA identifies the important parts, which are referred to as eigenfaces by eliminating the redundancy. Then, all the eigenfaces obtained from the database are combined to form an average face to contain the significant features obtained, like eyes, nose, mouth, etc. This allows us to represent all images as a linear combination of eigenfaces, which is achieved by adding the mean face obtained to them. After which, images are projected on the eigenspace and the Euclidean distance between the target face and the eigenspace image is calculated. If the calculated distance is below a certain threshold, the target image is classified as belonging to a certain class k. PCA performance is affected by orientation, lightning, and the size of the image. It can be used in real-time under constrained conditions, which allows it to achieve a maximum of 96% accuracy. In this method, orientation, size, and lightning in the image degrade the accuracy. This framework can be utilized as a real-time application in a confined environment. The steps of the algorithm are explained below:

- The image from the database is taken and its eigenface is calculated.
- A set of weights is obtained, each associated with the image projected in the eigenspace
- Then, the Euclidean distance is calculated between the target image and the projected image to look for the face.
- If the distance is below a certain threshold of a class, it considers the face as recognized, otherwise, it is declared unknown.

**Artificial Neural Network (ANN)**

ANN is one of the most go-to methods in this field. [14] entails an approach for ANN in which the images are first processed with PCA, then are fed to ANN. It gives good accuracy. Similarly in another approach [15], to deal with noisy images, ANN is trained in three stages, first with the ideal images, then with noisy images in the second stage, and at last again with ideal images, which allows it to achieve high precision in ideal images and average in noisy images. In [16], CNN-

based FR algorithms were implemented in Google's open-source TensorFlow software. Although ANN gives high accuracy, it gets very complex in the training stage and can take too much time. It is also difficult to decide the number of neurons in the secret layer as there is no way to arrive at a particular number.

**Gabor wavelet**

It is a method that stimulates human eyes due to having similar characteristics and orientation representation. Its representation of the image is obtained through the convolutions of the target image and Gabor filters. After image-processing, n weight points are obtained corresponding to wavelets placed along the region of importance in a portable grey map. These weights along with their set of frequencies are the required features that are used for the training and testing of neural networks. It is a feed-forward neural network that uses backpropagation. This method remains unaffected due to glasses, facial hair, pose, and illumination.

**Linear Discriminant Analysis (LDA)**

LDA is an algorithm that starts by reducing the dimensionality of pictures. It divides the pictures into various classes where each picture is represented as pixel vectors. By doing this, it finds a linear combination of images, which is later used to reduce dimensionality. After which, Euclidean distance-based classification is used.

**3D model-based FR**

In this method, the system takes images and measures the facial curves on a micrometer level. Then, a face template is created where each feature has a unique number. The same was done with database images before, allowing the system to directly compare their codes. Database images don't need to be in 3D as 2.5D scans can also be compared with 3D scans. The z-value is at max one for each (x, y) pair, making comparison possible. Constraints like illumination, lightening, and pose can be ignored while recognizing faces in this method

**Conclusion**

We have made a great deal of progress in the last two decades in the FDR region, but the problems faced have also increased, due to which the accuracy and results obtained so far are not acceptable. The primary factors which affect the results are pose variation, brightening, lightening, and environmental elements which cannot be controlled. Many face datasets have a variable composition that can be used in evaluating FR algorithms. Because of the distinction between these data sets, face datasets selection is a significant thought for research groups working at FR. They have to consider factors like dataset size, image resolution, and quality of images. Other factors are which of the above-mentioned factors are they considering and what factors that particular dataset covers. This is also the reason due to which researchers cannot actually compare their work, making it difficult to determine whose approach made more progress in FDR.

The algorithms can be classified into three categories, depending upon their approach to the image. These are appearance-based, feature-based, and the combination of two known as hybrid methods. Each category has its pros and cons depending upon the dataset used and the application of the area. Appearance-based approaches give faulty results for pose variation and low-resolution video, feature-based take too much memory and require clean and labeled datasets, while the combination of two, which is a hybrid, is very difficult to implement. The majority of the technique doesn't give satisfactory results for non-front facing faces, i.e., faces with various posture dissemination. Furthermore, the factors like plastic surgery, aging, and occlusion also influence the performance of the FR system. After this review, we have learned that the current results obtained are still not sufficient to build a dependable FR system that can work in unconstrained environment without being affected by above factors.

**References**


1. T. Fu, W. Chiu and Y. F. Wang, "Learning guided convolutional neural networks for cross-resolution face recognition," 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP), 2017, pp. 1-5
2. W. W. W. Zou and P. C. Yuen, "Very Low Resolution Face Recognition Problem," in IEEE Transactions on Image Processing, vol. 21, no. 1, pp. 327-340, Jan. 2012
3. Tin, H.H.K.: Removal of noise by median filtering in image processing. In: 6th Parallel and Soft Computing (PSC 2011) Removal (2011)
4. Z. Luo, J. Hu, W. Deng and H. Shen, "Deep Unsupervised Domain Adaptation for Face Recognition," 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), 2018, pp. 453-457
5. N. A. Abdullah, M. J. Saidi, N. H. A. Rahman, C. C. Wen, and I. R. A. Hamid, "Face recognition for criminal identification: An implementation of principal component analysis for face recognition", AIP Conference Proceedings, vol. 1891, no. 1, pp. 020002, 2017.
6. C. Ding, C. Xu and D. Tao, "Multi-Task Pose-Invariant Face Recognition," in IEEE Transactions on Image Processing, vol. 24, no. 3, pp. 980-993, March 2015
7. P. Wang, W. Lin, K. Chao and C. Lo, "A Face-Recognition Approach Using Deep Reinforcement Learning Approach for User Authentication," 2017 IEEE 14th International Conference on e-Business Engineering (ICEBE), 2017, pp. 183-188
8. T. Ahonen, A. Hadid and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037-2041, Dec. 2006
9. S. Z. Li, R. Chu, S. Liao and L. Zhang, "Illumination Invariant Face Recognition Using Near-Infrared Images," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 4, pp. 627-639, April 2007
10. S. Banerjee et al., "To Frontalize or Not to Frontalize: Do We Really Need Elaborate Pre-processing to Improve Face Recognition?," 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 2018, pp. 20-29

11. Y. Gao and H. J. Lee, 'Cross-pose face recognition based on multiple virtual views and alignment error', Pattern Recognition Letters, vol. 65, pp. 170–176, 2015.
12. Z. Li, U. Park and A. K. Jain, "A Discriminative Model for Age Invariant Face Recognition," in IEEE Transactions on Information Forensics and Security, vol. 6, no. 3, pp. 1028-1037, Sept. 2011
13. M. Turk and A. Pentland, "Eigenfaces for Recognition," in Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86
14. N. Jamil, S. Lqbal and N. Iqbal, "Face recognition using neural networks," Proceedings. IEEE International Multi Topic Conference, 2001. IEEE INMIC 2001. Technology for the 21st Century., 2001, pp. 277-281
15. A. Reda and B. Aoued, "Artificial neural network-based face recognition," First International Symposium on Control, Communications and Signal Processing, 2004., 2004, pp. 439-442
16. L. Yuan, Z. Qu, Y. Zhao, H. Zhang and Q. Nian, "A convolutional neural network based on TensorFlow for face recognition," 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2017, pp. 525-529
17. B. S. Manjunath, R. Chellappa and C. von der Malsburg, "A feature based approach to face recognition," Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1992, pp. 373-378
18. P. Campadelli, R. Lanzarotti and C. Savazzi, "A feature-based face recognition system," 12th International Conference on Image Analysis and Processing, 2003.Proceedings., 2003, pp. 68-73
19. A. Khan et al., "Forensic Video Analysis: Passive Tracking System for Automated Person of Interest (POI) Localization," in IEEE Access, vol. 6, pp. 43392-43403, 2018
20. Z. Huang et al., "A Benchmark and Comparative Study of Video-Based Face Recognition on COX Face Database," in IEEE Transactions on Image Processing, vol. 24, no. 12, pp. 5967-5981, Dec. 2015
21. A. A. Fathima, S. Ajitha, V. Vaidehi, M. Hemalatha, R. Karthigaiveni and R. Kumar, "Hybrid approach for face recognition combining Gabor Wavelet and Linear Discriminant Analysis," 2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS), 2015, pp. 220-225
22. Z. Lei, C. Wang, Q. Wang and Y. Huang, "Real-Time Face Detection and Recognition for Video Surveillance Applications," 2009 WRI World Congress on Computer Science and Information Engineering, 2009, pp. 168-172
23. Z. Lei, C. Wang, Q. Wang and Y. Huang, "Real-Time Face Detection and Recognition for Video Surveillance Applications," 2009 WRI World Congress on Computer Science and Information Engineering, 2009, pp. 168-172
24. R. Singh, M. Vatsa, A. Ross and A. Noore, "A Mosaicing Scheme for Pose-Invariant Face Recognition," in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 37, no. 5, pp. 1212-1225, Oct. 2007
25. S. Anwarul and S. Dahiya, "A Comprehensive Review on Face Recognition Methods and Factors Affecting Facial Recognition Accuracy", Jan. 2020, pp. 495–514.
26. P. Borisagar, S. Jani, Y. Agrawal and R. Parekh, "An Efficient and Compact Review of Face Recognition Techniques," 2020 IEEE International Students'

Conference on Electrical,Electronics and Computer Science (SCEECS), 2020, pp. 1-5
27. G. Castaneda and T. M. Khoshgoftaar, "A Review of Performance Evaluation on 2D Face Databases," 2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService), 2017, pp. 218
28. M. S. Ejaz, M. R. Islam, M. Sifatullah and A. Sarker, "Implementation of Principal Component Analysis on Masked and Non-masked Face Recognition," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019, pp. 1-5
29. H. Fredj, S. Bouguezzi, and C. Souani, "Face recognition in unconstrained environment with CNN", The Visual Computer, vol. 37, Feb. 2021.
30. H. Aung, A. V. Bobkov and N. L. Tun, "Face Detection in Real Time Live Video Using Yolo Algorithm Based on Vgg16 Convolutional Neural Network," 2021 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), 2021, pp. 697-702
31. S. Mhadgut, "Masked Face Detection and Recognition System in Real Time using YOLOv3 to combat COVID-19," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021, pp. 1-7
32. S. Gupta, K. Thakur, and M. Kumar, "2D-human face recognition using SIFT and SURF descriptors of face's feature regions", The Visual Computer, vol. 37, no. 3, pp. 447–456, Mar. 2021.
33. M. Taskiran, N. Kahraman, and C. Eroglu Erdem, "Hybrid face recognition under adverse conditions using appearance-based and dynamic features of smile expression", IET Biometrics, vol. 10, no. 1, pp. 99–115, 2021

*2022 3rd International Conference for Emerging Technology (INCET)*
*Belgaum, India. May 27-29, 2022*

# Sign Language to Text Conversion using Deep Learning

Dr. Aruna Bhat
*Computer Science and Engineering*
*Delhi Technological University*
New Delhi, India
aruna.bhat@dtu.ac.in

Vinay Yadav
*Computer Science and Engineering*
*Delhi Technological University*
New Delhi, India
vinayyadav_2k18co387@dtu.ac.in

Vishesh Dargan
*Computer Science and Engineering*
*Delhi Technological University*
New Delhi, India
visheshdargan_2k18co389@dtu.ac.in

Yash
*Computer Science and Engineering*
*Delhi Technological University*
New Delhi, India
yash_2k18co401@dtu.ac.in

2022 3rd International Conference for Emerging Technology (INCET) | 978-1-6654-9499-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/INCET54531.2022.9824885

***Abstract*—Signs and gestures are a way of communication that, instead of spoken words, makes use of movements and gestures made by hands along with changes in facial expression and changes in bodily movements. There are many distinct sign languages throughout the world, just like there are many different spoken languages. Signs and gestures are primarily used by people who cannot hear and speak and use this to express their emotions and interact with each other**

**In ancient times, gestures and signs were the only way of interacting with each other, since there were no spoken languages and was considered an obvious and natural way of communicating. However, as the time passed, the utilization of sign language started becoming more prevalent only for the people who are deaf or have any hearing impairments.**

**Currently, over 80 million people in the world use gestures made by hands and symbols, which is a small part of the 780 crores people living on Earth as of 2020. These people depend greatly on sign languages to learn, access services and be a part of the communities. However, there is a lack of support in terms of providing basic services because of the fewer sign interpreters available.**

**To overcome the challenges and issues created due to the dearth of information and knowledge relating to sign languages and to provide services that people deserve, there is a need to spread the use of sign languages among the general public. However, this involves a lot of effort and does not provide the results that are equivalent to the labour done. Detecting the correct signs in images can be a complex process and can include a variety of pre-processing techniques to be performed. In this project, various images containing the signs of the American-Sign-Language have been processed and utilised for implementing a CNN model to classify these images as their alphabetical equivalent.**

***Keywords—Convolution Neural Network, Image Classification, Sign Language Recognition, Callbacks, Deaf & Mute***

## I. INTRODUCTION

Gestures and signs are one form of verbal communication which is leveraged by individuals who have lost their ability to speak or hear. According to World Health Organization data, more than 6% of the total people living on earth – 3600 lakh people – suffer from inability to hear and comprehend what others are saying. More and more deaf individuals are losing their jobs and the rate of unemployment for deaf workers, in particular, is around 70%, creating a lot of pressure from society on them. People who do not face any problems hearing and talking find it tougher to communicate with mute and deaf people. Signs and hand gestures which together form sign language is the paramount means of interacting for people who cannot talk and have difficulty listening to other people. Hearing loss that is more than 40 dB in the stronger ear is mentioned as debilitating hearing loss. Hence with the rising number of individuals who are deaf, the demand for interpreters is also increasing. Minimizing the verbal communication gap between people who have difficulty in hearing and non-hearing impaired individuals becomes a desire to ensure that everyone can communicate effectively. Sign language translation is one of the quickest-growing fields of study currently, and it is the most natural mode of communication for those who are partially deaf.

Deep learning is a data-driven machine learning methodology that uses a variety of neural network architectures to perform various tasks like object detection, segmentation, and classification in imaging. Deep learning methods use automatic feature extraction as a method of learning and have been widely used for the purpose of classification. This also includes the classification of sign languages which can further be extended to sign to text conversion.

One thing that needs to be taken into consideration during image classification is that the data should be of good quality and should have as little noise as possible. This means that in the case of the classification of sign images, any noise or background elements can affect the performance of the classification. To avoid this, high-quality images can be used and using a large dataset can also benefit the process of classification. A good classification model must be able to utilize even small size of data to identify the signs and provide the corresponding alphabets or characters.

## II. RELATED WORK

Izutov and Evgeny[1] trained the ASL gesture recognition model was that could identify ASL gestures in a live stream and were trained using the metric-learning framework. Furthermore, the authors suggested residual spatio-temporal attention with the auxiliary self-supervised loss for improved model robustness to appearance changes.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:24:18 UTC from IEEE Xplore. Restrictions apply.

The results from their study suggested that the proposed gesture recognition model may be utilised to recognise ASL signs in a real-world situation.

Pigou, Lionel, Mieke Van Herreweghe, and Joni Dambre [2] concluded that deep residual networks are capable of learning different patterns in continuous gesture and sign language films using basic RGB cameras and with very little preprocessing. They consider sign language & gesture recognition as a framewise classification problem. They have used temporal convolutions as well as current breakthroughs in the area of deep learning such as batch normalisation, residual networks, and (ELUs)exponential linear units to solve it. To test the trained models, the authors employed two different sign language corpora and the biggest known gesture dataset. The ConGD, the Dutch Sign Language Corpus and the Flemish Sign Language Corpus (Corpus VGT) were used to test the models built by the authors (Corpus NGT). The accuracy of the Corpus NGT was 73.3 percent, whereas the accuracy of the Corpus VGT was 55.7 percent.

Enriquez, Manuel, Jose et.al [3] presented the usage of the skeleton-based MS-G3D architecture for Isolated Sign Language Recognition. The argument behind this decision is that it allows for a more reliable semantic link between body parts and hands in sign and gesture language dynamics. The authors compared their technique to the S3D, a SOTA solution based on raw RGB input. MS-G3D outperforms S3D if we compare it in independent and fusion training on shared ISLR (AUTSL) data, with accuracy comparable to the best on the AUTSL validation set of the 2021 ChaLearn LAP LSSII SLR Challenge (RGB TRACK)6. The researchers also looked at the impacts of transfer learning while training MS-G3D models, and found that a model that had been previously trained and has pre trained weights on a much larger dataset in a dissimilar language and under somewhat non identical acquisition conditions did not help with the ISLR task for medium vocabulary. The top 1% accuracy for the baseline model came out to be 42.58%. Moreover, the accuracy for the S3D model for the top 1% came out to be 90.27%. Whereas the top 1% accuracy for MS-G3D came out to be 95.38%

TABLE I. PERFORMANCE OF DIFFERENT ARCHITECTURES IN [3]

| Model | Stream | Top1(%) | Top5(%) |
|---|---|---|---|
| Baseline | RGB | 42.58 | - |
| S3D | RGB | 90.27 | 97.98 |
| MS-G3D | Joints | 95.38 | 99.37 |
| | Bones | 94.50 | 99.07 |
| | Joint-motion | 92.92 | 99.16 |
| | Bone-motion | 90.22 | 98.60 |

Al-Hammadi, Muneer, Muhammad, Mansour et. al [4] studied 3DCNN which is being used to identify hand gestures. During the preprocessing stage, linear sampling was used to normalise the temporal dimension of hand motion data. To normalise the spatial dimensions, the length of the recognised face and human body part ratios were used. Then, in two techniques, the authors employed 3DCNN for feature learning. In the first technique, a single 3DCNN instance was trained to extract hand motion features from the whole video. In the second approach, three instances of the 3DCNN structure were trained to extract hand motion features from the beginning, middle, and end of the video clip. Before being supplied to the classifier, these region-based characteristics were merged. MLP, LSTM, and an autoencoder were used for feature fusion. For categorization, the authors employed a SoftMax active layer in both techniques. Various datasets were used to assess the suggested methods. The three datasets performed excellently in both signer-dependent and signer-independent modes. Six different state-of-the-art methodologies from the literature were compared to the proposed approaches. They excelled in four of these strategies while being on par with the other two. On the KSU-SSL Dataset, an accuracy of 96.69 percent was attained in the signer dependent mode. On the KSU-SSL dataset, an accuracy of 72.35 percent was reached in signer independent mode.

Bantupalli and Xie [11] have proposed a model for extracting spatial characteristics from the video stream using a CNN (Inception) for Sign Language Recognition. After that, temporal information from video sequences are retrieved using an LSTM (Long Short-Term Memory).

TABLE II. PERFORMANCE OF [11] WITH DIFFERENT SAMPLE SIZES

| # of Signs | Accuracy with Soft Layer | Accuracy with Pool Layer |
|---|---|---|
| 10 | 90% | 55% |
| 50 | 92% | 58% |
| 100 | 93% | 58% |
| 150 | 91% | 55% |

Ankit Ojha, Ayush Pandey et. al [9], have proposed a SOTA sign language to text and speech conversion method using convolutional neural networks. The images are generated from OpenCV video streams with the images processed as grayscale having dimensions of 50x50. A 95 percent accurate fingerspelling sign language translator for American Sign Language has been obtained.

Gore, Sayali, Namrata Salvi, and Swati Singh [10], developed a method for transforming sign language to text based on CNN. This algorithm is particularly useful for improving recognition accuracy when faced with difficult conditions such as scale, rotation, and translation. The dataset used is quite large, which improves the result's accuracy.

Pushpalatha, M. N., A. Parkavi, R. S. Sachin, Amy S. Vadakkan, Aakifha Khateeb, and K. P. Deeksha. [12], developed a model to classify some of the most important words used in American Sign Language with a 92 percent accuracy. Our approach allowed people with speech impairments to communicate with others without difficulty. Because present technologies are both expensive and inconvenient to use, the proposed solution employs posenet and transfer learning to allow users to freely express their opinions.

Yogya Tewari, Payal Soni, Shubham Singh, Murali Saketh Turlapati and Avani Bhuva [14] Using image processing and deep learning techniques, designed a new system for physically challenged people to recognise American Sign Language hand gestures. The Canny edge detection technique is used to retrieve the hand's edges. To speed up the procedure, the image is shrunk and processed. The Convolutional Neural Networks (CNNs) were trained to



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:24:18 UTC from IEEE Xplore. Restrictions apply.

predict gestures with 99 percent accuracy. The proposed mechanism's performance changes depending on the input image to the CNN for training.

Kausar, Sumaira Javed, Muhammad PY [15], have tried to investigate the shortcomings and challenges while carrying out sign language recognition. The authors understood that the following problems act like a boulder while building a sign language recognition system: variety in signs and gestures, feature selection and after that, extraction, segmentation and invariance. After carefully studying about the above-mentioned problems, the authors have laid out further guidelines for scientists who are working in the subject of identification of sign languages. The authors concluded that rather than using an external aid for segmentation, it should be done on the basis of vision. The authors also concluded that techniques for categorization and segmentation that do not impose significant limits on the signer's surroundings should be used. For example, if segmentation is based on both geometrical and skin colour aspects, limits on backdrop colours and the signer's clothing can be reduced.

## III. METHODOLOGY

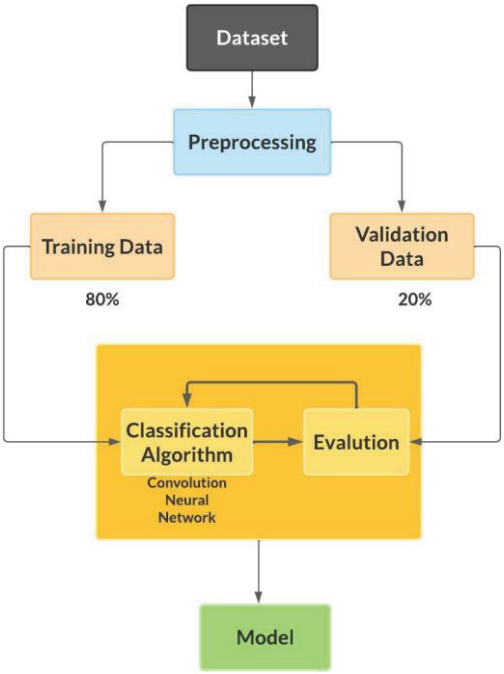



Fig. 1 Work Flow

The data was preprocessed by first converting the coloured images to greyscale images. Apart from that, gaussian blur was used to reduce image noise. The preprocessing steps also included binarization of images and the images were resized to a size of 128 x 128 from a size of 400 x 400. This data was divided into training data and validation have a split of 80 : 20 within the data. Once the preprocessing steps were completed, the data was passed into the model.

Alogirthm:

1. **Pre-processing** of Image data
   - **a.** Convert the image from rgb to gray scale
   - **b.** Applied Gaussian Blur to smooth the image
   - **c.** Segmentation of image using adaptive thresholding
   - **d.** Convert image to binary image for better retrieval of information
   - **e.** Resize the image to (128 X 128).
2. divide the dataset in training and validation data with a split of **80:20**.
3. **Input:** $\{X_1 \dots \dots X_m\}, \{I_1 \dots \dots I_m\}$
4. divide input into batches and forward it for training
5. **for** k = 1: K **do**
6. train CNN $f(X; \theta_k)$
7. **end**
8. **for** m = 1: M **do**
9. **for** k = 1: K **do**
10. obtain $P_m^k = f(X_m; \theta_k)$
11. **end**
12. **end**
13. **Output:** $\{\{P_1^1 \dots . P_1^K\} \dots \{P_M^1 \dots . P_M^K\}\}$

## IV. DATASET

There are a large number of choices for the language selected for the conversion of sign language into text. However, English being one of the most widely spoken languages was chosen in the scope of this project. In terms of English, there are a number of signs and gestures being used in different regions of the world. This includes American, British and the Indian Sign Language among many others. In our project, the American-Sign-Language is chosen as the set of choices as large datasets are easily available.

American-Sign-Language (ASL) can be considered as a natural language with linguistic characteristics similar to spoken languages and a grammar slightly distinct from English. Hand gestures and face movements are used to express ASL. ASL was not created by a single individual or any known committee. Although the original origins of ASL are unspecified, some speculate that it evolved more than 200 years ago through the mingling of local sign languages with French-Sign Language. The current ASL incorporates parts of French-Sign_Language as well as the original local sign languages, which have evolved into a rich, compound, and mature language over time.

To create the data for the project, a number of datasets have been used that are publicly available. Most of the datasets available for American Sign Language contain all of the alphabets separated into different directories. We have extended the dataset and have also added images for all the digits from 0 to 9 to create a combination of alphabets and numbers. This has helped us develop data that provides more utility and is more inclusive. A subset of [14] has been used as a source for the images of all the alphabets and the images of all the digits have been added to create the final dataset for the project. The data contains 2515 images which have been divided into a ratio of 80:20 to create the training set and validations set.

The images given below show the initial dataset which is in coloured format. This is followed by images on which the model has been trained.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:24:18 UTC from IEEE Xplore. Restrictions apply.

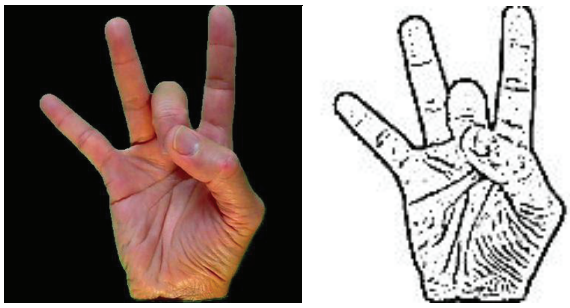


Fig. 2 Initial and Preprocessed Image

## V. Data Preprocessing

Preprocessing data means bringing the data into a predictable and analyzable form for further processing of data. In our project, information is in the form of images. In image-data preprocessing, some fundamental analysis and transformation are performed over a predetermined image. This preprocessed image is more useful for performing some other, more meaningful analytical task afterwards.

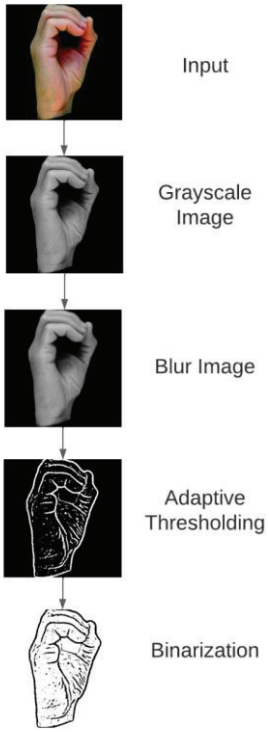



Fig. 3 Preprocessing Framework

The image preprocessing framework used in the preprocessing of the image is some simple image processing techniques. These techniques are:

**Converting an image into a grayscale image:** A grayscale image (also known as a grey level image) is one in which the only colours used are grey hues. When compared to other colour photographs, such images require less information to be delivered for each pixel. In RGB space, a 'grey' colour is one in which the red, green, and blue components are all equal in intensity. In contrast to the three distinct intensities necessary to identify each pixel in a full-color image, it becomes imperative to designate a single intensity value for each pixel.

**Use of Gaussian Blur:** A Gaussian blur is the effect of blurring a picture with a Gaussian function (also known as "Gaussian smoothing"). It is frequently used to reduce image noise and detail.

**Applying Adaptive Thresholding:** Adaptive thresholding is a technique for segmenting an image by giving all pixels with intensity levels greater than a threshold a foreground value and the remainder a background value. To distinguish desired foreground visual items from the background, the difference in pixel intensities of each region is employed.

**Binarization:** The process of transforming a grayscale (multi-tone) image to a black-and-white (two-tone) image is known as binarization. Determine the grey scale threshold value and whether or not a pixel has a certain grey value to begin the binarization process. The pixels are changed to white if their grey value exceeds the threshold. Similarly, if the grey value of the pixels falls below the threshold, they are transformed to black.

**Resizing:** The scaling of a picture is referred to as resizing. You'll need to resize a picture if you want to adjust the overall number of pixels in it.

The available images are of size 400x400. First, noise is reduced from each image using gaussian blur then the segmentation is done using adaptive thresholding to get the desired part of the image. This is followed by binarization which converts each image into black and white. At last the image size is reduced to 128x128. The specific dimension was chosen because it prevents too much information from being lost and helps to shorten the training period.

## VI. Training the Model

The model takes input as an image of shape 128x128x1. It consists of two convolution layers along with max-pooling for dimensionality reduction. The two layers resulted in a total of 9568 trainable parameters. Except for the last layer, each layer is followed by a batch normalization layer and a ReLU activation layer. The output layer uses the SoftMax function. The total number of trainable parameters generated by the model is 3,557,654. While creating our model, the pooling layer that we have used is the Max-Pooling layer with a pool size of 2x2 which means that the highest value inside a 2x2 array is kept.

In every of the layers, we have leveraged the ReLu (Rectified Linear Unit).For every input pixel, max(x,0) is calculated by ReLu . This provides the formula with some additional nonlinearity and makes it easier to learn more intricate features. It also helps in the removal of the vanishing gradient problem as well as decreasing the time taken in the calculations during training.

We have also used the softmax activation function in the last layer of the network which is a dense layer. In neural network models that predict a multinomial probability distribution, we have to use softmax as the activation in the output layer. Softmax is therefore utilised as the activation function in classification problems having more than 2 target classes.

To avoid the overfitting of our model we utilized callbacks. A callback is a type of object that can be used to perform actions at different points of the training



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:24:18 UTC from IEEE Xplore. Restrictions apply.

process. The number of epochs can be monitored with callbacks. We set up the 'val accuracy' to be checked with a patience of 10 epochs and a min delta value of 0% using a custom Early Stopping Callback. The model created has been trained for 100 epochs on the dataset that we generated. **The model achieved an accuracy of 84%.**

```
Model: "sequential_3"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_7 (Conv2D)           (None, 126, 126, 32)      320

 max_pooling2d_7 (MaxPooling  (None, 63, 63, 32)       0
 2D)

 conv2d_8 (Conv2D)           (None, 61, 61, 32)        9248

 max_pooling2d_8 (MaxPooling  (None, 30, 30, 32)       0
 2D)

 flatten_3 (Flatten)         (None, 28800)             0

 dense_12 (Dense)            (None, 128)               3686528

 dropout_6 (Dropout)         (None, 128)               0

 dense_13 (Dense)            (None, 96)                12384

 dropout_7 (Dropout)         (None, 96)                0

 dense_14 (Dense)            (None, 64)                6208

 dense_15 (Dense)            (None, 36)                2340

=================================================================
Total params: 3,717,028
Trainable params: 3,717,028
Non-trainable params: 0
```

Fig. 4 Model Summary

## VII. Result And Analysis

The DL model that we created was trained on the dataset we have generated. Binary cross entropy loss function was used to calculate the loss while training the model. The loss function is supposed to be decreased with consecutive epochs. The loss is calculated using the following equation:

$$Loss = -(y^i \log(y) + (1 - y^i) \log(1 - y))$$

where, $y^i$ is $i^{th}$ vector in output, y is the predicted probability.

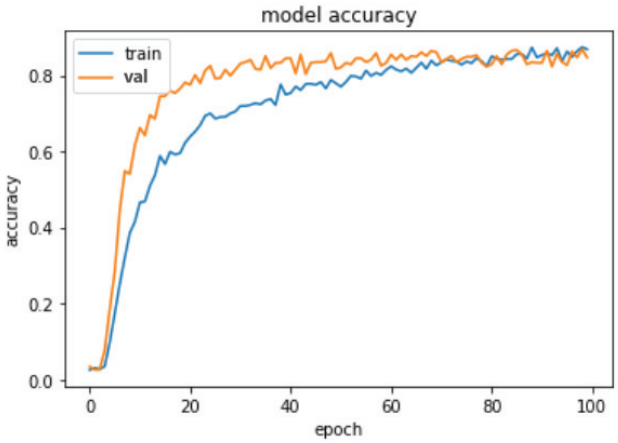



Fig. 5 Accuracy Curve

Fig.5 shows the accuracy per epoch for the phase of validation and training for the model. It can be inferred from the accuracy plot that with consecutive epochs, model accuracy tends to increase, but in some cases, unstable performance was noticed while comparing validation and training results.

Fig.6 shows the loss per epoch for the training phase and testing phase for the model. The plots show us that with consecutive epochs, the loss rate decreases, but in some cases, unstable performance was noticed while comparing testing and training results. The validation loss started out at 3.58% and decreased drastically with the increasing epochs to just 0.39%. The training loss also followed a similar trajectory. It started out at 3.60% and kept on decreasing until 0.40%.

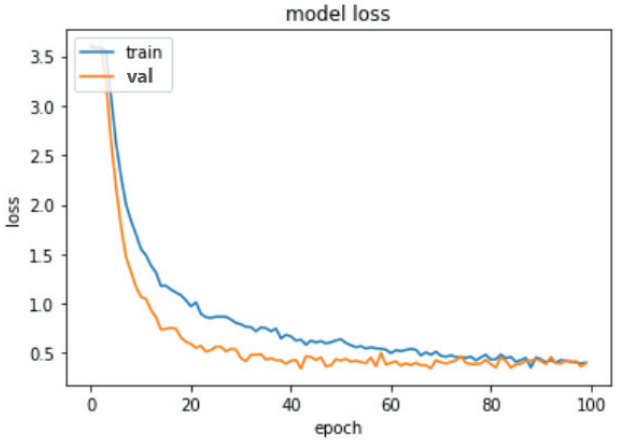



Fig. 6 Loss Curve

The following is the Classification report for our deep learning model. In terms of precision, most of the classes had a precision score above 85%. Through the results, we can infer those classes depicting {8:'8', 9:'9', 15:'F', 17:'H'} were detected with very high precision and F - 1 scores.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.50 | 0.77 | 0.61 | 70 |
| 1 | 0.69 | 1.00 | 0.82 | 70 |
| 2 | 0.77 | 0.84 | 0.80 | 70 |
| 3 | 0.99 | 1.00 | 0.99 | 70 |
| 4 | 0.89 | 0.81 | 0.85 | 70 |
| 5 | 0.83 | 1.00 | 0.91 | 70 |
| 6 | 0.87 | 0.19 | 0.31 | 70 |
| 7 | 0.94 | 0.94 | 0.94 | 70 |
| 8 | 1.00 | 0.84 | 0.91 | 70 |
| 9 | 1.00 | 0.93 | 0.96 | 70 |
| 10 | 0.80 | 0.87 | 0.84 | 70 |
| 11 | 0.98 | 0.80 | 0.88 | 70 |
| 12 | 0.96 | 0.91 | 0.93 | 70 |
| 13 | 0.89 | 0.94 | 0.92 | 70 |
| 14 | 0.94 | 0.84 | 0.89 | 70 |
| 15 | 1.00 | 0.89 | 0.94 | 70 |
| 16 | 0.93 | 0.94 | 0.94 | 70 |
| 17 | 1.00 | 0.93 | 0.96 | 70 |
| 18 | 0.98 | 0.91 | 0.95 | 70 |
| 19 | 0.99 | 0.94 | 0.96 | 70 |
| 20 | 0.75 | 0.84 | 0.79 | 70 |
| 21 | 0.96 | 1.00 | 0.98 | 70 |
| 22 | 0.68 | 0.94 | 0.79 | 70 |
| 23 | 0.90 | 0.53 | 0.67 | 70 |
| 24 | 0.49 | 0.37 | 0.42 | 70 |
| 25 | 0.97 | 0.99 | 0.98 | 70 |
| 26 | 0.93 | 1.00 | 0.97 | 70 |
| 27 | 0.76 | 0.90 | 0.82 | 70 |
| 28 | 0.87 | 0.89 | 0.88 | 70 |
| 29 | 0.81 | 0.89 | 0.85 | 65 |
| 30 | 0.84 | 0.37 | 0.51 | 70 |
| 31 | 0.87 | 0.64 | 0.74 | 70 |
| 32 | 0.51 | 0.99 | 0.67 | 70 |
| 33 | 0.97 | 0.84 | 0.90 | 70 |
| 34 | 0.93 | 1.00 | 0.97 | 70 |
| 35 | 0.97 | 0.90 | 0.93 | 70 |
| accuracy | | | 0.84 | 2515 |
| macro avg | 0.87 | 0.84 | 0.84 | 2515 |
| weighted avg | 0.87 | 0.84 | 0.84 | 2515 |

Fig. 7 Classification Report



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:24:18 UTC from IEEE Xplore. Restrictions apply.

The relevance of precision in our study is to identify the quality of positive predictions. The average precision of the model is 0.87 which depicts that the model was good at predicting the actual class of a symbol out of the classes predicted for a symbol. The high precision scores of the classes indicates that the number of true positives were very high for the respective classes. The classes having higher precisions and recall also have higher F1 Score since the f1 score is the harmonic mean of precision and recall.

However, the classes belonging to {0:0, 1:1, 2:2, 24: O, 27: R} have low precision scores. The possible reason for the same accounts to the fact that the size of data per class is relatively less as of now, which leads to the model being unable to learn how to classify those classes correctly.

## VIII. Conclusion

Sign language is one of the most useful tools for facilitating communication between deaf and mute people and the rest of society. Though sign language can be used to communicate, the target person must have a basic understanding of the language, which is not always possible. As a result, our project lowers these barriers. This project was created as a proof-of-concept to see if it was possible to recognise sign language. This project allows ordinary people to communicate with deaf or dumb people using sign language, with the text being converted to images.

In a number of scenarios, a sign language interpreter becomes an object of high utility. In educational institutions, hospitals, airports, and courts, anybody may use this technology to comprehend and communicate in sign language. It allows persons with normal hearing to communicate with others who have trouble hearing.

In this project, we have developed a model that converts sign language to text using deep learning. The algorithm used to develop the model is Convolution Neural Network. The dataset used is a combination of multiple datasets. The model created by us has been trained for 100 epochs on the dataset that we generated. **The model achieved an accuracy of 84%.** The model is performing quite well with the amount of data we are using. By increasing the dataset size, we can improve the accuracy.

## IX. Future Work

Sign language is part and parcel in the lives of deaf and mute people.With over 5% of the world's population i.e more than 70 million people are deaf and dumb and have to face challenges at each and every stage of life. Hence, it becomes imperative to make advancements in the arena of sign language and gesture identification so that it becomes easier for people who cannot hear and speak. In this paper, we have successfully designed a model which converts sign language to text for not only the English alphabet but also the numerical numbers from 0 to 9.

In the study that we have conducted, we have trained a DL model from scratch to build a sign language and gesture recognition system. With the recent advancements in pre-trained models in deep learning, we aim to use transfer learning on different pre-trained classification models to build our sign language recognition system which will help us increase the accuracy of the model. Some of the pre-trained models that aim to use are VGG 16, Resnet 50, Inception v3, EfficientNet etc.

The model that we have trained is pretty lightweight and can be exported. Hence in the future, we also aim to deploy our model in the form of a web app or an android app with a simple and interactive GUI which will make it easier for the deaf and dumb people to use it and interact with each other in a simple and hassle free way.

## Acknowledgment

We would like to express our deepest appreciation for Dr. Aruna Bhat for providing us with the opportunity to write a research paper on the topic "Sign Language to Text Conversion using Deep Learning". We would like to thank her for her continued support and guiding us in every stage of this research paper.

## References

[1] Izutov, Evgeny. "ASL Recognition with Metric-Learning based Lightweight Network." arXiv preprint arXiv:2004.05054 (2020).

[2] Pigou, Lionel, Mieke Van Herreweghe, and Joni Dambre. "Gesture and sign language recognition with temporal residual networks." In Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 3086-3093. 2017.

[3] Vazquez-Enriquez, Manuel, Jose L. Alba-Castro, Laura Docio-Fernandez, and Eduardo Rodriguez-Banga. "Isolated Sign Language Recognition With Multi-Scale Spatial-Temporal Graph Convolutional Networks." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3462-3471. 2021.

[4] Al-Hammadi, Muneer, Ghulam Muhammad, Wadood Abdul, Mansour Alsulaiman, Mohamed A. Bencherif, and Mohamed Amine Mekhtiche. "Hand gesture recognition for sign language using 3DCNN." IEEE Access 8 (2020): 79491-79509.

[5] Truong, Vi NT, Chuan-Kai Yang, and Quoc-Viet Tran. "A translator for American sign language to text and speech." In 2016 IEEE 5th Global Conference on Consumer Electronics, pp. 1-2. IEEE, 2016.

[6] Albawi, Saad, Tareq Abed Mohammed, and Saad Al-Zawi. "Understanding of a convolutional neural network." In 2017 International Conference on Engineering and Technology (ICET), pp. 1-6. Ieee, 2017.

[7] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521, no. 7553 (2015): 436-444.

[8] Kartik, P. V. S. M. S., Konjeti BVNS Sumanth, VNV Sri Ram, and P. Prakash. "Sign Language to Text Conversion Using Deep Learning." In Inventive Communication and Computational Technologies, pp. 219-227. Springer, Singapore, 2021.

[9] Ankit Ojha, Ayush Pandey, Shubham Maurya, Abhishek Thakur, Dr. Dayananda P, 2020, Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCAIT – 2020 (Volume 8 – Issue 15)

[10] Gore, Sayali, Namrata Salvi, and Swati Singh. "Conversion of Sign Language into Text Using Machine Learning Technique." International Journal of Research in Engineering, Science and Management 4, no. 5 (2021): 126-128.

[11] Bantupalli, Kshitij, and Ying Xie. "American sign language recognition using deep learning and computer vision." In 2018 IEEE International Conference on Big Data (Big Data), pp. 4896-4899. IEEE, 2018.

[12] Pushpalatha, M. N., A. Parkavi, R. S. Sachin, Amy S. Vadakkan, Aakifha Khateeb, and K. P. Deeksha. "Sign Language Converter Using Feature Extractor and PoseNet." Webology 19, no. 1 (2022).

[13] Joy, J., Balakrishnan, K. & Madhavankutty, S. SignText: a web-based tool for providing accessible text book contents for Deaf learners. Univ Access Inf Soc (2021). https://doi.org/10.1007/s10209-021-00801-7

[14] Yogya Tewari, Payal Soni, Shubham Singh, Murali Saketh Turlapati and Avani Bhuva Conference: 2021 International Conference on



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:24:18 UTC from IEEE Xplore. Restrictions apply.

Communication information and Computing Technology (ICCICT), Year: 2021, Page 1

DOI: 10.1109/ICCICT50803.2021.9510094

[15] TY - BOOK AU - Kausar, Sumaira AU - Javed, Muhammad PY - 2011/12/01 SP - 95 EP - 98 T1 - A Survey on Sign Language Recognition DO - 10.1109/FIT.2011.25 JO - Proceedings - 2011 9th International Conference on Frontiers of Information Technology, FIT 2011 ER



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2022 at 04:24:18 UTC from IEEE Xplore. Restrictions apply.

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/0025-1747.htm



# Stimulating CSR learning collaboration by the mentor universities with digital tools and technologies – an empirical study during the COVID-19 pandemic

Received 31 December 2021
Revised 1 May 2022
12 June 2022
Accepted 5 July 2022

Namita Jain
*Shri Ram College of Commerce, University of Delhi, New Delhi, India*
Asha Thomas
*Department of Operations Research and Business Intelligence, Wroclaw University of Science and Technology, Wrocław, Poland*
Vikas Gupta
*Delhi School of Management, Delhi Technological University, Delhi, India*
Mario Ossorio
*Department of Economics, University of Campania Luigi Vanvitelli, Capua, Italy, and*
Daniele Porcheddu
*Department of Economics and Business Studies, University of Sassari, Sassari, Italy*

## Abstract

**Purpose** – The research aims to measure the effectiveness of collaborative learning exchanges transpired through digital tools and technologies (DT&Ts) employed by the mentor universities during the COVID-19 pandemic by conducting an empirical study on undergraduate students in Indian higher educational institutions (HEIs) under the mentorship program based on the corporate social responsibility (CSR) initiative. The pandemic scenario, its impact on the mentor university's social responsibility and the way DT&Ts can assist are investigated in this article.

**Design/methodology/approach** – The interactions with experts and students were conducted to explore the DT&Ts for learning exchanges. Next, structural equation modeling (SEM) was performed to validate the model and perform regression analysis. The quantitative data collection was made through questionnaires during the second deadly wave of COVID-19 that hit India.

**Findings** – The independent variables (IVs) such as the IT infrastructure support (IT_IS), virtual collaborative tools (VCTs) and future-oriented technologies (FOTs) have a significant impact on the CSR learning outcomes (CSR_LOs) of undergraduate students under the mentorship program. However, IV research instruments for innovation could not make a significant effect.

**Research limitations/implications** – The IVs IT_IS, VCTs and FOTs influence the CSR_LOs, while RII does not have an influential impact.

**Practical implications** – As the online learning environment is expected to stay at least in a blended form, adequate CSR funding in infrastructure is necessitated to harness the full potential of this important resource, technology. The results of this empirical investigation affirm that IT_IS, VOTs and FOTs significantly impact CSR_LOs during the crisis. The study findings would encourage the mendtor universities and their stakeholders, including the mentee universities, to evolve and create an ecosystem for effective management of these resources to attain positive outcomes. The study findings can guide the mentor universities in managing uncertainties like pandemics and effectively using the earlier-mentioned critical resources for social responsibility. This research also allows the development of future applications adnd models in mentor-mentee universities for social responsibility, post-pandemic transformation and resilience.

**Social implications** – The DT&Ts came to the immediate rescue during the pandemic and positively affected collaborative CSR_LOs by the mentor universities, but they have not evolved to a level where offline learning can be replaced entirely. Hence, it can be inferred that a hybrid model is preferable. The study also improves the understanding of how DT&Ts are being harnessed to aid collaborative learning in fulfilling the

Management Decision
© Emerald Publishing Limited
0025-1747
DOI 10.1108/MD-12-2021-1679

mentors' CSR in fatal emergencies. The purpose is to equip the education system through mentorship so that universities can sustain, innovate and grow even in trying times. Also, it discusses the dynamics of various DT&Ts for creating a sustainable learning environment and utilizing them to make the teaching prolific and influential.

**Originality/value** – There is a scarcity of literature regarding the learning outcomes realized through CSR initiatives and collaboration between mentor-mentee institutions. There is a need to understand how these knowledge exchanges continued despite the physical restrictions during the pandemic. In this direction, this study helps to understand how the DT&Ts played a critical role in continuing learning and keeping abreast in a knowledge society from the perspective of resource-based view (RBV) in these precarious situations.

**Keywords** Digital tools, Education, Student learning, Corporate social responsibility (CSR), Covid-19, Knowledge exchange, Resource-based view

**Paper type** Research paper

## 1. Introduction

Universities have a greater economic and social development responsibility, in addition to teaching and research, to strengthen stakeholder engagement (Drucker and Goldstein, 2007; Thomas and Paul, 2019). Through education and community service, universities have consistently engaged in corporate social responsibility (CSR) (Dima *et al.*, 2013; Irby, 2015; Gangi *et al.*, 2018; Rahman *et al.*, 2019). Also, for a productive system and the development of the knowledge society, higher educational institutions (HEIs) and universities must provide excellent education and innovation in key techniques, volumes and forms (Ranga and Etzkowitz, 2015; Thomas, 2021). Mentoring is an essential representation of a CSR initiative (Brodeur, 2013). Mentoring as social responsibility is defined as the mentor investing in and acting responsibly toward the mentee to benefit both the mentee and the broader good of humanity. A mentee is "anyone who wants to learn and seeks valuable advice from someone who knows to grow professionally and personally." (UGC, 2021). Like CSR, mentorship should be led or guided by values (Irby, 2015). To be a good mentor institution, one must respect the mentee, have a deep understanding and respect for all humans, know global human needs, concern for all stakeholders, be accountable for actions and transparency with the mentee (Irby, 2015).

The mentor-mentee initiative's CSR is to teach, engage, motivate and enable mentee university students to collaborate on innovative ideas in the coming years. The Government of India's Ministry of Education's Innovation Cell (MIC) has selected institutes across India as "mentor institutions" under the mentor-mentee program. The mentorship program is required to improve the educational quality of the mentee institution. To develop an excellent education and innovation ecosystem, high-performing institutions will act as mentors to institutions that need direction and support. Mentors assist mentees in becoming more self-sufficient, sustainable and productive members of the next generation (Hastings and Sunderman, 2020). Mentors actively participate in research, course development, internships and decision-making. They also provide resources to assist mentee institutions in growing. During the pandemic, it was difficult to supervise and guide these students. Teachers had to do their best to accommodate all the students while working within limited resources in these situations. India's situation is unique; in that, it was one of the few countries that were severely impacted by the coronavirus. The high student-teacher ratio made social distancing more difficult than in low-population-density countries. Keeping these concerns in mind, schools and colleges in the affected districts were expected to remain closed. Examinations were frequently rescheduled, leaving students stranded.

In the face of accelerating turbulence and change, mentors need to find new ways of thinking to sustain performance and growth (Sinha *et al.*, 2012, 2015; Thomas and Gupta, 2022b). In these unusual circumstances, switching to a different mode of instruction was critical (Rajhans *et al.*, 2020). As a result, all on-site activities were ceased and replaced with remote internet-based methods. For example, in-person training has evolved into

videoconferencing, seminars into webinars, peer discussion into blogs and hands-on training into a simulation. Physical equipment was replaced by digital technology and communities of practice transitioned from physical gatherings to virtual teams. The pivotal point of this shift in technology has aided in preserving a knowledge society under such conditions. One of the most effective ways to communicate with students at mentee institutions, understand their challenges and assist them as needed is to use digital tools and technologies (DT&Ts) to steer sessions.

Despite the widely acknowledged potential of technology in learning, there has been little research on how DT&Ts aid collaborative learning in meeting CSR (Chigona, 2015; Wang *et al.*, 2020). The authors intend to examine the effectiveness of DT&Ts in this COVID-19 emergency in five mentee universities in India. Using India as a case study allows us to better understand how students at these mentee universities in the country learn. With over 20 million students as stakeholders, India's higher education system is the world's second-largest (Study in India, 2021). India, like other developing countries, has a digital divide. On the one hand, India has technologically empowered mentor institutions with the availability of resources, and on the other, they have weaker institutions with vulnerable technological capabilities. They struggle with infrastructure, research labs, admissions and are forced to close. This schism became more pronounced during the COVID-19 pandemic when all physical activities shifted to the digital realm (Barbagila *et al.*, 2021). In such a difficult situation, mentor universities assisted mentee universities in navigating, overcoming and learning during the COVID-19 global crisis. The goal is to assess how DT&Ts aid in the digitalization of exchanges, impacting the learning outcomes of the mentee institutions.

This study both theoretically and empirically contributes to the existing literature on RBV and CSR. To the best of the authors' knowledge, this is the first study that examines the CSR_LOs through DT&Ts in the context of mentor-mentee institutions during the pandemic. Thus, the main objective of this paper is to assess the impact of DT&Ts on the CSR_LOs of students at mentee universities and attempt to fill the research gap. Thus, this paper will clarify the learning outcome based on which future proactive initiatives could be designed and contribute to social and practical implications.

The remainder of this work is organized in the following manner. Section 2 presents the literature review, followed by section 3 on formulating the study's constructs and hypotheses. Section 4 outlines the methods for conducting the research, data preparation and statistical analysis. Section 5 closes with concluding findings. Finally, section 6 concludes with implications and future research avenues.

## 2. Literature review

The complex business environment and demanding work profiles require strong and effective mentoring of students from experts in their area. There has been a rise in the use of DT&Ts and their impact has been measured in literature. However, there is still a scarcity of literature on the learning outcome under the CSR mentorship program during COVID-19. This pandemic is a recent phenomenon wherein DT&Ts are utilized as the sole medium of learning because physical classes were unthinkable in such a scenario. A study by García-Morales and Rojas (2021) revealed that the usage of virtual platforms increased extensively in a post-pandemic teaching-learning environment. Impey and Formanek (2021), during the lockdown period of COVID-19, noticed a surge in massive open online courses (MOOCs) enrollments. Ghada Refaat El Said (2021) compared the exam results of 372 students enrolled in face-to-face versus online business education programs in Egypt and determined that there is no significant difference in students' performance. It was proved that online education can be effective as well. Zou *et al.* (2020) expressed that though online learning has become the new normal, still it is not able to achieve the efficiency equivalent to that of physical face-to-

face classes. Sharma and Alvi (2021) discovered that students' learning experience in pre-pandemic blended learning is better than in post-pandemic online learning. Campillo-Ferrer and Miralles-Martínez (2021) suggested that online learning can also be made effective, exciting and engaging using gamification and other active learning methodologies as a part of flipped classroom model. In this direction, using new technologies like Qbots powered by artificial intelligence (AI) and augmented reality can make problem-solving easier and comprehension faster in the future (Curtin, 2021).

The Department of Education and Training (2021) defined DT&Ts as the electronic devices, software, applications and systems that store, share, present and process information in various forms for its users. Table 1 lists some of the key DT&Ts with their elements and source as mentioned in the literature.

From Table 1, it can be inferred that though research has been conducted to study DT&Ts, they were not done to empirically test their contribution during lockdown when the learning was solely dependent on them. A survey is also necessitated to determine their usage and familiarity among the students.

The paper's objective is to determine the significant DT&Ts for student learning and knowledge exchanges under CSR initiatives during the COVID-19 pandemic. For this purpose, an empirical study was conducted and data were collected through a structured

| DT&Ts | Elements | Source |
|---|---|---|
| IT infrastructure support | Hardware like laptop/PC, printer and scanner<br>System and application software<br>Efficiency of the Internet connectivity | Duncan (1995), Walmiki and Ramakrishnegowda (2009), Byrd and Turner (2001), Rockart *et al.* (1996), Sharma and Singh (2010), OECD (2010b), IFC (2021) |
| Virtual collaborative tools | Classroom communities like WhatsApp group, Google group<br>Online tutoring and webinars with MS Teams, Google Meet, Zoom<br>Peer editing tools<br>MOOCs/video tutorials | Prensky (2001), Amin *et al.* (2018), Brady *et al.* (2010), Bates (2014), Kumar *et al.* (2021), Zou *et al.* (2020), Radu *et al.* (2020), Naresh (2020), Kaup *et al.* (2020), Thomas (2011), Correia *et al.* (2020) |
| Future-oriented technologies | Recent technologies like<br>Virtual Reality<br>Augmented Reality<br>Artificial intelligence<br>Gamification<br>Flipped classroom | Fisk (2017), Gupta and Jain (2017), Ahmad (2015), Brandeis University (2015), MIT (2016), Gopalan (2021), Teräs *et al.* (2020), Kupe and Ouma (2021), Nieto-Escamez and Roldán-Tapia (2021) |
| Research instruments for innovation | Plagiarism software<br>Data analysis software<br>Knowledge repositories/ databases<br>E-journals | Evering and Moorman (2012), Chandere *et al.* (2021), Sukula *et al.* (2021), Apostol (2020), Rathee and Sarkar (2020), Bogdandy *et al.* (2020), Jameel (2020), Mishra *et al.* (2020) |
| CSR learning outcomes | Ability to grasp the concepts taught online<br>Ability to apply the concepts taught online<br>Ability to solve the problems accurately | Fry *et al.* (2008), Bertling *et al.* (2020), Caspersen *et al.* (2017) |

**Table 1.**
Review of DT&Ts

**Source(s):** Compiled by the authors

questionnaire. The paper attempted to answer which DT&Ts significantly affect learning exchanges in times of pandemic under the CSR initiatives. As the pandemic is a novel phenomenon wherein all educational pursuits went completely online, not many studies are available measuring the impact of DT&Ts and social responsibility on CSR_LOs. The study identifies and discusses the various DT&Ts and contributes to their usage and effective management in testing times.

## 3. Background and construct development

For over three decades, the RBV has served as one of the underpinnings of management research (Nason and Wiklund, 2018). Organization and management of corporate resources are crucial for developing a company's competitive advantage (Barney, 1991; Wernerfelt, 1989; Battisti *et al.*, 2022). RBV focuses on fostering skills, knowledge and corporate culture, which can provide the company with a sustainable competitive advantage (Vrontis *et al.*, 2021; Battisti *et al.*, 2022; Gaur and Gupta, 2021). This also increases a company's intangible assets and performance, creating firm-specific resources that are inimitable, exclusive and irreplaceable (Barney, 1991; Thomas, 2021).

Theoretically, the authors have proposed a conceptual model accentuating the resources capable of inducing a competitive and community advantage and improving the performance of the mentor university through CSR_LOs. This paper also contributes to the literature on stakeholder theory and CSR (Freeman, 1994). According to stakeholder theory, CSR strategy and its outcomes may mitigate risk due to a willingness to respond to CSR needs and expectations of stakeholders, resulting in increased stakeholder support and loyalty (Martnez and Rodrguez del Bosque, 2013; Nirino *et al.*, 2022). The ability to build stakeholder loyalty reduces risk, especially during difficult times for the company, while preventing shareholder overreactions (Helm and Tolsdorf, 2013; Waheed and Zhang, 2020). In this regard, the CSR initiatives implemented by the mentor universities enhanced the learning of the mentee institutions even during a pandemic. These virtual partnerships aided mentee institutions in enhancing their emergency management capabilities. It has compelled weaker institutions to alter their established procedures and become receptive to novel methods.

The conceptual model presented in the paper adopts and adapts its constructs IT infrastructure support (IT_IS), virtual collaborative tools (VCTs) and future-oriented technologies (FOTs) from technology, social interaction and pedagogy of the generic model and constructs RII from the extended awareness incentives demand and support (AIDS) model. The generic model of the education system can guide the integration of ICT into the teaching and learning process (Kirschner *et al.*, 2004; Wang, 2008). The model has its roots in constructivist learning theories, interactivity design and usefulness. Constructivist learning theories posit that an efficient pedagogical design combined with relevant content permits knowledge transmission and knowledge creation. Nowadays, the blend of futuristic technologies with pedagogy has given birth to novel pedagogical innovations enhancing the quality of education. The component "social interaction" suggests that learning is facilitated through the interactive process of knowledge sharing, collaboration and conversation between members (Thomas *et al.*, 2020; Gupta and Thomas, 2019). The parameter "technology" renders usability by providing an easy-to-use interface for supporting the learning endeavors of students. Realizing the importance of research for innovation and discovery in the educational landscape, the generic model is extended to the AIDS Model to aid teaching, learning and study for progression in the current field (Asabere *et al.*, 2017).

*3.1 Independent variables*

*3.1.1 IT infrastructure support (IT_IS).* The epidemic has compelled educators to increase their use of ICT and embrace online education as an essential component of daily teaching and instruction at the institution. While teaching is undeniably one of the university's primary responsibilities, the teacher's role is changing as online learning environments emerge (Laurillard, 2002; Lawrence and Tar, 2018). Core competencies necessitate identifying the pedagogical and technological skills required for effective teaching in web-based learning environments. Professional development and IT_IS for pedagogical, technological support workers and learning are critical in this situation. Any study on technology is incomplete without considering the availability of a robust and efficient IT infrastructure. This resource provides the foundation for the smooth operation of learning activities in times of lockdown and physical distancing. According to Duncan (1995), IT infrastructure comprises "hardware and operating system, network and telecommunication technologies, data and core software applications." In brief, the hardware, software and networks are pillars of IT infrastructure. Hardware comprises tangible devices like mobiles, laptops, PC, printers and scanners used for online learning. This hardware can only run with the help of system and application software. Robust and dependable Internet connectivity is mandatory to run apps and stay connected worldwide. Therefore, it is observed that adequate funding under the CSR initiative for developing this resource is critical to student learning (IFC, 2021; OECD, 2010b). Soong *et al.* (2001) and Alsabawy *et al.* (2013) investigated, observed and confirmed the critical factors in online courses and mentioned that IT_IS was a significant valid factor in the success of online courses and e-learning systems. The present research demonstrates the critical role of IT_IS services in the success of e-learning systems.

Thus, drawing from the literature, it is suggested that IT_IS can significantly impact the learning outcome (CSR_LO) of students under the mentorship program.

*H1.* IT_IS has a significant effect on CSR_LO in HEIs.

*3.1.2 Virtual collaborative tools (VCTs).* VCTs such as Google groups, WhatsApp groups, shared docs and online courses were already prevalent in a pre-pandemic environment but gained significant momentum during the pandemic. Accepting new technologies opens an excellent opportunity for virtual collaborative systems in HEIs. This latest form of exchange creates possibilities for enhancing students' learning (Alanis-Funes *et al.*, 2011).

The learning activities during COVID-19 would not have been possible had the video conferencing tools not come to the rescue. The popular virtual classroom tools are Zoom, MS Teams, Google Meet, Webex and Skype (Zou *et al.*, 2020; Radu *et al.*, 2020). These tools replicated the physical classes, allowing virtual face-to-face interaction through video, audio and screen sharing capabilities. These resources also incorporated additional features such as attendance download, assignment submission, polling, file sharing and hand-raising. Collaborative peer editing tools like Google Docs, Microsoft One Drive, SharePoint and Anydesk permit online editing and organized compilation of material on the Cloud without physical contact (Naresh, 2020; Jain, 2018a, b). Through this collaborative software, mentor-mentee members and many people can simultaneously work on the same file and share their ideas and views in a distributed learning space. Co-authoring is also possible.

Online groups or communities of practice are the online social networks in which "people with common interests, goals, or practices interact to share information and knowledge, and engage in social interactions" (Chiu *et al.*, 2006). It supports a more "personalized communicative form of the World Wide Web that emphasizes active participation, connectivity, collaboration, and sharing of knowledge and ideas among users." (McLoughlin and Lee, 2007). Using some of the online COPs like WhatsApp groups, Google groups, Facebook and YouTube communities, mentor and mentee members can create content, share knowledge, collaborate and conveniently communicate (Mishra *et al.*,

2020; Thomas and Gupta, 2022a; Gupta, 2020). Online collaborative tools are elemental for fostering student interaction and stimulating intra-group emotional support. These aspects directly correspond to meaningful collaborative learning (Hernández-Sellés *et al.*, 2019).

MOOC is an open-access course based on the distributed peer learning model. Its key features include short video-based lectures, automatic assessments and online forums for discussion with teachers, peers and MOOC providers (Gupta and Jain, 2017; George Glance *et al.*, 2013; Baturay, 2015). In these emergent situations, when mentees could not provide the much-needed exposure to experts in the area, many universities could not adjust to the online framework due to inadequate infrastructure and training, and MOOCs became a viable option. This is because MOOCs providers were already accustomed to teaching online and all other allied activities like assignment, evaluation, student progress and attendance were tracked online and guidance was available from them.

Studies have confirmed that well-structured collaboration promotes learning through VCTs found in socio-emotional presence, enhancing learning and individual performance (King, 2007; Medina and Suthers, 2008; Kwon *et al.*, 2014; Hernández-Sellés *et al.*, 2019). Based on the earlier discussion, the following hypothesis is proposed:

*H2.* VCTs have a significant effect on CSR_LO in HEIs.

*3.1.3 Future-oriented technologies (FOTs).* Our lives are overwhelmed by interesting technical inventions and innovations (Sinha *et al.*, 2009). The pandemic has fastened the speed of progression towards Education 4.0, where the blockades between man and technology would blur and the future trending technologies would become routine in education (Fisk, 2017; Jain and Gupta, 2019a, b). The prominent examples are virtual reality (VR), AI, augmented reality, gamification and flipped classrooms. These FOTs enhance the learning experience with their advanced features and capabilities.

In the case of VR, the user is transported to a digital world disconnected from the real world. The head-mounted displays create an immersive experience that enthralls the human senses. Using VR, the teachers can walk students to the height of mountains, the depth of oceans, unbounded galaxies and gigantic outer space which was not possible before (Bamford, 2012).

AI refers to the development of smart machines that mimic human acumen and thinking. AI-enabled devices are capable of reasoning, problem-solving and self-learning the way humans do (ColdFusion, 2016; Munfarid, 2018). Some examples of AI are auto-suggestion, pattern recognition, language translation, visual perception and decision-making. In education, AI-powered grading software can be used to evaluate a student's assignment. As part of AI, machine learning can be used to monitor progress and provide real-time feedback to students. Similarly, bots can answer basic questions and automate recurring mechanical chores (Singh, 2016).

Augmented reality overlays digital content or objects to the real-world view (Hajizada, 2019). In this form of experiential learning, students are no longer passive audiences but participate, perform tasks and learn real-world skills, which were impossible in a conventional classroom (Davies, 2017; Jain, 2018a, b). AR has applications in training, gaming and prototype generation. The much-needed experience and exposure to work on machines or their prototypes can be granted online to mentee institutions through DT&Ts.

Gamification refers to the use of fun learning activities like puzzles, quizzes, escape rooms and avatars to engage and seek maximum student participation in classroom activities. In the absence of physical interaction, they are said to boost virtual social communication through 3D interactions, simulations and AI. Various ed-tech gaming applications such as Kahoot, Classcraft, Classdojo and Curious World are used in an online environment (Nieto-Escamez and Roldan-Tapia, 2021).

The flipped classroom is a form of pedagogical innovation which inverts the traditional class (Shi-Chun *et al.*, 2014). Students watch lectures from experts, giving them ample time for activities and applying concepts and utilize this application-based learning later in classes-virtual or physical (Bechter and Swierczek, 2017). The pre-recorded videos offer the flexibility of pausing, rewinding, or fast-forwarding. However, they may fail to provide personalized attention to students, which can be addressed through in-class activities. Based on this discussion, the following hypothesis is proposed:

*H3.* FOTs have a significant effect on CSR_LO in HEIs.

*3.1.4 Research instruments for innovation (RII).* Research occupies a prominent position in the teaching-learning process. Knowledge exchange between mentor and mentee is expected to create a perennial flow of new knowledge for advancement in the field. If integrated successfully, the research tools can enrich learning and broaden knowledge paradigms for problem-solving, publications and successful completion of innovation projects in the HEIs. Knowledge repositories such as Prowess, Jstor, Shodhsindhu and e-journals are critical to imparting food for thought and research-based activities (Novak, 2017). Data analytics tools like SPSS, Stata, R, Amos and Eviews aid data crunching and analysis with their algorithms and advanced features. Similarly, while researching writing, plagiarism checks using tools like Urkund and Turnitin have become mandatory (Chandere *et al.*, 2021). The research instruments aid data collection, analysis and reporting to deepen our understanding of the subject. Based on this, the following hypothesis is proposed:

*H4.* RII has a significant effect on CSR_LO in HEIs.

### *3.2 Dependent variable*

*3.2.1 CSR learning outcomes (CSR_LOs).* The students' learning outcome through CSR in the mentee university is the dependent variable (DV) of the study. Learning can be created through greater exposure, opportunities and direction, which the mentor institution can provide to the mentee institution. The ultimate aim of any knowledge exchange is to better understand the world we live in. Learning can be measured by comprehending the problem and developing innovative solutions. A learned student is not only able to master the concepts but apply them wisely for sound decision-making (Fry *et al.*, 2008). Learning helps decode unstructured problems and derive meaning from data to reason and comprehend further. Learning is not just a transfer of knowledge but a reflection and thought systemization process for value creation (Bratianu and Vatamanescu, 2017; Singh and Gupta, 2021).

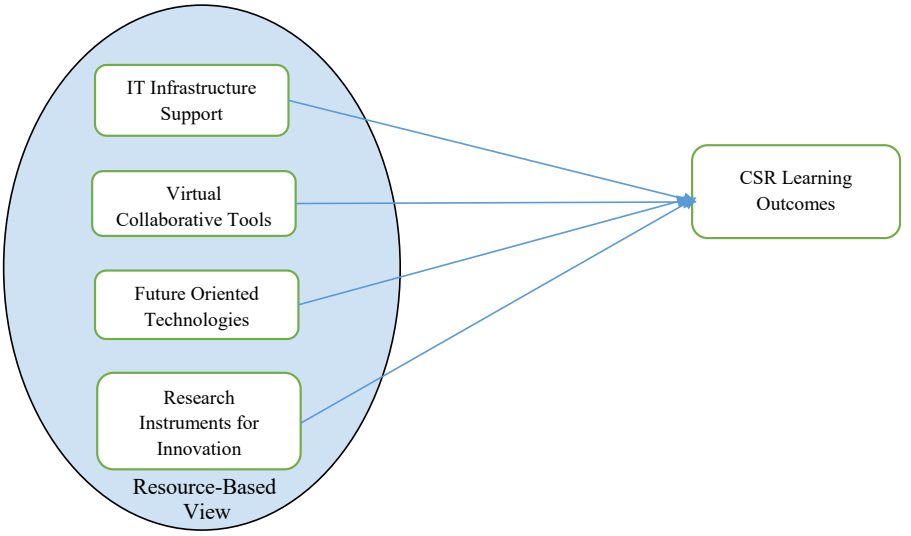
During the pandemic, offline teaching-learning suddenly pivoted to online learning. Though e-learning was already growing (Dominici and Palumbo, 2013a), it gained unthinkable thrust during the spread of COVID-19. Thus, high dependence on these digital technologies was observed. Therefore, research on the usage and adoption of DT&Ts is demanded, keeping in mind the growing learning needs of students.

### *3.3 The conceptual model*

Figure 1 is a diagrammatic illustration of the four independent variables (IVs), IT_IS, VCTs, FOTs and RII, impacting the DV CSR_LOs of undergraduate students under the mentorship program. On the basis of the above discussion, a conceptual model is proposed:

## 4. Research methodology

The present research is descriptive and cross-sectional. The authors intend to evaluate the DT&Ts at five different Indian universities. The use of India as a case study aids in understanding how these mentee students learn. Around 20 million students are stakeholders




**Source(s):** Compiled by the authors

**Figure 1.**
Conceptual model (Proposed by authors)

in India's higher education system (Study in India, 2021). Courses in this country are inexpensive and diverse. In the "Worldwide Education for the Future Index 2019," India is ranked 35th (Sheth, 2020) because critical thinking, problem-solving, collaboration, innovation and entrepreneurship are valued. The educational system is being redesigned to be more result-oriented and global. The new education policy (NEP) aims to improve learning gains and literacy rates in rural areas (Kant, 2021). According to the Indian Brand Equity Foundation, the Indian education market is worth $100 billion, with higher education accounting for nearly half of it. The ed-tech market in the country is thriving, ranking second only to that of the United States (Ibef, 2021). International students flock to the country because of its reputation as a global educational center. Satya Nadella, Sundar Pichai and Indra Nooyi, the CEOs of Microsoft, Google and Pepsi, were all educated in India (Study in India, 2021). India also ranked second in terms of Internet users after China in the first quarter of 2021 (Johnson, 2021). The epidemic, the government's Digital India initiatives and the growing digital device industry contribute to this increase (Keelery, 2021; Thomas and Chopra, 2020). As a result, the world looks to India. Indian science and education are significant globally.

Convenience sampling with no pre-selected individuals is employed. Undergraduate students from HEIs (mentee universities) were approached for data collection for a period ranging from May 2021 to July 2021, when the country was confronting the second deadly wave of COVID-19. Though the data were collected from 726 respondents, only 675 responses could be identified as valid for the final analysis after cleaning.

The primary data collection was done through the structured questionnaire on a 7-point Likert scale ranging from 1-strongly disagree to 7-strongly agree, indicating the strength of agreement and the extent of ICT usage. The secondary sources in journals, books, articles, newspaper reports and websites were scanned thoroughly to develop constructs and the questionnaire. The formulated questionnaire examined the availability and usage of various DT&Ts and the CSR_LOs. It also reflected the future preference of the mode of classes – online, offline or hybrid.

A trial run in the form of a pilot survey was conducted on a sample of 120 respondents selected from the mentee universities. The questionnaire was created using Google Forms and distributed through digital channels like WhatsApp and e-mail because of strict lockdown.

### *4.1 Data preparation*

After data collection, screening was done to identify missing values and outliers. Thereafter, the data were appropriately cleaned and the assumptions were checked. The reliability and validity of the questionnaire were checked to determine the consistency and accuracy of the data collected. To determine reliability, Cronbach's alpha was calculated which was above the threshold level of 0.7 (Bagozzi and Yi, 1988; Kline, 1999) for all the constructs. The Cronbach's alpha values of FOTs, RII, VCTs and IT_IS are 0.894, 0.887, 0.785 and 0.797, respectively. Statements having low reliability were removed from the final analysis. Experts in the field were approached to check the content validity. To determine sample adequacy, KMO and Bartlett test was performed. The sample size for the study is 675, which is above the accepted sample size requirements. According to Hair *et al.* (2010), a minimum of five observations should be made for each independent variable in the variate for the sample to be representative. In this case, the KMO value is 0.879, which depicts great sample adequacy (Hutcheson and Sofroniou, 1999).

### *4.2 Statistical tools and analysis*

After data preparation, exploratory factor analysis was performed to derive key factors from the data collected on several variables. In this process, cross or low loadings items were dropped (Field, 2009).

The structural model is drawn to test the dependence relationship between the constructs (Malhotra and Dash, 2010). In this study, SEM is performed to determine the causal relationships between the IVs IT_IS, VCTs, FOTs and RII with the DV CSR_LO. First, a measurement model is drawn to check composite reliability (CR), convergent validity and discriminant validity (Anderson and Gerbing, 1988). In the analysis, all values were acceptable, i.e. CR > 0.7, average variance extracted (AVE) > 0.5 and CR > AVE, thus depicting convergent validity, which shows how well the observed variables in a construct correlate with each other (Hair *et al.*, 2010). The value of maximum shared variance (MSV) is less than AVE in the analysis confirming discriminant validity, i.e. a construct is distinct from other constructs and makes a unique contribution (Hair *et al.*, 2010).

After testing validity in the measurement model, the structural model was constructed and model fit was determined. The CMIN/DF value or chi-square is calculated to verify the model fit, which is how the expected values differ from observed values. In this case, the value is 2.996, which is below the threshold level of 3. The goodness of fit measure CFI is 0.954 and the badness of fit measure RMSEA is 0.054, which is within the acceptable limits (Table 2).

### *4.3 Regression results*

The hypotheses were tested to determine whether DT&Ts, namely, IT_IS, VCTs, FOTs and RII have a significant effect on learning in the HEIs. For this purpose, SEM in AMOS was performed which gave the following results (Table 3):

It was discovered that IT_IS, VCTs and FOTs have a significant effect on learning through CSR, while RII could not have a significant effect. The structural model in Figure 2 depicts the relationship of the IVs with the DV.

## 5. Discussion and conclusion

Experts have indicated that COVID-19 will deepen inequalities, therefore CSR researchers must re-examine theories of value creation for stakeholder identification and prioritization (Crane and Matten, 2021). The paper highlights how the mentor universities can find novel ways to generate value for stakeholders and add fresh perspectives in social responsibility theories using DT&Ts.

The DT&Ts significantly and positively affected learning through CSR during the COVID-19 pandemic. The same can also be verified from the $R$ square value of 55%. An efficient IT_IS positively and significantly affects learning performance in HEIs ($p < 0.001$, beta 0.23). This observation is in line with several studies that demonstrated the benefits of ICT for "improving the quality of education by facilitating access to resources and services as

| Measure | Estimate | Threshold | Interpretation |
|---|---|---|---|
| CMIN | 422.482 | – | – |
| DF | 141 | – | – |
| CMIN/DF | 2.996 | Between 1 and 3 | Excellent |
| CFI | 0.954 | >0.95 | Excellent |
| SRMR | 0.047 | <0.08 | Excellent |
| RMSEA | 0.054 | <0.06 | Excellent |
| PClose | 0.108 | >0.05 | Excellent |

**Source(s):** Hu and Bentler (1999), Gaskin and Lim (2016)

**Table 2.** Model fit measures

| DV | | IVs | Estimate | S.E. | C.R. | $p$ | Decision |
|---|---|---|---|---|---|---|---|
| CSR_LO | ← | IT_IS | 0.125 | 0.035 | 3.55 | *** | Supported |
| CSR_LO | ← | VCTs | 0.173 | 0.037 | 4.706 | *** | Supported |
| CSR_LO | ← | FOTs | 0.155 | 0.027 | 5.676 | *** | Supported |
| CSR_LO | ← | RII | 0.001 | 0.021 | 0.052 | 0.959 | Not supported |

**Table 3.** Regression results

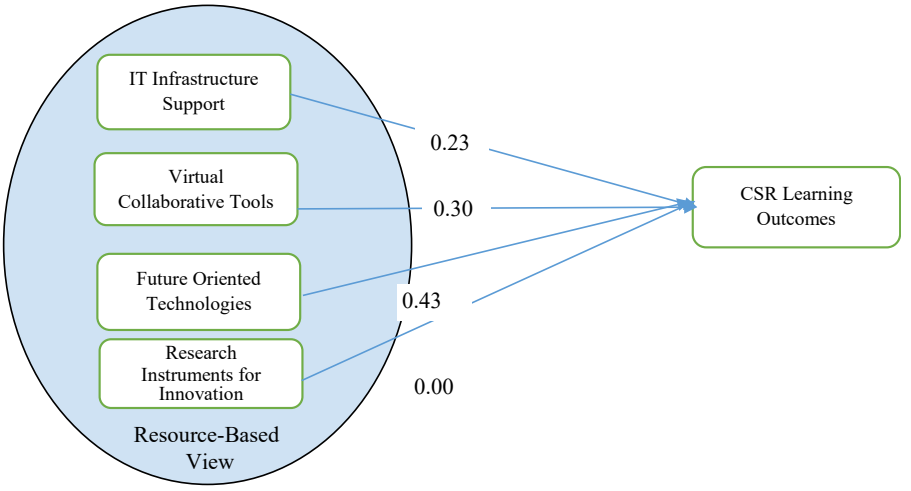



**Figure 2.** Structural model with path coefficients

well as remote exchange and collaboration" (Commission of the European Communities, 2001; IFC, 2021). ICT is viewed as a "major tool for building knowledge societies" (UNESCO, 2003). The latest equipment with supportive software facilitates students and teachers, renders convenience and reduces their stress by saving time and effort.

A good infrastructure is essential to meet the demands of a fast-paced, dynamic and tech-savvy world. Adequate availability of software, hardware and networks is vital to scale up the productivity of both teachers and students. The Internet has offered infinite opportunities to communicate, collaborate and co-create. Without it, online applications would not budge and the entire system would come to a halt. Technological interventions keep even a remote mentee institution connected to researchers and academicians worldwide while adopting the latest technologies redefines possibilities in the digital age. But, due to the pandemic, robust and seamless home IT_IS could not be accessible to all the students because of the digital divide, Internet instability and technical glitches.

*The virtual collaborative tools* have a significant and positive impact on the learning performance of its members ($p < 0.001$, beta 0.30). They have a high usage rate denoting that students use these tools for group communication in the mentee universities. With its acceptability, convenience and low cost, netizens have got accustomed to it (Prensky, 2001; Amin *et al.*, 2018; Jain and Gupta, 2019a, b).

Online communities of practice like WhatsApp groups and Google groups are accredited to have united people in a group. These communities allow members to share, create and post information with flexibility and convenience. Members collaborate for knowledge and support (De Vries *et al.*, 2000; Wenger, 2011; Forkosh-Baruch and Hershkovitz, 2012; Singh and Gupta, 2020, 2022). This sense of community improves teamwork and productivity as all members work towards a common goal and in tandem with each other.

Video conferencing tools like Zoom, MS Teams, Google Meet and WebEx are subscribed by mentors for interaction and are employed for tutoring, webinars and conferences. These tools have options available for recording, chatting and calling as well. These tools are synchronous and make real-time video interaction possible. They also support the office applications of text, presentations and mathematical calculations. Peer editing tools like Google Docs and SkyDrive make knowledge sharing, brainstorming and idea generation possible by establishing meaningful connections among mentors and mentees through teamwork (Al-Samarraie and Saeed, 2018). These exchanges promote the intellectual, social and emotional development of members. Peer editing saves time as it frees individuals from the hassles of mailing repeatedly after revisions and corrections. A shared document allows members to review, reflect and track the progress of other members. They can work autonomously and both synchronously and asynchronously with others. They can also collaborate for problem-solving, critical thinking and knowledge generation.

The pandemic and lockdown of schools and colleges acted as a catalyst in promoting online learning through MOOCs. The experts from the mentor universities were able to interact with the students of mentee institutions through MOOCs and these courses contributed to bridging the skill gap and temporal and spatial differences for learning continuity and enhancement. The lockdown period is efficiently utilized by many students of the mentee universities to supplement class learning and acquire new skills. When students were isolated, they used the discussion forums in MOOCs to learn virtually and share their opinions (Kumar *et al.*, 2021). In the current study, it has been found that FOTs have low usage in developing countries, but they significantly affect learning performance ($p < 0.05$, beta 0.43).

At the beginning of the nationwide lockdown, there was a kind of asynchronous learning wherein the mentors posted study material via mail or uploaded video lectures on LMS or YouTube. Though effective in maintaining continuity, there was no peer interaction or live discussion. In this context, the use of in-class activities could engage learners. The in-class

activities could be conducted in a flipped classroom model-both offline and or online (Hew *et al.*, 2020). The in-class activities would be synchronous and promote active learning through live discussions, debates, projects, quizzes, role-playing, lab work, presentations and feedback. Even breakout rooms could be developed for group-level interactions that aid clear understanding and concept building through active social interactions. During these activities, the comprehension, analysis and application helped reinforce the concepts learned.

In online distance learning, it is observed that recipients' concentration drops quickly as they passively receive knowledge. So, gamification was used to pep up dropped interest levels. Games are naturally appealing and can generate positive feelings toward learning. Not only do they motivate the learners but also immerse their senses in a thrilling experience. With its inherent challenge, feedback and competence, Gamification helps learners stay focused in an environment that can otherwise be inhibited by distractions.

VR, AR and AI thought of as emerging technologies, registered their presence in the educational landscape during the pandemic (Kupe and Ouma, 2021). Though their existence was limited in developing countries and remote areas, their scope and application saw a sharp rise, especially in elite mentor institutions. Though the power of AI is not fully harnessed in many HEIs, the technology was used for auto-suggestion, pattern recognition, language translation, visual perception, data analysis and reporting. Similarly, VR was used to virtually transport students to distant locations, work fields, museums and laboratories for experiential learning, as physical movement was restricted at the time of lockdown. AR overlays virtual objects like models, machines and aircraft in real-world environments for three-dimensional inspection and training. These trending technologies immersed the user and enthralled them when they were otherwise isolated and separated by time and distance.

The *research instruments for innovation,* such as the database and software subscriptions, is the e-resources necessary to aid mentee institutions' teaching, learning and research endeavors. The Center for Monitoring Indian Economy (CMIE) prowess database stocks financial data of companies, while JSTOR houses scholarly content and research publications. Turnitin detects plagiarism and improves student outcomes by laying the foundation for critical thinking, original writing and academic integrity practices. SPSS is used for data crunching and statistical analysis. Subscription to all these software and databases would provide institutions with a knowledge edge that will translate into valued research publications, better lecture notes, enhanced student engagement (Dwivedi *et al.*, 2019) and better grades or results.

RII is the only factor that does not significantly affect CSR_LO ($p > 0.05$, beta 0.00). Many reasons can be attributed to it. The research software is expensive and requires a regular subscription and up-gradation. As society was wrecked with various physical and mental health issues, the focus was on strengthening fundamental concepts rather than advancements in the field. Moreover, the students at the undergraduate level use basic office applications rather than advanced research tools.

The paper also investigated the preferred mode of classes (online, offline, or hybrid) for students' exchange in a post-pandemic environment in the mentee universities. It has been found that the DT&Ts have given immense power and benefits to the students in their learning endeavors, but it has not reached a level where they can be completely banked upon and physical interactions can be eliminated. This is clear from the survey results, wherein 48.2% of students from these universities opted for blended learning; 41.9% opted for physical knowledge exchanges, and only 9.9% of students opted for online interactions.

The students are against the notion of completing online classes due to the lack of motivation, fatigue and isolation. But they suggested a careful blend of online for basic conceptual learning and offline for high-level conceptual and experiential learning to reap the benefits of both.

## 6. Implications

The CSR initiatives undertaken by the mentor universities boosted learning amongst the mentee institutions even at the time of the pandemic. These virtual collaborations assisted mentee institutions in upgrading their disaster management capabilities. It has pushed weaker institutions to change their traditional practices and become receptive to innovative approaches.

Due to the emergent lockdown, the age-old tradition of physical classes had to give way to online classes. However, the trend is expected to continue (Ketchell, 2021). The education model has been transformed and a redesign of activities is demanded to build a hybrid model that combines the advantages of both online classes and offline modes. In this way, the new normal will overwhelm the conventional limitations in education and provide opportunities for creative mindsets. The intent is not to provide an entirely new learning method but to create understanding amongst the researchers and educators regarding learning in the context of mentor-mentee universities. Also, mentoring benefits help students and employers or mentors who gain leadership abilities through this process. Based on this premise, implementing e-mentoring in HEIs requires educational leaders to be willing to assume the role of mentor to advise both teachers and students, be trained to serve as mentors and invest in their training. This will also include work on connecting students and teachers with other leaders who could serve as mentors in institutional projects and be integrated into mentoring development to provide a managerial and global vision. For example, under the "Paramarsh" scheme of the University Grants Commission (UGC), the mentor institution Stella Maris College in Chennai state is mentoring the mentee institution St. Joseph's College of Kovur and St. Anne's College in Madhavaram to help them achieve quality standards to get accredited with a better score (Stella Maris College, 2021; UGC, 2019). The mentors conducted a training session on Technology-enabled Learning to keep the mentees abreast with various FOTs and VCTs of teaching and learning. Furthermore, there should be a greater emphasis on IT infrastructure services in aiding students in attaining diverse aims. Consider, for instance, students' evaluations and feedback on the performance of the ICT division; adapt additional channels for students to communicate with ICT staff, such as chat; and provide students with some online courses or educational videos and lectures on how to use e-learning systems and their primary functions.

In the current research, the mentor institutions augment the RBV by organizing and managing technological resources for the mentee institutions as a part of their CSR initiatives. The findings of this paper could help HEIs understand the dynamics of various DT&Ts for creating a rich learning environment for the students. The research also concentrates on utilizing these tools to make the knowledge transfer prolific and influential. HEIs have to skill students for the jobs of tomorrow and serve the community at large. In this endeavor, academia can get immense support from mentors' guidance available through DT&Ts. The CSR_LOs from past experimentation help the mentors to deliver a better learning experience because they are better equipped. The mentors have to consider both the synchronous and asynchronous learning modes and blend them to enhance student engagement. The bridging could be done by embracing the rich diversity of DT&Ts for collaborative learning. A rewarding hybrid class discussion could be created by leveraging DT&Ts. The mantra to success is the same as it has always been encouraging students to organize and participate, be engaged, question and listen well and create social capital through DT&Ts (Schiano, 2021). This study is unique and the first to develop empirical insights into DT&Ts in academia and their impact on CSR_LOs.

According to Carroll (2021), economic responsibility is a baseline requirement that must be met in a competitive business environment and nowhere is this more evident than in the current pandemic world, where many businesses are struggling to survive, let alone thrive. The pandemic has widened the divide between the haves and have-nots. The mentor universities are better placed in DT&Ts and can perform better. On the other hand, mentee

universities do not have access to such resources and are more prone to contraction. In such situations, it is the social responsibility of the mentor university to share resources with the mentee university and assist pro-actively.

The resource-based perspective identifies physical, social, organizational, human and financial resources as the sources of competitive advantage (Greene and Brown, 1997). Thereafter, subsequent research suggested that technology could also be a strategic resource (Venkataraman, 2004; Wade and Hulland, 2004; Bakar and Ahmad, 2010). But post the COVID-19 pandemic, technological resources have occupied a premium place amongst other resources due to its tremendous growth. As Microsoft's Satya Nadella remarked, due to COVID-19, "two years of digital transformation have happened in two months" (Spataro, 2020). It is also evident from the results of the study that technological resources are a source of competitive advantage and strengthen teaching-learning endeavors. The mentor institutions that had invested in IT_IS, VCTs and FOTs not only benefited themselves but also supported mentee institutions by enabling effective learning even during the complete lockdown. In the realm of CSR, the idea is not only competition but mutual growth through sharing resources with the various stakeholders. Collaboration and knowledge sharing create synergies that enable institutions to create a better world through their commitment to CSR initiatives. The management should measure the worth of the earlier-mentioned resources not only in terms of their competitive advantage but also community advantage. In light of this notion, the definition of the RBV should change accordingly. The positive impact of these resources on the community needs to be gauged, considering the digital divide and growing inequities in resource division in the post-pandemic world.

## 7. Recommendations

As the online learning environment is expected to stay at least in a blended form, adequate CSR funding in infrastructure is necessitated to harness the full potential of technology. The survey results showed that students experience slow or unreliable Internet connectivity and that causes hiccups in learning. An enhancement in Internet bandwidth and the provision of necessary hardware equipment can contribute to a better online learning experience.

The IT_IS having adequate availability of software, hardware and network is critical for collaborative learning. The research experimentations, though ignored at the peak of a wave for the students at the undergraduate level, need to be reinstated firmly in the quest for innovation and development. Though in-person social interactions were not possible due to the pandemic and resultant spatial and temporal barriers, the technology has enabled virtual collaborations, which must be effectively utilized for participation, engagement, knowledge creation, dissemination and active learning. VCTs have enhanced the way students interact, comprehend, contribute and acquire new relevant knowledge.

The emerging technologies received a significant boost during the lockdown, but their availability is limited in developing countries. Some of them, like AI, gaming, flipped classrooms and VR can be used to enhance teaching-learning and assessment. Even though the FOTs are not available in plenitude, these have provided an engaging and inundated experience overcoming the barriers of time and distance. To successfully integrate the emerging upcoming technologies, the students and teachers must be well-trained in their use and application.

## 8. Limitations, future scope and future of education

Like any other empirical research, this study also has some limitations. In this study, the authors have examined a few critical but not all of the CSR concepts and premises. The list of DT&Ts employed for this research is extensive but may not be exhaustive. The sample is

limited to 675 students from five universities in India, which could be enlarged in the future. This study manifests a framework for learning with DT&Ts in the Indian context. It has scope for further research in India and other parts of the world. Also, it will be interesting to examine how the innovations in DT&Ts can provide novel solutions to the challenges of tomorrow. Since India is one of the largest and thriving economies of the world, its endeavor at reforms through DT&Ts under the CSR initiative is going to revolutionize educational strategies for the world. Although all these factors can be applied for mentoring, they may vary in the extent to which they are suitable for collaborative learning.

Blended learning in this context can bring a major paradigm shift in higher education. This claim was validated through survey results of students. The switch to the online mode that initially began as a temporary recourse will become the "new normal" for the HEIs. The pandemic played a crucial role in legitimizing online collaboration and making it more acceptable in society. It has offered flexibility for delivery, location and remote collaboration through synchronous DT&Ts. It is now possible to maintain networks around the world for knowledge sharing and idea generation (Chopra and Gupta, 2019). Hence, a reallocation from the investment in physical infrastructure to the IT infrastructure could be observed (Ewing, 2021). Also, the Ed-tech industry is integrating new technologies like AI, VR, gamification, 3D printing and digital inking into education at all levels (Walker, 2018; Terä s *et al.*, 2020). However, despite advancements, technology still suffers from various limitations and cannot substitute offline learning, which is here to stay as it provides more personalized, synchronized, experiential and hands-on learning to the students. It has time-tested benefits of field-based projects, group assignments, research-based teaching and laboratory experiments. Moreover, offline learning sufficiently addresses the lack of motivation, direction, digital divide and health issues among learners. Thus, the future of higher education belongs to a collaboration-based hybrid model between mentor and mentee institutions that maximizes the learning synergies between HEIs on the path to growth trajectory toward a socially responsible and sustainable world. Forthcoming developments of DT&Ts for mentoring and knowledge-building purposes will shape the learning of future generations. Using various methodologies, there will be opportunities to conduct extended research in both developed and developing countries.

## References


Ahmad, T. (2015), "Preparing for the future of higher education", *On the Horizon*, Vol. 23 No. 4, pp. 323-330, doi: 10.1108/oth-06-2015-0029.

Al-Samarraie, H. and Saeed, N. (2018), "A systematic review of cloud computing tools for collaborative learning: opportunities and challenges to the blended-learning environment", *Computers and Education*, Vol. 124, pp. 77-91.

Alanis-Funes, G.J., Neri, L. and Noguez, J. (2011), "Virtual collaborative space to support active learning", *Frontiers in Education Conference (FIE)*, pp. F3C-1-F3C-6, doi: 10.1109/FIE.2011.6143078.

Alsabawy, A.Y., Cater-Steel, A. and Soar, J. (2013), "IT infrastructure services as a requirement for e-learning system success", *Computers and Education*, Vol. 69, pp. 431-451.

Amin, N.A.H.N., Almunawar, M.N., Hasnan, A.S. and Besar, N.N. (2018), "Preferences, benefits, and barriers of web 2.0 tools for knowledge sharing in Brunei Darussalam's tertiary education", *Management Strategies and Technology Fluidity in the Asian Business Sector*, IGI Global, pp. 253-276.

Anderson, J.C. and Gerbing, D.W. (1988), "Structural equation modeling in practice: a review and recommended two-step approach", *Psychological Bulletin*, Vol. 103 No. 3, pp. 411-423.

Apostol, A.C. (2020), "Students' perceptions about the impact of COVID-19 on learning process. A sociological approach", *Technium Social Sciences Journal*, Vol. 9, pp. 495-500.

Asabere, N., Togo, G., Acakpovi, A., Torgby, W. and Ampadu, K. (2017), "AIDS: an ICT model for integrating teaching, learning and research in Technical University Education in Ghana", *International Journal of Education and Development Using ICT*, Vol. 13 No. 3, pp. 162-183.

Bagozzi, R.P. and Yi, Y. (1988), "On the evaluation of structural equation models", *Journal of the Academy of Marketing Science*, Vol. 16 No. 1, pp. 74-94.

Bakar, L.J.A. and Ahmad, H. (2010), "Assessing the relationship between firm resources and product innovation performance: a resource-based view", *Business Process Management Journal*, Vol. 16 No. 3, pp. 420-435.

Bamford, A. (2012), "Page the 3D in education white paper", available at: http://webcache.googleusercontent.com/search?q=cache:KZX24QhrGt8J:www.eonreality.com/download/The%25203D%2520in%2520Education%2520White.pdf+&cd=1&hl=en&ct=clnk&gl=in (accessed 20 May 2016).

Barbagila, M., Buttice, V., Giudici, G., Mendy, J., Sarker, T., Sharma, G.D., Thomas, A. and Zutshi, A. (2021), "Supporting SMEs in sustainable strategy development post-Covid-19: challenges and policy agenda for the G20".

Barney, J. (1991), "Firm resources and sustained competitive advantage", *Journal of Management*, Vol. 17 No. 1, pp. 99-120.

Bates, T. (2014), "The role of communities of practice in a digital age", Tony Bates, Tonybates.ca, available at: https://www.tonybates.ca/2014/10/01/the-role-of-communities-of-practice-in-a-digital-age/ (accessed 28 April 2019).

Battisti, E., Nirino, N., Leonidou, E. and Thrassou, A. (2022), "Corporate venture capital and CSR performance: an extended resource based view's perspective", *Journal of Business Research*, Vol. 139, pp. 1058-1066.

Baturay, M.H. (2015), "An overview of the world of MOOCs", *Procedia – Social and Behavioral Sciences*, Vol. 174, pp. 427-433, doi: 10.1016/j.sbspro.2015.01.685.

Bechter, C. and Swierczek, F.W. (2017), "Digital storytelling in a flipped classroom for effective learning", *Education Sciences*, Vol. 7 No. 2, doi: 10.3390/educsci7020061.

Bertling, J., Rojas, N., Alegre, J. and Faherty, K. (2020), "A tool to capture learning experiences during COVID-19: the PISA global crises questionnaire module, OECD Publishing, Paris", OECD Education Working Papers, No. 232, doi: 10.1787/9988df4e-en.

Bogdandy, B., Tamas, J. and Toth, Z. (2020), "Digital transformation in education during covid-19: a case study", *2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, IEEE, Mariehamn, pp. 000173-000178, doi: 10.1109/CogInfoCom50765.2020.9237840.

Brady, K.P., Holcomb, L.B. and Smith, B.V. (2010), "The use of alternative social networking sites in higher educational settings: a case study of the e-learning benefits of Ning in education", *Journal of Interactive Online Learning*, Vol. 9 No. 2, pp. 151-170.

Brandeis University (2015), "How technology has changed the meeting", available at: http://projectmgmt.brandeis.edu/how-technology-has-changed-the-meeting/ (accessed 29 May 2016).

Bratianu, C. and Vatamanescu, E.M. (2017), "Students' perception on developing conceptual generic skills for business: a knowledge-based approach", *VINE Journal of Information and Knowledge Management Systems*, Vol. 47 No. 4, pp. 490-505, doi: 10.1108/VJIKMS-11-2016-0065.

Brodeur, D.R. (2013), "Mentoring young adults in the development of social responsibility", *Australasian Journal of Engineering Education*, Vol. 19 No. 1, pp. 13-25.

Byrd, T.A. and Turner, D.E. (2001), "An exploratory examination of the relationship between flexible IT infrastructure and competitive advantage", *Information and Management*, Vol. 39 No. 1, pp. 41-52.

Campillo-Ferrer, J.M. and Miralles-Martínez, P. (2021), "Effectiveness of the flipped classroom model on students' self-reported motivation and learning during the COVID-19 pandemic", *Humanities and Social Sciences Communications*, Vol. 8 No. 1, pp. 1-9.

Carroll, A.B. (2021), "Corporate social responsibility (CSR) and the COVID-19 pandemic: organizational and managerial implications", *Journal of Strategy and Management*, Vol. 14 No. 3, pp. 315-330.

Caspersen, J., Smeby, J.C. and Olaf Aamodt, P. (2017), "Measuring learning outcomes", *European Journal of Education*, Vol. 52 No. 1, pp. 20-30.

Chandere, V., Satish, S. and Lakshminarayanan, R. (2021), "Online plagiarism detection tools in the digital age: a review", *Annals of the Romanian Society for Cell Biology*, Vol. 25 No. 1, pp. 7110-7119.

Chigona, A. (2015), "Pedagogical shift in the twenty-first century: preparing teachers to teach with new technologies", *Africa Education Review*, Vol. 12 No. 3, pp. 478-492.

Chiu, C.M., Hsu, M.H. and Wang, E.T. (2006), "Understanding knowledge sharing in virtual communities: an integration of social capital and social cognitive theories", *Decision Support Systems*, Vol. 42 No. 3, pp. 1872-1888.

Chopra, M. and Gupta, V. (2019), "Linking knowledge management practices to organizational performance using the balanced scorecard approach", *Kybernetes*, Vol. 49 No. 1, pp. 88-115.

ColdFusion (2016), "What is artificial intelligence exactly?", available at: https://www.youtube.com/watch?v=kWmX3pd1f10&list=LLgAzVc7se-IArBd1L8fQqOw&index=11&t=105s.

Commission of the European Communities (2001), *The eLearning Action Plan. Designing Tomorrow's Education*, Directorate General for Education and Culture, Brussels, pp. 1-19.

Correia, A.P., Liu, C. and Xu, F. (2020), "Evaluating videoconferencing systems for the quality of the educational experience", *Distance Education*, Vol. 41 No. 4, pp. 429-452.

Crane, A. and Matten, D. (2021), "COVID-19 and the future of CSR research", *Journal of Management Studies*, Vol. 58, pp. 280-284.

Curtin, R. (2021), "Reimagining higher education: the post-covid classroom", available at: https://er.educause.edu/articles/2021/4/reimagining-higher-education-the-post-covid-classroom> (accessed 11 September 2021).

Davies, S. (2017), "Augmented reality and its potential effect on education", available at: http://www.iamwire.com/2017/06/augmented-reality-education/154484 (accessed 15 September 2021).

De Vries, S., Bloemen, P. and Roossink, L. (2000), "Online knowledge communities", *WebNet World Conference on the WWW and Internet*, Association for the Advancement of Computing in Education (AACE), San Antonio, TX, pp. 124-129.

Department of Education and Training (2021), "Teach with digital technologies", *Education.vic.gov.au*, available at: https://www.education.vic.gov.au/school/teachers/teachingresources/digital/Pages/teach.aspx> (accessed 25 January 2021).

Dima, A.M., Vasilache, S., Ghinea, V. and Agoston, S. (2013), "A model of academic social responsibility", *Transylvanian Review of Administrative Sciences*, Vol. 9 No. 38, pp. 23-43.

Dominici, G. and Palumbo, F. (2013), "How to build an e-learning product: factors for student/customer satisfaction", *Business Horizons*, Vol. 56 No. 1, pp. 87-96.

Drucker, J. and Goldstein, H. (2007), "Assessing the regional economic development impacts of universities: a review of current approaches", *International Regional Science Review*, Vol. 30 No. 1, pp. 20-46.

Duncan, N.B. (1995), "Capturing flexibility of information technology infrastructure: a study of resource characteristics and their measure", *Journal of Management Information Systems*, Vol. 12 No. 2, pp. 37-57.

Dwivedi, A., Dwivedi, P., Bobek, S. and Sternad Zabukovšek, S. (2019), "Factors affecting students' engagement with online content in blended learning", *Kybernetes*, Vol. 48 No. 7, pp. 1500-1515, doi: 10.1108/K-10-2018-0559.

Evering, L.C. and Moorman, G. (2012), "Rethinking plagiarism in the digital age", *Journal of Adolescent and Adult Literacy*, Vol. 56 No. 1, pp. 35-44.

Ewing, L.A. (2021), "Rethinking higher education post COVID-19", *The Future of Service Post-COVID-19 Pandemic*, Springer, Vol. 1, pp. 37-54.

Field, A. (2009), *Discovering Statistics Using SPSS*, 3rd ed., Sage publications, London.

Fisk, P. (2017), "Education 4.0 ... the future of learning will be dramatically different, in school and throughout life", *Geniusworks*, available at: https://www.thegeniusworks.com/2017/01/future-education-young-everyone-taught-together/ (accessed 16 June 2018).

Forkosh-Baruch, A. and Hershkovitz, A. (2012), "A case study of Israeli higher-education institutes sharing scholarly information with the community via social networks", *The Internet and Higher Education*, Vol. 15 No. 1, pp. 58-68.

Freeman, R.E. (1994), "The politics of stakeholder theory: some future directions", *Business Ethics Quarterly*, Vol. 4 No. 4, pp. 409-421.

Fry, H., Ketteridge, S. and Marshall, S. (2008), "Understanding student learning", in *A Handbook for Teaching and Learning in Higher Education*, Routledge, pp. 26-44.

Gangi, F., Mustilli, M. and Varrone, N. (2018), "The impact of corporate social responsibility (CSR) knowledge on corporate financial performance: evidence from the European banking industry", *Journal of Knowledge Management*, Vol. 23, pp. 110-134, doi: 10.1108/JKM-04-2018-0267.

García-Morales, V.J., Garrido-Moreno, A. and Martín-Rojas, R. (2021), "The transformation of higher education after the COVID disruption: emerging challenges in an online learning scenario", *Frontiers in Psychology*, Vol. 12, p. 196.

Gaskin, J. and Lim, J. (2016), *Master Validity Tool*, AMOS Plugin. Gaskination's StatWiki, Provo, UT.

Gaur, N. and Gupta, V. (2021), "Devising a knowledge culture", *Journal of Contemporary Issues in Business and Government*, Vol. 27 No. 1, pp. 1587-1623.

George Glance, D., Forsay, M. and Riley, M. (2013), "The pedagogical foundations of massive open online courses", *First Monday*, Vol. 18 No. 5.

Ghada Refaat El Said (2021), "How did the COVID-19 pandemic affect higher education learning experience? An empirical investigation of learners' academic performance at a university in a developing country", *Advances in Human-Computer Interaction*, pp. 1-10, doi: 10.1155/2021/6649524.

Gopalan, C. (2021), "Remote flipped teaching during the COVID-19 pandemic", *The FASEB Journal*, Vol. 35 No. S1, pp. 40-56, doi: 10.1096/fasebj.2021.35.S1.03727.

Greene, P.G. and Brown, T.E. (1997), "Resource needs and the dynamic capitalism typology", *Journal of Business Venturing*, Vol. 3 No. 12, pp. 161-173.

Gupta, V. (2020), "Understanding the dynamics of knowledge management tools in two public universities in Delhi, India", *Knowledge Management Practices in the Public Sector*, IGI Global, pp. 145-177, Chapter.

Gupta, V. and Jain, N. (2017), "Harnessing information and communication technologies for effective knowledge creation: shaping the future of education", *Journal of Enterprise Information Management*, Vol. 30 No. 5, pp. 831-855.

Gupta, V. and Thomas, A. (2019), "Fostering tacit knowledge sharing and innovative work behaviour: an integrated theoretical view", *International Journal of Managerial and Financial Accounting*, Vol. 11 Nos 3-4, pp. 320-346.

Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2010), *Multivariate Data Analysis*, 7th ed., Prentice-Hall, Upper Saddle River, NJ.

Hajizada, S. (2019), *Emerging Trends in Educational Technology*, Blog, available at: https://carleton.ca/edc/2019/blog-emerging-trends-in-educational-technology/.

Hastings, L.J. and Sunderman, H.M. (2020), "Evidence-based practices in mentoring for leadership development", *New Directions for Student Leadership*, Vol. 2020 No. 168, pp. 75-84.

Helm, S. and Tolsdorf, J. (2013), "How does corporate reputation affect customer loyalty in a corporate crisis? Corporate reputation, customer loyalty and crises", *Journal of Contingencies and Crisis Management*, Vol. 21 No. 3, pp. 144-152, doi: 10.1111/1468-5973.12020.

Hernández-Sellés, Muñoz-Carril and González-Sanmamed (2019), "Computer-supported collaborative learning: an analysis of the relationship between interaction, emotional support and online collaborative tools", *Computers and Education*, Vol. 138 No. 2019, pp. 1-12, doi: 10.1016/j.compedu.2019.04.012.

Hew, K.F., Jia, C., Gonda, D.E. and Bai, S. (2020), "Transitioning to the 'new normal' of learning in unpredictable times: pedagogical practices and learning performance in fully online flipped classrooms", *International Journal of Educational Technology in Higher Education*, Vol. 17 No. 1, pp. 1-22, doi: 10.1007/978-981-33-4126-5_3.

Hu, L. and Bentler, P.M. (1999), "Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives", *SEM*, Vol. 6 No. 1, pp. 1-55.

Hutcheson, G.D. and Sofroniou, N. (1999), *The Multivariate Social Scientist: Introductory Statistics Using Generalized Linear Models*, Sage, London.

Ibef (2021), "Education industry analysis", available at: https://www.ibef.org/industry/education-presentation.

IFC (2021), "What COVID-19 means for digital infrastructure in emerging markets", *Ifc.org*, available at: https://www.ifc.org/wps/wcm/connect/publications_ext_content/ifc_external_publication_site/publications_listing_page/what+covid+19+means+for+digital+infrastructure+in+emerging+markets> (accessed 8 August 2021).

Impey and Formanek (2021), "MOOCS and 100 Days of COVID: enrollment surges in massive open online astronomy classes during the coronavirus pandemic", *Social Sciences and Humanities Open*, Vol. 4 No. 1, pp. 1-9, 100177.

Irby, B.J. (2015), "Editor's overview: mentoring as social responsibility", *Mentoring and Tutoring: Partnership in Learning*, Vol. 23 No. 2, pp. 95-99.

Jain, K. (2018a), "Role of AR/VR in the education system: how technology is shaping India's learning space", *India Today*, available at: https://www.indiatoday. in/education-today/featurephilia/story/role-of-augmented-virtual-reality-in-education-1417739-2018-12-26.

Jain, N. (2018b), "Transforming accounting through cloud computing", *Vivekananda Journal of Research*, Vol. 7 No. 2, pp. 137-158.

Jain, N. and Gupta, V. (2019a), "The impact of knowledge management system on student performance: a case study of the University of Delhi", *VINE Journal of Information and Knowledge Management Systems*, Vol. 49 No. 1, pp. 115-135, doi: 10.1108/VJIKMS-07-2018-0065.

Jain, N. and Gupta, V. (2019b), "Communities of practice for digital knowledge management: a case study of web 2.0 in the university of Delhi", *International Journal of Public Sector Performance Management*, available at: https://www.inderscience.com/info/ingeneral/forthcoming.php?jcode=ijpspm (accessed 20 May 2020).

Jameel, A.S. (2020), "The role of information and communication technology on knowledge sharing among the academic staff during COVID-19 pandemic", *2020 2nd Annual International Conference on Information and Sciences (AiCIS)*, IEEE, pp. 141-147.

Johnson, J. (2021), *Most internet users by country | Statista*, Statista, available at: https://www.statista.com/statistics/262966/number-of-internet-users-in-selected-countries/> (accessed 14 September 2021).

Kant (2021), "The future of learning in India is ed-tech", available at: https://indianexpress.com/article/opinion/columns/the-future-of-learning-in-india-is-ed-tech-pandemic-online-classes-7381782/ (accessed 28 Aug 2021).

Kaup, S., Jain, R., Shivalli, S., Pandey, S. and Kaup, S. (2020), "Sustaining academics during COVID-19 pandemic: the role of online teaching-learning", *Indian Journal of Ophthalmology*, Vol. 68 No. 6, pp. 1220-1221.

Keelery, S. (2021), *India: Internet Inclusivity Index by Category 2021 | Statista*, Statista, available at: https://www.statista.com/statistics/1232322/internet-inclusivity-index-by-category-india/#statisticContainer> (accessed 14 September 2021).

Ketchell, M. (2021), "Post COVID-19: opportunity for universities to have a rethink", *The Conversation*, available at: https://theconversation.com/post-covid-19-opportunity-for-universities-to-have-a-rethink-149474> (accessed 11 September 2021).

King, A. (2007), "Scripting collaborative learning processes: a cognitive perspective", in Fischer, F., Kollar, I., Mandl, H. and Haake, J.M. (Eds), *Scripting Computer-Supported Collaborative Learning: Cognitive, Computational and Educational Perspectives*, Springer, New York, pp. 13-37, doi: 10.1007/978-0-387-36949-5_2.

Kirschner, P., Strijbos, J.W., Kreijns, K. and Beers, P.J. (2004), "Designing electronic collaborative learning environments", *Educational Technology Research and Development*, Vol. 52 No. 3, pp. 47-66.

Kline, T.J. (1999), "The team player inventory: reliability and validity of a measure of predisposition toward organizational team-working environments", *Journal for Specialists in Group Work*, Vol. 24 No. 1, pp. 102-112.

Kumar, M., Mamgain, P. and Pasumarti, S.S. (2021), "Students'e-satisfaction examined online community learning, technology fit, and compatibility sem–modelling during covid-19 pandemic", *Information Technology in Industry*, Vol. 9 No. 2, pp. 625-646.

Kupe, T. and Ouma, G. (2021), "Post COVID-19: opportunity for universities to have a rethink", *The Conversation*, available at: https://theconversation.com/post-covid-19-opportunity-for-universities-to-have-a-rethink-149474> (accessed 29 August 2021).

Kwon, K., Liu, Y. and Johnson, L. (2014), "Group regulation and social-emotional interactions observed in computer supported collaborative Learning: comparison between good vs poor collaborators", *Computers and Education*, Vol. 78, pp. 185-200, doi: 10.1016/j.compedu.2014.06.004.

Laurillard, D. (2002), *Rethinking university Teaching: a Conversational Framework for the Effective Use of Learning Technologies*, 2nd ed., Routledge Falmer, London.

Lawrence, J.E. and Tar, Usman .A. (2018), "Factors that influence teachers' adoption and integration of ICT in teaching/learning process", *Educational Media International*, Vol. 55 No. 1, pp. 79-105.

Malhotra, N.K. and Dash, S. (2010), *Marketing Research: an Applied Orientation*, 6th ed., Pearson Education, Noida.

Martínez, P. and Rodríguez del Bosque, I. (2013), "CSR and customer loyalty: the roles of trust, customer identification with the company and satisfaction", *International Journal of Hospitality Management*, Vol. 35, pp. 89-99, doi: 10.1016/j.ijhm.2013.05.009.

Medina, R. and Suthers, D. (2008), *Bringing Representational Practice from Log to Light*, University of Hawai'i.

McLoughlin, C. and Lee, M.J.W. (2007), "Social software and participatory learning: pedagogical choices with technology affordances in the Web 2.0 era", *ICT: Providing choices for learners and learning'. Proceedings ascilite Singapore*, available at: http://www.ascilite.org.au/conferences/singapore07/procs/mcloughlin.pdf.

Mishra, L., Gupta, T. and Shree, A. (2020), "Online teaching-learning in higher education during lockdown period of COVID-19 pandemic", *International Journal of Educational Research Open*, Vol. 1 No. 100012, pp. 1-8.

MIT (2016), "iCampus", available at: http://icampus.mit.edu/ (accessed 20 October 2016).

Munfarid (2018), "Artificial intelligence and mixed reality in education – gateway to future oriented classrooms", *Blog*, available at: https://munfarid.org/artificial-intelligence-mixed-reality-education.

Naresh, R. (2020), "Education after COVID-19 crisis based on ICT tools", *Purakala*, Vol. 31 No. 37, pp. 464-468.

Nason, R.S. and Wiklund, J. (2018), "An assessment of resource-based theorizing on firm growth and suggestions for the future", *Journal of Management*, Vol. 44 No. 1, pp. 32-60.

Nieto-Escamez, F.A. and Roldán-Tapia, M.D. (2021), "Gamification as online teaching strategy during COVID-19: a mini-review", *Frontiers in Psychology*, Vol. 12 No. 1, p. 1644, doi: 10.3389/fpsyg.2021.648552.

Nirino, N., Battisti, E., Ferraris, A., Dell'Atti, S. and Briamonte, M.F. (2022), "How and when corporate social performance reduces firm risk? The moderating role of corporate governance", *Corporate Social Responsibility and Environmental Management*, Special Issue, pp. 1-11.

Novak, A. (2017), "Knowledge management and organizational performance–literature review", *Management Challenges in a Network Economy: Proceedings of the MakeLearn and TIIM International Conference*, ToKnowPress, pp. 433-440.

OECD (2010), "Are the new millennium learners making the grade? Technology use and educational performance in PISA", available at: http://browse.oecdbookshop.org/oecd/pdfs/browseit/9609101E.PDF.

Prensky, M. (2001), "Digital natives, digital immigrants", part 1, *On the Horizon*, Vol. 9 No. 5, pp. 1-6.

Radu, M.C., Schnakovszky, C., Herghelegiu, E., Ciubotariu, V.A. and Cristea, I. (2020), "The impact of the COVID-19 pandemic on the quality of educational process: a student survey", *International Journal of Environmental Research and Public Health*, Vol. 17 No. 21, p. 7770.

Rahman, A.A., Castka, P. and Love, T. (2019), "Corporate social responsibility in higher education: a study of the institutionalisation of CSR in Malaysian public universities", *Corporate Social Responsibility and Environmental Management*, Vol. 26 No. 4, pp. 916-928.

Rajhans, V., Memon, U., Patil, V. and Goyal, A. (2020), "Impact of COVID-19 on academic activities and way forward in Indian Optometry", *Journal of Optometry, India Scenario*, Vol. 13, doi: 10.3390/ijerph17217770.

Ranga, M. and Etzkowitz, H. (2015), "Triple Helix systems: an analytical framework for innovation policy and practice in the Knowledge Society", *Entrepreneurship and Knowledge Exchange*, Vol. 27 No. 4, pp. 117-158.

Rathee, N. and Sarkar, C. (2020), "COVID-19 lockdown: how the pandemic bringing change in Indian education system", *Impact of COVID-19 and Pandemic Lockdown in India*, Eureka Publications, Balewadi.

Rockart, J.F., Earl, M.J. and Ross, J.W. (1996), "Eight imperatives for the new IT organization", *Inventing the Organizations of the 21st Century*, pp. 1-30.

Schiano, B. (2021), "Making the most of the hybrid classroom", available at: https://hbsp.harvard.edu/inspiring-minds/making-the-most-of-the-hybrid-classroom (accessed 12 April 2022).

Sharma, A. and Alvi, I. (2021), "Evaluating pre and post COVID 19 learning: an empirical study of learners' perception in higher education", *Education and Information Technologies*, Vol. 26, pp. 1-18.

Sharma, D. and Singh, V. (2010), "ICT infrastructure and human resource performance–A study of universities in the western Himalayan region of India", *International Journal of Advanced Engineering and Application*, Vol. 7 No. 1, pp. 316-321.

Sheth, H. (2020), "India jumps five ranks to 35th in worldwide education for the future Index 2019", *@businessline*, available at: https://www.thehindubusinessline.com/news/education/india-jumps-5-ranks-to-35th-in-worldwide-education-for-the-future-index-2019/article30851541.ece> (accessed 14 September 2021).

Shi-Chun, D., Ze-Tian, F. and Yi, W. (2014), "The flipped classroom–advantages and challenges", *2014 International Conference on Economic Management and Trade Cooperation (EMTC 2014)*, Atlantis Press.

Singh, R. (2016), "How a bunch of Indian Startups have taken an early lead in Chatbots", *The Economic Times*, available at: http://economictimes .indiatimes.com/small-biz/startups/how-a-

bunch-of-indian-startups-have-taken-an-early-lead-in-chatbots/articleshow/52592715.cms (accessed 13 October 2016).

Singh, M.K. and Gupta, V. (2020), "Critical types of knowledge loss in military organisations", *VINE Journal of Information and Knowledge Management Systems*, Vol. 51 No. 4, pp. 618-635.

Singh, M.K. and Gupta, V. (2021), "IAF transition towards a network-centric-knowledge based force", *Journal of the United Service Institution of India*, Vol. CLI No. 623, pp. 93-100.

Singh, M.K. and Gupta, V. (2022), "An empirical study of knowledge environment and suitability of performance measures of a civil organization for a knowledge-based military force", *Kybernetes*, Vol. ahead-of-print No. ahead-of-print, In press.

Sinha, N., Kakkar, N.K. and Gupta, V. (2009), "Unleash the power of creativity and innovation", *International Journal of Sustainable Strategic Management*, Vol. 1, pp. 417-433.

Sinha, N., Kakkar, N.K. and Gupta, V. (2012), "Uncovering the secrets of the twenty-first-century organization", *Global Business and Organizational Excellence*, Vol. 31, pp. 49-56.

Sinha, N., Kakkar, N.K. and Gupta, V. (2015), "Harnessing the power of knowledge management for innovation", *International Journal of Knowledge and Learning*, Vol. 10 No. 2, pp. 124-146.

Soong, M.H., Chan, H., Chua, B. and Loh, K. (2001), "Critical success factors for on-line course resources", *Computers and Education*, Vol. 36 No. 2, pp. 101-120.

Spataro, J. (2020), "2 years of digital transformation have happened in two months", available at: https://www.microsoft.com/en-us/microsoft-365/blog/2020/04/30/2-years-digital-transformation-2-months/ (accessed 10 June 2022).

Stella Maris College (2021), "Report on UGC - PARAMARSH scheme", Chennai, pp. 1-27, available at: https://stellamariscollege.edu.in/assets/documents/Paramarsh%202021.pdf.

Study in India (2021), *Education*, Ministry of Education, Government of India, India, available at: https://www.studyinindia.gov.in (2022).

Sukula, S.K., Awasthi, S. and Dwivedi, A.K. (2021), "Transcending learning and information access by academic libraries in the perspective of digital accessibility and policies in higher education during covid-19 in India", *International Journal of Research in Library Science*, Vol. 7 No. 1, pp. 1-14.

Teräs, M., Suoranta, J., Teräs, H. and Curcher, M. (2020), "Post-Covid-19 education and education technology 'solutionism': a seller's market", *Postdigital Science and Education*, Vol. 2 No. 3, pp. 863-878.

Thomas, P.Y. (2011), "Cloud computing: a potential paradigm for practising the scholarship of teaching and learning", *The Electronic Library*, Vol. 29 No. 2, pp. 214-224.

Thomas, A. (2021), "Business beyond COVID-19: towards open innovation", *Globalization, Deglobalization, and New Paradigms in Business*, Palgrave Macmillan, Cham, pp. 189-212.

Thomas, A. and Chopra, M. (2020), "On how big data revolutionizes knowledge management", *Digital Transformation in Business and Society*, Palgrave Macmillan, Cham, pp. 39-60, Ftec.

Thomas, A. and Gupta, V. (2022a), "Tacit knowledge in organizations: bibliometrics and a framework-based systematic review of antecedents, outcomes, theories, methods and future directions", *Journal of Knowledge Management*, Vol. 26 No. 4, pp. 1014-1041, doi: 10.1108/JKM-01-2021-0026.

Thomas, A. and Gupta, V. (2022b), "The role of motivation theories in knowledge sharing: an integrative theoretical reviews and future research agenda", *Kybernetes*, Vol. 51 No. 1, pp. 116-140, doi: 10.1108/K-07-2020-0465.

Thomas, A. and Paul, J. (2019), "Knowledge transfer and innovation through university-industry partnership: an integrated theoretical view", *Knowledge Management Research and Practice*, Vol. 17 No. 4, pp. 436-448.

Thomas, A., Cillo, V., Caggiano, V. and Vrontis, D. (2020), "Drivers of social capital in enhancing team knowledge sharing and team performance: moderator role of manager's cultural intelligence", *International Journal of Managerial and Financial Accounting*, Vol. 12 Nos 3-4, pp. 284-303.

UGC (2019), "Paramarsh - UGC scheme for mentoring NAAC accreditation aspirant institutions to promote quality assurance in higher education", pp. 1-12, available at: https://www.ugc.ac.in/pdfnews/7009610_PARAMARSH.pdf.

UGC (2021), "Induction and mentorship for teacher of higher education guidelines about improve quality of service, area, key issues, foundation etc, New Delhi", pp. 1-48, available at: https://www.ugc.ac.in/pdfnews/3830777_Mentorship-Guidelines-.pdf.

UNESCO (2003), *Communiqué of the Ministerial Roundtable on 'Towards Knowledge Societies'*, UNESCO, Paris.

Venkataraman, S. (2004), "Regional transformation through technological entrepreneurship", *Journal of Business Venturing*, Vol. 19 No. 1, pp. 153-167.

Vrontis, D., Christofi, M., Battisti, E. and Graziano, E.A. (2021), "Intellectual capital, knowledge sharing and equity crowdfunding", *Journal of Intellectual Capital*, Vol. 22 No. 1, pp. 95-121.

Wade, M. and Hulland, J. (2004), "The resource-based view and information systems research: review, extension, and suggestions for future research", *MIS Quarterly*, Vol. 28 No. 1, pp. 107-142.

Waheed, A. and Zhang, Q. (2020), "Effect of CSR and ethical practices on sustainable competitive performance: a case of emerging markets from stakeholder theory perspective", *Journal of Business Ethics*, Vol. 175, pp. 837-855, doi: 10.1007/s10551-020-04679-y.

Walker, S. (2018), "3 futuristic technologies to support blended learning: artificial intelligence, virtual reality, and augmented reality", Written by Sherri Walker [Blog], available at: https://www.imaginelearning.com/blog/2018/10/3-futuristic-technologies-support-blended-learning-artificial-intelligence-virtual.

Walmiki, R.H. and Ramakrishnegowda, K.C. (2009), "ICT infrastructure in university libraries of Karnataka", *Annals of Library and Information Studies*, Vol. 56, pp. 236-241.

Wang, Q. (2008), "A generic model for guiding the integration of ICT into teaching and learning", *Innovations in Education and Teaching International*, Vol. 45 No. 4, pp. 411-419.

Wang, C., Zhang, Q. and Zhang, W. (2020), "Corporate social responsibility, green supply chain management and firm performance: the moderating role of big-data analytics capability", *Research in Transportation Business and Management*, Vol. 37, p. 100557.

Wenger, E. (2011), *Communities of Practice: A Brief Introduction*, National Science Foundation (US), pp. 1-5.

Wernerfelt, B. (1989), "From critical resources to corporate strategy", *Journal of General Management*, Vol. 14, pp. 4-12.

Zou, C., Zhao, W. and Siau, K. (2020), "COVID-19 pandemic: a usability study on platforms to support eLearning", *International Conference on Human-Computer Interaction*, Cham, Springer, pp. 333-340.

## Further reading

El Said, G.R. (2021), "How did the COVID-19 pandemic affect higher education learning experience? An empirical investigation of learners' academic performance at a university in a developing country", *Advances in Human-Computer Interaction*, Vol. 2021, doi: 10.1155/2021/6649524.

Thomas, A., Gupta, V., Riso, T., Briamonte, M.F., Usai, A. and Fiano, F. (2021), "Enhancing innovative behavior at the workplace: the moderating role of entrepreneurial orientation and web 2.0", *2021 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)*, IEEE, pp. 1-8.

## About the authors

Namita Jain has a PhD from Delhi Technological University and is working as an assistant professor in Shri Ram College of Commerce, University of Delhi. She has teaching experience of more than ten years. She has published research papers related to commerce and management in Scopus, SSCI and ABDC

listed Journals and participated in National and International Conferences. Her research and teaching interests include Knowledge management, Information Technology and Business Management.

Asha Thomas is an assistant professor at the Department of Operations Research and Business Intelligence, Wroclaw University of Science and Technology, Poland. Her Ph.D. is in knowledge management. Her areas of research interest include knowledge management, open innovation, big data, grass-root innovation and SMEs. She has about 12 years of experience in teaching and over three years of experience in the IT and Telecom industries. She has several research papers published in top journals, including the "Journal of Knowledge Management", "Knowledge Management Research and Practices", and "Tourism Management Perspectives". She has published 15 books at a national level and book chapters with publishers like "Palgrave Macmillan" and "Springer".

Dr Vikas Gupta is a Gold Medalist (MBA) and has more than 20 years of experience. He is currently working as an Assistant Professor, Delhi School of Management, Delhi Technological University (Former Delhi College of Engineering). He has many research papers and articles in his credit which are published in Scopus, SCI, SSCI and ABDC listed Journals.

His research interests are in the area of knowledge management, innovation and corporate social responsibility. He has been awarded the Best Faculty Award in the area of management. He has also been associated with many consultancy projects. Vikas Gupta is the corresponding author and can be contacted at: vikasguptadtu@gmail.com

Mario Ossorio is an assistant professor in Management at Università degli Studi della Campania "Luigi Vanvitelli", where he received his PhD in Entrepreneurship and Innovation. He earned his MSc in Economics and Finance from Università di Napoli "Federico II". His primary research interests include innovation, internationalization, family business and mergers and acquisitions. Recently he published in *Management Decision, Journal of Management and Governance, International Journal of Managerial and Financial Accounting* and *EuroMed Journal of Business*.

Daniele Porcheddu is an associate professor of marketing at the department of economics and business at the University of Sassari. His research interests currently are focused on consumers' perceptual processes in retail environment. His papers have been published in journals such as the Journal of Retailing, Journal of Advertising Research, Psychology of Consciousness and Frontiers in Psychology.

For instructions on how to order reprints of this article, please visit our website:
**www.emeraldgrouppublishing.com/licensing/reprints.htm**
Or contact us for further details: **permissions@emeraldinsight.com**

# Sustainable Reservoir Sediment Management in Indian River Basins

(*By Amrendra Kumar Singh and Anuj Kanwal, Pub. in Water & Energy International, Oct.2021,A-172, pp.6-13*)

**DISCUSSOR PROF. S.K. MAZUMDER**[1]

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 14.139.251.198 on Wed, 15-Jun-2022

Preserving the existing reservoirs and increasing its useful life is essential for sustainable development in order to meet the future water requirements of the country. In a vision paper on reservoir sedimentation , Schleiss et al [1] stated "*an inadequate sediment management can greatly accelerate the process of eutrophication and organisms dissemination at levels which compromise the functions of the reservoir. Dam construction creates an extremely efficient sink of sediments in the valley. Without efficient sediment management, it strongly alters the natural equilibrium between sediment dynamics and morphology in a river catchment. Over the years, as the sediments accumulate, the reservoir loses the storage capacity for which it was initially designed. The installed water storage capacity of reservoirs world wide is about 7000 km³, from which some 4000 km³ are used for energy production, irrigation, and water supply . The mean age of existing reservoirs is between 30 and 40 years, and it is estimated that 0.5–1% of the worldwide water storage capacity is lost annually due to sedimentation".*

Accurate estimation of sedimentation rate is essential for the realistic determination of useful life of the reservoir. In this paper authors have made commendable efforts in collecting sediment data in different basins all over the country. As rightly observed by the authors *"The old assumption that the average annual decrease in the Dead Storage of the Dam by 1% at the design stage has been negated in practice with Siltation entering in the Live storages"*. Variation of sediment deposition along depth of reservoir, as shown in Fig.1, depends upon type of reservoir [2] .

In a paper, [3] Deepa et. al predicted the life of Gobindsagar/Bhakra reservoir, constructed on the river Sutlej in Himachal Pradesh. Time-series analysis of the sedimentation rate from a long-term data for the period 1958–2003 obtained from hydrographic survey showed an increasing trend in annual sedimentation rate and reduction of remaining life of the reservoir. The useful life of Gobindsagar reservoir was computed as 300 years in contrast with 500 years estimated at the planning stage. Economic life of the reservoir with accelerated sedimentation was found to be about 165 years only [4]. Actual depth wise distribution of sediments deposited in the Bhakra and some other reservoirs is shown in Fig.1[5]. It is apparent from the figure that the full life of live storage of the reservoir as well as its useful life ( at 50% loss of live storage) are reduced considerably due to sediment deposition in live and surcharge space.

CWC [6] undertook a coordinated research scheme of reservoir sedimentation across the country. Sedimentation rate in the major reservoirs (storage capacity >100 MCM) which completed 50 years of useful life was found to vary from 0.3 to 4.89 ha-m/100 km²/year and those reservoirs which have served less than 50 years of useful life, it varied from 0.34 to 27.85

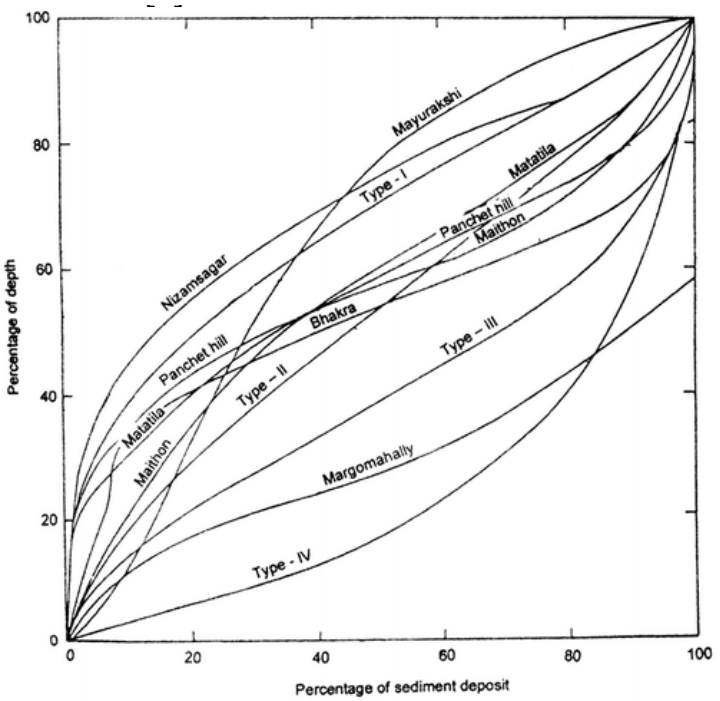



**Fig. 1** : Boreland-Miller Curves for Sediment Distribution in Different Types (I,II,III &IV) of Reservoirs (Note: Comparison made with Some Important Reservoirs in India)

1. Former AICTE Emeritus Professor of Civil Engineering, Delhi Technology University (Former Delhi College of Engineering)

ha-m/100 km²/year. With sedimentation , as the reservoir capacity reduces, the capacity- inflow ratio and the trap efficiency [7,8] decreases (Fig.2) resulting in reduction of sedimentation rate thereby prolonging full life of reservoir up to dam crest. Table-1 shows Sedimentation rate and expected useful life which was found to vary region-wise.

The results of useful life computation of Bhakra /Govindsagar reservoir by different methods are presented in Table 2. The useful life estimated using different criteria is significantly different. Using average sedimentation rate, the useful life can approximately be taken as 140 years. According to Bhakra Management Board, the time period for 100% depletion of dead storage is 260 years which is much higher. The discrepancy may be due to the assumptions made during the planning stage that entire fine silt including clay i.e. 10% of suspended silt load and 5% of the remaining silt load will be passing over the spillway during flood and through dam sluice outlets as density currents. With C/I value of 0.59, the trap efficiency of the reservoir is nearly 99% which makes the assumption invalid. Most of the suspended silt load is trapped in the live and dead storage zones, reducing the useful/economic life of the reservoir.

The rate of sedimentation for reservoirs in the Himalayan region (Indus, Ganga, Brahmaputra basin) was found to vary from 5.658 to 27.85 ha-m/100 km²/year. Based on the different approaches, the Bhakra Project Authority computed the useful life as shown in Table 2

Using Churchil's sedimentation index method, Mazumder [11], studied the life of a Himalayan reservoir in Sikim province and determined the life by area incremental method to find the loss of reservoir capacity at an interval of every 6 years (Fig. 3). Sedimentation rate was found to be decreasing with time.

Satellite Remote Sensing (SRS) technique was used by Santosh Wagh et.al [12] and Pandey,et.al [13] to estimate spatial, temporal, and spectral attributes for evaluating the sediment distribution pattern . The main advantage of satellite data over the conventional procedure includes its repetitive coverage over a long period. The results obtained by remote sensing survey are relevant with the hydrographical survey. The remote sensing technique is time and cost-effective for the development of revised elevation–area–capacity curves for a reservoir with the objective of determining the loss of reservoir space due to sedimentation.

MIKE-3 hydrodynamic model, a commonly used mathematical model system, was applied to simulate cohesive sediments, flows, water quality, and ecology in lakes, rivers, estuaries, bays, coastal areas, and seas in three dimensions[14] . The model was developed to estimate the

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 8 on dated 17-Aug-2022

**Table 1** : Sedimentation rate of some of the Indian reservoirs

| Sl. No | Name of reservoir | Gross storage capacity (M-m³) & (year) | | Average sedimentation rate (%) | Seriousness of sedimentation | Loss of gross capacity till 2014 (%) | Calendar year in which useful life shall be |
|---|---|---|---|---|---|---|---|
| | | Initial survey | Last survey | | | | |
| 1 | Matatila | 1133 (1956) | 764 (1994) | 0.86 | Serious | 49.76 | 2014 |
| 2 | Hirakud | 8105 (1957) | 6146 (1994) | 0.65 | Serious | 37.24 | 2034 |
| 3 | Maithon | 1349 (1955) | 1085 (1994) | 0.50 | Serious | 29.60 | 2055 |
| 4 | Tungabhadra | 3751 (1953) | 3158 (1993) | 0.40 | Significant | 24.13 | 2078 |
| 5 | Koyna | 2798 (1961) | 2779 (1986) | 0.03 | Insignificant | 2.73 | 3628 |

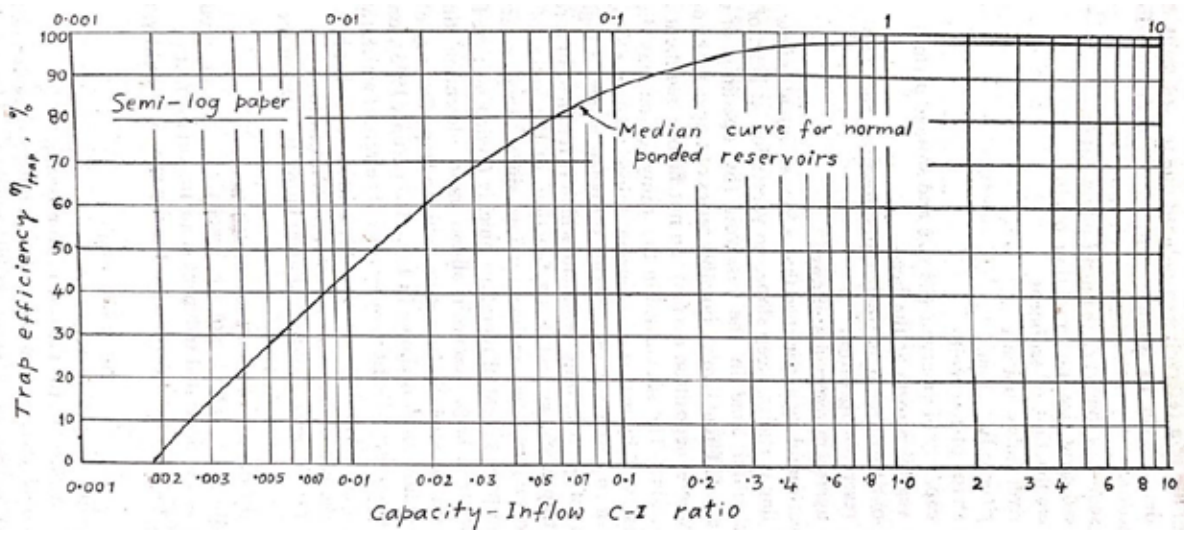



**Fig. 2 :** Variation of Trap Efficiency with Capacity- Inflow Ratio

**Table 2 :** Useful Life of Gavindsagar/Bhakra Reservoir Differerrt Approaches [9,10]

| Data & period | Methodology | Approach & criteria | Useful life (yrs) | Calendar year in which useful life shall be over |
|---|---|---|---|---|
| Sedimentation rate based on Hydrographic survey data (1958 & 1998) | Extrapolation | 100% depletion of dead storage capacity | 138 | 2096 |
| | | @ 17.62 M-m$^3$/year | 109 | 2067 |
| | | 50% depletion of gross storage capacity@ 34.76 M-m$^3$/year | 142 | 2100 |
| Sedimentation rate based on Hydrographic survey (1958–2003) | Extrapolation | Time-series analysis using increasing trend in annual sedimentation | 127 | 2085 |
| Annual sedimentation inflow observation (1958–1998) | Gill's approach | 50% depletion of gross storage capacity assuming a dominant coarse sediment inflow | 143 | 2101 |
| Annual sedimentation inflow observation during planning stage (prior to 1957) | Brune's step method | 50% depletion of gross storage capacity considering variation in trap efficiency with time | 150 | 2108 |
| Design criteria employed during planning stage (prior to 1957) | Analogy of Lake Mead and Elephant Butte reservoirs | 100% depletion of dead storage capacity based on the advance of delta formation towards the dam | 260 | 2218 |

www.IndianJournals.com
Members Copy, Not for Commercial Sale

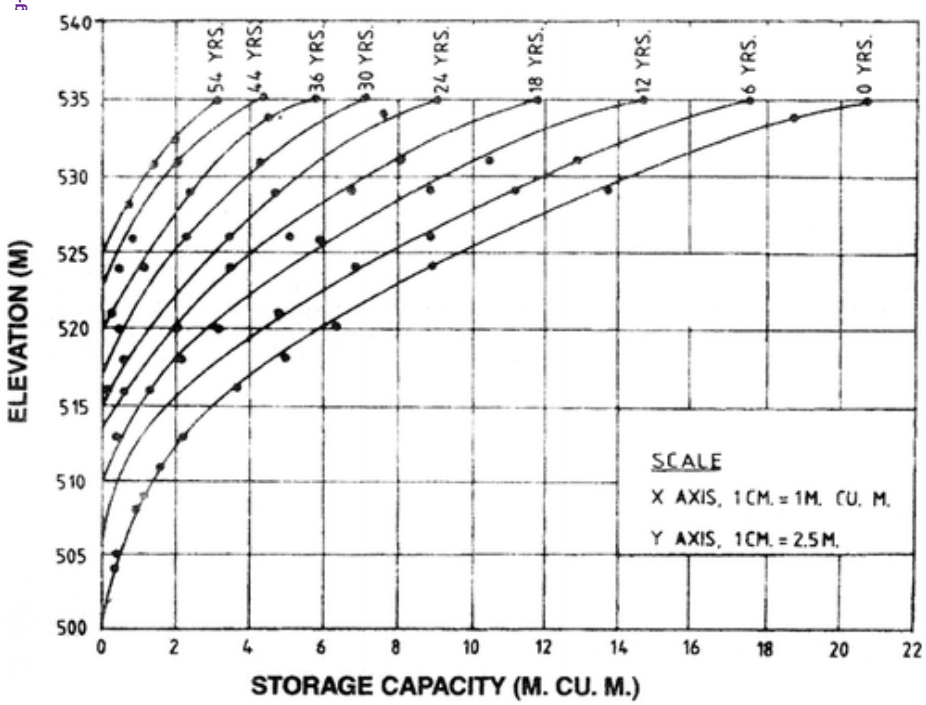



**Fig. 3** : Loss of Storage Capacity in Kirichu Dam in Sikim (Using Churchil's Method)

sedimentation in the Xiaolangdi Reservoir on the Yellow River, China. The flow and sedimentation were well estimated by the model. The model results revealed that the sedimentation elevation is lower than the permissible value.

***Discusser wishes to bring to the notice of the authors following points stated clause wise below:***

## Clause 2 :

- Statement regarding 2% annual storage loss has not been established
- The following statement "*The Change in the Hydraulics and storages for flood cushion which may lead to Overtopping, or dam disasters. Sediment Induced hydraulic wave and induced seismicity, in which weight increase by silt layers add stress to both sub-strata and the power-dam itself*".

  Silt pressure and Hydro-dynamic pressure due to earthquake are considered in stability analysis/ stress analysis of rigid dams like gravity dams. It may be explained how sediment content in water can further aggravate the situation.

## Clause 3 :

- Apart from the amount of sediment concentration, Sediment deposited in a reservoir depends also on trap efficiency governed by C/I ratio and type of reservoir
- Sediment delivery ratio (not rate) depends on sediment yield of basin governed by several basin parameters and reservoir characteristics
- No link has been established between urbanization and sedimentation of reservoirs as mentioned by the authors in the abstract.

References given within the body of the manuscript should have been listed at the end as per CBIP norms.

The value of the paper would have further increased if authors could add a Para or so regarding sediment removal /flushing

from reservoirs to extend their useful life.Fig. 4 shows Sediment Deposit in Lewis and Clark Lake - reservoir created by Gavins Point Dam on the Missouri River. It's 30 percent full of sediment and could be half full by 2045 [15] . Discusser mentioned few methods of sediment removal during his talk in ISWEE 16].

## REFERENCES


1. Schleiss A.J.,Mário J. F, Carmelo J. and Giovanni D. C.(2016) "Vision Paper" pub. online in the J. of Hyd. Research, IAHR, (*Laboratory of Hydraulic Constructions, École Polytechnique Fédérale de Lausanne,Lausanne, Switzerland* -http://www.tandfonline.com)
2. Borland, W.M. and Millar,C.R. (1958) "Distribution of Sediment in Large Reservoirs", J. of Hyd. Divn., Proc. ASCE, Vol.84, No.HY2,Pt.1,pp.1587-1 to 9
3. Deepa S. K., U.C. Chaube and A. Pandey (2015 ) "Revisiting the Useful Life Computation of Gobindsagar (Bhakra) Reservoir" (2015), ISH Journal of Hydraulic Engineering, Vol.22 No.2 ,pp. 124-126
4. Mazumder, S. K. (2016) "Discussion of Revisiting the useful life computation of Gobindsagar Bhakra reservoir by Deepa S Koomullil U C Chaube and Ashish Pandey" pub. in ISH J. of HYD Engineering Vol. 22, No. 2, PP. 124-126.
5. Garde, R.J., and Kothyari, U.C. (1987). "Estimation of sediment yield." J. Irrig.and Power, Central Board of Irrigation and Power (CBIP), India, 44(3), 97–121
6. CWC. (2001). *Compendium of silting of reservoirs in India 2001*, Central Water mmission (CWC) Publication, Government of India, New Delhi
7. Murthy, B.N. (1977). *Life of reservoir*. Technical report no. 19, Central Board of Irrigation and Power (CBIP), New Delhi
8. Raghunath,H.M.(1985) "Hydrology" , A Text book pub. by Wiley Eastern Ltd., New Delhi
9. Jothiprakash, V., and Garg, V. (2009). "Reservoir sedimentation estimation using artificial neural network." *J. Hydrologic Eng.*, ASCE, vo.14, no.4,pp.1035-40
10. Garg, V., and Jothiprakash, V. (2012). "Sediment yield assessment of a large basin using PSIAC approach in GIS environment." *Water Resour. Manage.*, 26(3), pp.799–840
11. Mazumder, S.K., "Life of a Reservoir in the Himalayan Region" Paper pub. In CBI & P J. of Water and Engergy International, Vol. 57, No. 8, July-Sep.2000
12. Santosh Wagh and Vivek Manekar "Assessment of Reservoir Sedimentation using Satellite Remote Sensing Technique (SRS)" J. of The Institution of Engineers (India), CE Divn. May 2021
13. A. Pandey, U.C. Chaube, S.K. Mishra, D. Kumar (2016), Assessment of reservoir sedimentation using remote sensing and recommendation for desilting patratu reservoir India. J. Hydrol Sci. J. https://doi.org/10.1080/02626667.2014.993988
14. Zhang, J., Wang, Y., and Zhang, H. (2002). "Numerical simulation of deposition process of the Xiaolangdi Reservoir in the early stage of operation." J. Hydraul. Eng., Vol. 7, No.7, pp.110–115.
15. 16. Morris, G.L., and Fan, J. (1998). "Sediment deposit in reservoirs." *Reservoir sedimentation handbook*, L. Hager and P. Lamb, eds., McGraw-Hill Book, New York, NY,
16. Mazumder,S.K.(2020) "Sediment Problems in Rivers and Reservoirs"Invited Audio- Video Talk Deivered in the International Symposium on Water,Ecology and Environment (ISWEE-20) Organised by Jiatong University, Bejing, China During 6-8 Dec.2020 in 5th Session on ˜Hydrology and Water Resources.


www.IndianJournals.com
Members Copy, Not for Commercial Sale
m IP - 14.139.251.98 on dated 1-Aug-2022

**Fig. 4** : Sediment Depostion in Rreservoir Behind Gavins Point Dam on the Missouri River

Journal of Molecular Structure 1269 (2022) 133780

Contents lists available at ScienceDirect

Journal of Molecular Structure

journal homepage: www.elsevier.com/locate/molstr

# Synthesis, structural characterizations, electrochemical properties and DFT calculations of highly fluorescent phosphorus(V) corroles

Ritika Kubba[a], Omprakash Yadav[b], Suman Maji[c], Natalia Fridman[d], Anil Kumar[a,*]

[a] Department of Applied Chemistry, Delhi Technological University, 42, Delhi, India
[b] Department of Chemistry, ARSD College, University of Delhi, Delhi, India
[c] School of Chemical Engineering and Physical Sciences, Lovely Professional University, Punjab, India
[d] Schulich Faculty of Chemistry, Technion-Israel Institute of Technology, Haifa, Israel

ARTICLE INFO

*Article history:*
Received 11 May 2022
Revised 13 July 2022
Accepted 21 July 2022
Available online 21 July 2022

*Keywords:*
Corrole
Phosphorus
Electrochemistry
DFT
Fluorescence

ABSTRACT

The synthesis, characterizations, electrochemical properties, single crystal X-ray diffraction studies and DFT calculations of highly fluorescent phosphorus corroles are reported. Nine phosphorus corroles have been synthesised by the reaction of metal-free corroles with $PCl_3$ at reflux in pyridine. Synthesised phosphorus (V) corrole complexes have been well characterized with the aid of UV-visible spectroscopy, HRMS, $^1$H NMR, $^{31}$P NMR and single crystal X-ray diffraction. Their absorption spectra depicted that phosphorus corroles show a strong band at 600 nm which indeed could be beneficial in biological applications. Exclusive fluorescent features of synthesised phosphorus corroles have been made evident by photophysical studies. To get insight into more structural information, single crystal X-ray diffraction of four complexes have been carried out. Detailed discussion on the electrochemistry of phosphorus (V) corroles has been incorporated showing the effect of *meso*-phenyl substitution on the oxidation and reduction potentials of phosphorus (V) corrole complexes. DFT calculations have also been employed, highlighting the HOMO-LUMO gap variation with the *meso*-carbon substituents in newly synthesised nine phosphorus corroles.

© 2022 Elsevier B.V. All rights reserved.

## 1. Introduction

Phosphorus corroles are relatively less explored but when it comes to the subject of metallocorroles, they prove out to be of great significance. Heading to the discovery of first corrole by Johnson and Kay in 1964 [1], it took a reasonably long time to reach out to the main group metal complexes of corrole [2–9]. After about 34 years, synthesis of first phosphorus corrole was reported by Paolesse and co-workers in 1998 [7]. Gross research group and Kadish research group [8,9] then further took the lead on carrying forward the research in the field of phosphorus corrole chemistry [10]. Beginning from the synthesis to detailed characterization (including UV-VIS, NMR, HRMS and single crystal X-ray diffraction) to photophysical [11,12] and electrochemical studies [12–15] to biological applications [16,17], phosphorus corroles have been dwelled upon all the aforementioned aspects and have also been highlighted in a recent review article by Di Natale et al. [18]. Thanks to its high fluorescent quantum yields and electrochemical versatility, phosphorus corroles have attracted a lot of limelight in the past few years. On the one hand axial ligand dissociation of $P^v$ corrole complexes grabbed a lot of importance [12,19], on the other hand reports on fluorescence properties [20,21], electrochemistry and density functional theory (DFT) calculations [13,22–25] are equally in quested. Due to their intrinsically strong emission and electrochemical features, phosphorus corroles have been extensively utilized in numerous applications including cancer cell imaging [16], photocatalysis [26,27], photodynamic therapy (PDT) [28–31] and photo electrochemistry [32] and many more. It was found that many accounts based on $A_3$ type P(V) corroles have been reported, leaving the ABA type corroles unexplored. Where $A_3$ type refers to the corroles having same *meso* substituent group at $C_5$, $C_{10}$ and $C_{15}$ positions, while ABA type corresponds to those having same *meso* substituent group at $C_5$ and $C_{15}$ positions and different *meso* substituent group at $C_{10}$ position. We wish to widen the studies focussed on ABA type corroles rendering variety of exciting features and in turn various applications. Here, we have synthesized nine ABA type $P^v$ corroles aiming at further apprehending the field of phosphorus corroles. Each corrole complex is bis-hydroxyl ligated where $C_5$ and $C_{15}$ *meso* positions are p-nitrophenyl substituted while $C_{10}$ position has either electron releasing or electron withdrawing substituent. In this article, we report the synthesis, structural characterizations, photophysical and electrochemical

* Corresponding author.
*E-mail address:* anil_kumar@dce.ac.in (A. Kumar).

https://doi.org/10.1016/j.molstruc.2022.133780
0022-2860/© 2022 Elsevier B.V. All rights reserved.

properties, single crystal X-ray structures and density functional theory (DFT) of a series of phosphorus corroles. In essence, we wish to present a detailed explanation of numerous properties of well characterized series of ABA $P^v$ corroles, in turn opening the door wide-open for innumerable applications for future.

## 2. Experimental section

### 2.1. Instrumentation

#### 2.1.1. Nuclear magnetic resonance spectroscopy

Jeol Model: JNM-EXCP 400 operating at 400 MHz spectrometer and Bruker Avance III 400 MHz operating at 400.4 MHz was used for obtaining $^1$H NMR data and Bruker Avance III 400 spectrometer with automated tuning, operating at 162.08 MHz was used for $^{31}$P NMR data.

Chemical shift values for hydrogens are reported with respect to residual solvent peaks at 7.26 ppm for $CDCl_3$.

*UV–Vis spectra:* UV-1800 Shimadzu spectrophotometer was used for recording UV–Vis absorption spectra. Calculations of molar absorptivity coefficients of phosphorus corroles were done in dichloromethane (DCM).

*High resolution mass spectrometry:* Bruker Maxis Impact mass spectrometer was used for recording high-resolution mass spectra, and APCI (atmospheric pressure charge ionization) was used as a direct probe in either positive or negative mode.

*Cyclic voltammetry:* Cyclic voltammetry studies were carried out by using Autolab Potentiostat Galvanostat (AUT204, Netherlands) supported by its software, NOVA 2.1. Three electrode system *viz.* glassy carbon working electrode, platinum wire counter electrode and an Ag/AgCl reference electrode were used for obtaining the cyclic voltammograms. Acetonitrile (HPLC) was utilized as solvent for all the CV recordings with 0.1 M tetrabutylammonium perchlorate (TBAP) as the supporting electrolyte.

*Fluorescence techniques:* Emission quantum yield and life time investigations were performed on Horiba Jobin Yvon, Fluorolog-3 spectrofluorometer at room temperature in toluene. The fluorescence quantum yields ($\Phi$) of compounds **1P-9P** were evaluated from their absorption and emission spectra by employing the comparative method [33] with tetraphenylporphyrin (TPP) as a standard ($\Phi_{tol}$ =0.13) [34].

*X-ray crystallography:* Single crystals of $C_{37}H_{21}F_2N_6O_6P$ (**2P**), $C_{37}H_{21}Cl_2N_6O_6P$ (**5P**), $C_{37}H_{19}Cl_2F_2N_6O_4P$ (**5P($F_2$)**) and $C_{40}H_{27}N_6O_8P$ (**6P**) were grown by slow evaporation from dichloromethane/heptane solutions. An appropriate crystal of (**2P**), (**5P**) and (**5P($F_2$)**) was selected and mounted on a diffractometer (Bruker ApexII) and an appropriate crystal of (**6P**) was selected and mounted on a diffractometer (Rigaku Oxford Diffraction) at 200 K. Data collection was performed using monochromated Mo K$\alpha$ radiation, $\lambda$ = 0.71073 Å, using $\varphi$ and $\omega$ scans to cover the Ewald sphere. Accurate cell parameters were obtained with the amount of indicated reflections. Using Olex2 [35], the structure was solved with the olex2.solve [36] structure solution program using Charge Flipping and SHELXL [37] refinement package was used for refinement using the Least Squares minimization. All non-hydrogen atoms were refined with anisotropic displacement parameters. The hydrogen atoms were refined isotropically on calculated positions using a riding model with their $U_{iso}$ values constrained to 1.5 times the $U_{eq}$ of their pivot atoms for terminal $sp^3$ carbon atoms and 1.2 times for all other carbon atoms.

**Crystal Data for** (**2P**) $C_{75}H_{44}Cl_2F_4N_{12}O_{12}P_2$ ($M$ = 1514.06 g/mol): triclinic, space group P-1 (no. 2), $a$ = 7.6887(14) Å, $b$ = 11.136(2) Å, $c$ = 19.999(4) Å, $\alpha$ = 77.978(4)°, $\beta$ = 79.640(4)°, $\gamma$ = 75.752(5)°, $V$ = 1608.2(5) Å$^3$, $Z$ = 1, $T$ = 200.15 K, $\mu$(MoK$\alpha$) = 0.242 mm$^{-1}$, *Dcalc* = 1.563 g/cm$^3$, 12,189 reflections measured (3.83° ≤ 2Θ ≤ 50.216°), 4572 unique ($R_{int}$ = 0.0767, $R_{sigma}$ = 0.1484) which were used in all calculations. The final $R_1$ was 0.0723 ($I > 2\sigma(I)$) and $wR_2$ was 0.2199 (all data). **CCDC No: 2162132**.

**Crystal Data for** (**5P**) $C_{75}H_{44}Cl_6N_{12}O_{12}P_2$ ($M$ = 1579.86 g/mol): triclinic, space group P-1 (no. 2), $a$ = 7.939(3) Å, $b$ = 10.828(4) Å, $c$ = 20.525(7) Å, $\alpha$ = 95.961(9)°, $\beta$ = 100.316(9)°, $\gamma$ = 106.973(10)°, $V$ = 1637.2(10) Å$^3$, $Z$ = 1, $T$ = 200.15 K, $\mu$(MoK$\alpha$) = 0.391 mm$^{-1}$, *Dcalc* = 1.602 g/cm$^3$, 5834 reflections measured (3.988° ≤ 2Θ ≤ 47.29°), 4261 unique ($R_{int}$ = 0.0428, $R_{sigma}$ = 0.1366) which were used in all calculations. The final $R_1$ was 0.0983 ($I > 2\sigma(I)$) and $wR_2$ was 0.2577 (all data). **CCDC No: 2162130**

**Crystal Data for (5P($F_2$))** $C_{75}H_{40}Cl_6F_4N_{12}O_8P_2$ ($M$ = 1587.83 g/mol): triclinic, space group P-1 (no. 2), $a$ = 7.9933(7) Å, $b$ = 10.7033(9) Å, $c$ = 20.5770(18) Å, $\alpha$ = 96.244(2)°, $\beta$ = 99.574 (3)°, $\gamma$ = 107.652(4)°, $V$ = 1630.3(2) Å$^3$, $Z$ = 1, $T$ = 200.15 K, $\mu$(MoK$\alpha$) = 0.397 mm$^{-1}$, *Dcalc* = 1.617 g/cm$^3$, 12,842 reflections measured (4.052° ≤ 2Θ ≤ 49.074°), 5331 unique ($R_{int}$ = 0.0985, $R_{sigma}$ = 0.0911) which were used in all calculations. The final $R_1$ was 0.0916 ($I > 2\sigma(I)$) and $wR_2$ was 0.2343 (all data). **CCDC No: 2162131**.

**Crystal Data for** (**6P**) $C_{52}H_{31}Cl_8N_6O_{12}P$ ($M$ = 1246.40 g/mol): triclinic, space group P-1 (no. 2), $a$ = 13.5168(3) Å, $b$ = 14.8591(7) Å, $c$ = 16.5130(6) Å, $\alpha$ = 64.859(4)°, $\beta$ = 75.736(3)°, $\gamma$ = 81.962(5)°, $V$ = 2907.5(2) Å$^3$, $Z$ = 2, $T$ = 200.15 K, $\mu$(MoK$\alpha$) = 0.479 mm$^{-1}$, *Dcalc* = 1.424 g/cm$^3$, 38,549 reflections measured (6.424° ≤ 2Θ ≤ 52.744°), 11,551 unique ($R_{int}$ = 0.0544, $R_{sigma}$ = 0.0783) which were used in all calculations. The final $R_1$ was 0.0930 ($I > 2\sigma(I)$) and $wR_2$ was 0.2924 (all data). **CCDC No: 2162133**

### 2.2. Materials and methods

*Chemicals:* All chemicals and solvents were used without further purification. For column chromatography of synthesised corroles, silica gel (100–200 µm mesh) was used. Pyrrole and substituted aldehydes (liquid) were distilled prior to use. Tetrabutylammonium perchlorate was obtained from Alfa Aesar. Deuterated chloroform ($CDCl_3$) was used as supplied from Sigma-Aldrich. Pyridine (99%) was obtained from Spectrochem Pvt. Ltd.

*Synthesis of free-base corroles:* 5,15-bis-nitrophenyl-10-(4-carboxymethylphenyl) free base corrole (**1**); 5,15-bis-nitrophenyl-10-(2,6-difluorophenyl) free base corrole (**2**); 5,15-bis-nitrophenyl-10-pentafluorophenyl) free base corrole (**3**); 5,15-bis-nitrophenyl-10-(3,4,5-trifluorophenyl) free base corrole (**4**); 5,15-Bis-nitrophenyl-10-(2,6-dichlorophenyl) free base corrole (**5**); 5,15-Bis-nitrophenyl-10-(3,4,5-trimethoxyphenyl) free base corrole (**6**); 5,15-Bis-nitrophenyl-10-(4-chlorophenyl) free base corrole (**7**); 5,15-Bis-nitrophenyl-10-(4-methoxyphenyl) free base corrole (**8**); 5,15-Bis-nitrophenyl-10-(4-bromophenyl) free base corrole (**9**) were synthesised as reported in the literature [38].

All the analytical data of compounds **1–9** matches with the previously reported work from our lab [39].

*Synthesis of phosphorus(V) corroles:* Nine phosphorus(V) ABA corrole complexes were synthesized using previously reported conventional method [13].

*Synthesis of 5,15-bis-nitrophenyl-10-(4-carboxymethylphenyl) phosphorus corrole, 1P:* A solution of 5,15-bis-nitrophenyl-10-(4-carboxymethylphenyl) free base corrole (**1**) (50 mg) in dry pyridine (10 mL) was prepared, to which 300 $\mu$L of $PCl_3$ (100 $\mu$L every 10 min) was added and the reaction mixture was stirred continuously. The color of the solution instantly switched from a deep green to reddish-pink with a greenish red tint. After the addition of required amount of $PCl_3$, 15–20 mL $H_2O$ was added into the reaction mixture followed by continuous stirring for about 10 min. Extraction of the reaction mixture was carried out by using dichloromethane ($CH_2Cl_2$) and was washed thoroughly with

distilled water. Desired product was obtained from the organic layer. Thin layer chromatography (TLC) and absorption spectroscopy were employed as an important tool in order to monitor the reaction progress. Solvent was removed from the product by using rotary evaporator. The solid thus obtained was subjected to column chromatography on silica gel and the required compound was collected from a greenish reddish-pink-colored band by eluting with $CHCl_3$. Using rotary evaporator, the solvent was removed in order to get a deep red-green solid, obtained in 65% yield. $^1$H NMR (400 MHz, $CDCl_3$): $\delta$ 9.37 (dd, $J$ = 4.4, 2.5 Hz, 2H), 9.02 (dd, $J$ = 7.3, 3.6 Hz, 2H), 8.99 (dd, $J$ = 9.2, 4.8 Hz, 2H), 8.85 (dd, $J$ = 4.9, 3.0 Hz, 2H), 4.09 (s, 3H); $^{31}$P NMR ($CDCl_3$): $\delta$: −100.56 ppm, $\delta$: −190.75 ppm. HRMS: m/z (%) observed: 736.2952 $[M]^+$, 719.2710 $[M-OH^-]^+$; calculated: 736.2700 $[M]^+$, 719.2700 $[M-OH^-]^+$. UV–vis ($CH_2Cl_2$) $\lambda_{max}$ ($\varepsilon$) [nm (× $10^4$ $cm^{-1}$ $M^{-1}$)]: 427 (2.45), 536 (0.25), 572 (0.30), 613 (0.70).

Likewise, compounds **2P-9P** were synthesised using similar synthetic route, as mentioned for compound **1P**.

Analytical details for compound **2P-9P** have been given below.

5,15-bis-nitrophenyl-10-(2,6-difluorophenyl) phosphorus corrole, **2P** The reddish-pink color solid was obtained in 60% yield. $^1$H NMR (400 MHz, $CDCl_3$): $\delta$ 9.34 (dd, $J$ = 4.4, 2.6 Hz, 2H), 9.03 (dd, $J$ = 7.2, 3.5 Hz, 2H), 8.96 (dd, $J$ = 5.1, 2.9 Hz, 2H), 8.81 (dd, $J$ = 7.4, 3.6 Hz, 2H); $^{31}$P NMR ($CDCl_3$): $\delta$: −100.64 ppm, $\delta$: −190.39 ppm. HRMS: m/z (%) observed: 697.2446 $[M-OH^-]^+$; calculated: 697.2700 $[M-OH^-]^+$. UV–vis ($CH_2Cl_2$) $\lambda_{max}$ ($\varepsilon$) [nm (× $10^4$ $cm^{-1}$ $M^{-1}$)]: 425 (3.10), 535 (0.35), 570 (0.40), 608 (0.85).

5,15-bis-nitrophenyl-10-pentafluorophenyl phosphorus corrole, **3P.** The reddish-pink color solid was obtained in 75% yield. $^1$H NMR (400 MHz, $CDCl_3$): $\delta$ 9.37 (dd, $J$ = 4.5, 2.5 Hz, 2H), 9.06 (dd, $J$ = 5.0, 3.6 Hz, 2H), 8.98 (dd, $J$ = 4.3, 3.6 Hz, 2H), 8.75 (dd, $J$ = 4.7, 3.0 Hz, 2H); $^{31}$P NMR ($CDCl_3$): $\delta$: −190.97 ppm. HRMS: m/z (%) observed: 751.2333 $[M-OH^-]^+$; calculated: 751.2500 $[M-OH^-]^+$. UV–vis ($CH_2Cl_2$) $\lambda_{max}$ ($\varepsilon$) [nm (× $10^4$ $cm^{-1}$ $M^{-1}$)]: 426 (2.60), 534 (0.30), 570 (0.35), 606 (0.60).

5,15-bis-nitrophenyl-10-(3,4,5-trifluorophenyl) phosphorus corrole, **4P.** The reddish-pink color solid was obtained in 58% yield. $^1$H NMR (400 MHz, $CDCl_3$) $\delta$ 9.37 (dd, $J$ = 4.5, 2.5 Hz, 2H), 9.04 (dd, $J$ = 5.0, 3.6 Hz, 2H), 8.99 (dd, $J$ = 4.5, 3.5 Hz, 2H), 8.85 (dd, $J$ = 5.0, 3.0 Hz, 2H); $^{31}$P NMR ($CDCl_3$): $\delta$: −100.69 ppm, $\delta$: −190.73 ppm. HRMS: m/z (%) observed: 713.8889 $[M-OH^-]^+$; calculated: 715.55 $[M-OH^-]^+$. UV–vis ($CH_2Cl_2$) $\lambda_{max}$ ($\varepsilon$) [nm (× $10^4$ $cm^{-1}$ $M^{-1}$)]: 425 (3.30), 536 (0.25), 573 (0.35), 611 (0.90).

5,15-Bis-nitrophenyl-10-(2,6-dichlorophenyl) phosphorus corrole, **5P.** The reddish-pink color solid was obtained in 45% yield. $^1$H NMR (400 MHz, $CDCl_3$) $\delta$ 9.34 (dd, $J$ = 4.5, 2.5 Hz, 2H), 9.01 (dd, $J$ = 4.9, 3.6 Hz, 2H), 8.96 (dd, $J$ = 4.2, 3.8 Hz, 2H), 8.65 (dd, $J$ = 5.0, 2.9 Hz, 2H); $^{31}$P NMR ($CDCl_3$): $\delta$: −100.70 ppm, $\delta$: −190.49 ppm. HRMS: m/z (%) observed: 722.2772 $[M- OH^-]^+$; calculated: 729.0600 $[M-OH^-]^+$. UV–vis ($CH_2Cl_2$) $\lambda_{max}$ ($\varepsilon$) [nm (× $10^4$ $cm^{-1}$ $M^{-1}$)]: 426 (2.65), 537 (0.20), 573 (0.35), 614 (0.70).

5,15-Bis-nitrophenyl-10-(3,4,5-trimethoxyphenyl) phosphorus corrole, **6P.** The reddish-pink color solid was obtained in 66% yield. $^1$H NMR (400 MHz, $CDCl_3$): $\delta$ 9.37 (dd, $J$ = 4.5, 2.5 Hz, 2H), 9.04 (dd, $J$ = 5.0, 3.6 Hz, 2H), 8.99 (dd, $J$ = 4.5, 3.5 Hz, 2H), 8.85 (dd, $J$ = 5.0, 3.0 Hz, 2H); $^{31}$P NMR ($CDCl_3$): $\delta$: −100.30 ppm, $\delta$: −190.84 ppm. HRMS: m/z (%) observed: 768.3442 $[M]^+$, 751.3227 $[M-OH^-]^+$; calculated: 768.3901 $[M]^+$, 751.3901 $[M-OH^-]^+$. UV–vis ($CH_2Cl_2$) $\lambda_{max}$ ($\varepsilon$) [nm (× $10^4$ $cm^{-1}$ $M^{-1}$)]: 426 (2.55), 536 (0.20), 572 (0.25), 609 (0.65).

5,15-Bis-nitrophenyl-10-(4-chlorophenyl) phosphorus corrole, **7P.** The reddish-pink color solid was obtained in 50% yield. $^1$H NMR (400 MHz, $CDCl_3$): $^1$H NMR (400 MHz, $CDCl_3$): $\delta$ 9.36 (dd, $J$ = 4.5, 2.5 Hz, 2H), 9.02 (dd, $J$ = 5.0, 3.6 Hz, 2H), 8.97 (dd, $J$ = 4.4, 3.5 Hz, 2H), 8.86 (dd, $J$ = 5.0, 3.0 Hz, 2H); $^{31}$P NMR ($CDCl_3$): $\delta$: −100.55 ppm, $\delta$: −190.79 ppm. HRMS: m/z (%) observed: 713.8967 $[M]^+$; calculated: 713.8067 $[M]^+$. UV–vis ($CH_2Cl_2$) $\lambda_{max}$ ($\varepsilon$) [nm (× $10^4$ $cm^{-1}$ $M^{-1}$)]: 414 (2.45), 522 (0.25), 564 (0.30), 590 (0.80).

5,15-Bis-nitrophenyl-10-(4-methoxyphenyl) phosphorus corrole, **8P.** The reddish-pink color solid was obtained in 40% yield. $^1$H NMR (400 MHz, $CDCl_3$) $\delta$ 9.35 (dd, $J$ = 4.4, 2.5 Hz, 2H), 9.00 (dd, $J$ = 4.4, 2.6 Hz, 2H), 8.92 (dd, $J$ = 4.9, 2.8 Hz, 2H), 8.69 (dd, $J$ = 5.7, 2.9 Hz, 2H), 3.89 (s, 3H); $^{31}$P NMR ($CDCl_3$): $\delta$: −100.02 ppm, $\delta$: −190.88 ppm. HRMS: m/z (%) observed: 691.2786 $[M-OH^-]^+$; calculated: 691.2975 $[M-OH^-]^+$. UV–vis ($CH_2Cl_2$) $\lambda_{max}$ ($\varepsilon$) [nm (× $10^4$ $cm^{-1}$ $M^{-1}$)]: 424 (3.0), 537 (0.30), 574 (0.45), 616 (0.80).

5,15-Bis-nitrophenyl-10-(4-bromophenyl) phosphorus corrole, **9P.** The reddish-pink color solid was obtained in 46% yield. $^1$H NMR (400 MHz, $CDCl_3$): $\delta$ 9.36 (dd, $J$ = 4.5, 2.5 Hz, 2H), 9.02 (dd, $J$ = 5.0, 3.6 Hz, 2H), 8.97 (dd, $J$ = 4.4, 3.5 Hz, 2H), 8.86 (dd, $J$ = 5.0, 3.0 Hz, 2H); $^{31}$P NMR ($CDCl_3$): $\delta$: −100.36 ppm, $\delta$: −190.80 ppm. HRMS: m/z (%) observed: 757.2380 $[M]^+$, 758.2431 $[M+1]^+$; calculated: 757.2980 $[M]^+$, 758.2980 $[M+1]^+$. UV–vis ($CH_2Cl_2$) $\lambda_{max}$ ($\varepsilon$) [nm (× $10^4$ $cm^{-1}$ $M^{-1}$)]: 426 (3.60), 536 (0.10), 573 (0.25), 613 (1.00).

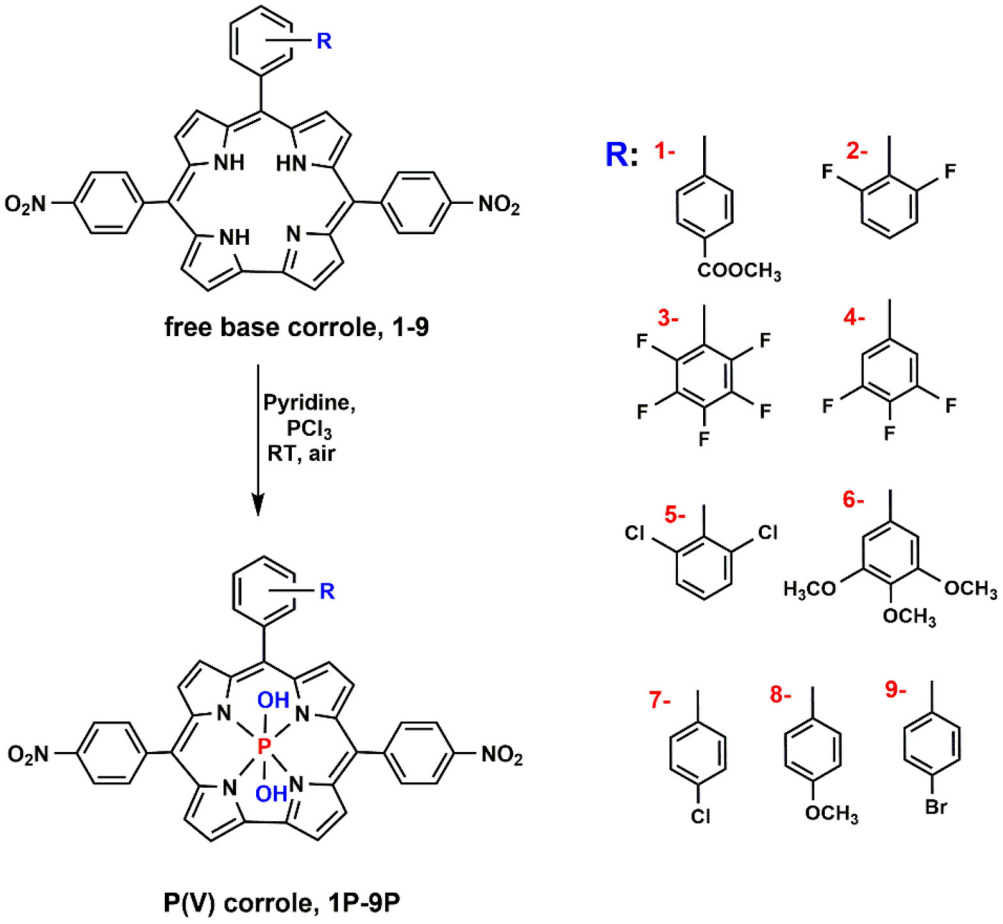
### 2.3. DFT calculations

Density functional theory (DFT) method has been employed for structural optimization of all the geometries using Gaussian09, revision D.01, suits of program [40]. No constraints have been added for the geometry optimization. The calculation of frequency has also been performed on all the optimized geometry, in order to reach the global minima. The crystal coordinates of 15-(m,p,p-triflurophenyl)phosphorus corrole have been used as a primary structure, all the subsequent input structures were made by changing the substituents using GaussView 6.0 software [41]. Geometrical optimizations have been carried out using the unrestricted mode of DFT calculation, bearing B3LYP functional *viz.* Becke's three parameter hybrid exchange correlation functional, given by the Lee et al. [42]. 6–31G(d,p) basic set was utilized for geometry optimization and frequency calculations while 6–311G(d,p) was used for energy as single point calculations on the optimized geometries [43,44].

## 3. Results and discussion

### 3.1. Synthesis

The synthesis of free-base corroles **1–9** was carried out according to the method previously reported by Koszarna and coworkers [38]. This method proved out to be quite facile, efficient and helped us to obtain a wide series of corroles in sufficient yields. All synthesised corroles contain one $-NO_2$ group on $C_5$ *meso* carbon and another $-NO_2$ group on $C_{15}$ *meso* carbon. These corroles differed only at $C_{10}$, where we have introduced different aryl substituents in order to study the effects of electron-releasing and electro-withdrawing group on $C_{10}$ *meso* group and hence oriented a series of ABA free-base corroles. The compounds **1–9** were highly soluble in dichloromethane and their analytical details matched well with the previously reported work by our research group [39]. Insertion of phosphorus into the cavity of free-base corroles, **1–9**, was easily achieved by dissolving them in pyridine at room temperature followed by careful addition of $PCl_3$ (Scheme 1). Color change from dark green to light green was observed in the reaction mixture. After standard workup and column chromatographic purification desired compounds **1P-9P** were collected as reddish-pink colored solids in good yields. Also, we have conducted axial hydroxyl exchange in compound **5P**. In our previous report [19] we have carried out exchange of axial -OH in the case of compound **3P** highlighting, axial ligand based penta-hexa coordination behavior of phosphorus corroles. Similarly, we have obtained **5P($F_2$)**, via




**Scheme 1.** Synthetic scheme of phosphorus corroles, **1P-9P**.

addition of excess amount of HF to a solution of **5P** in $CH_2Cl_2$. In accordance with our preceding work [19], **5P($F_2$)** also depicted a much stable hexa-coordinated phosphorus corrole. Likewise, all the phosphorus corroles *viz* **1P-9P** could be stabilized in hexa-coordinated form via axial-ligand stabilization. Structural characterizations of synthesised phosphorus complexes were carried out using UV-visible spectroscopy, mass spectrometry, NMR and single crystal X-ray techniques.

### 3.2. Absorption studies

Synthesised free-base corroles, **1–9** and phosphorus corroles, **1P-9P** were dissolved in DCM and UV-vis spectra was recorded (see the Supporting Information, Table T1). Fig. 1 shows the absorption spectra of compounds **1P-9P** recorded in DCM and comparison of UV-Vis spectra of compound **1** and **1P** (inset). In UV–visible spectra of free base corroles Soret band appeared at 425–452 nm and Q-band at 585–650 nm. However, phosphorus (V) corroles have Soret band in the region 412–415 nm and three Q-bands at 522–600 nm. On close observation, it was realized that the Soret band of phosphorus corroles, **1P-9P** have undergone blue shift while the Q band has red shifted as compared to free-base corroles, **1–9** (see Fig. 1 (inset)).

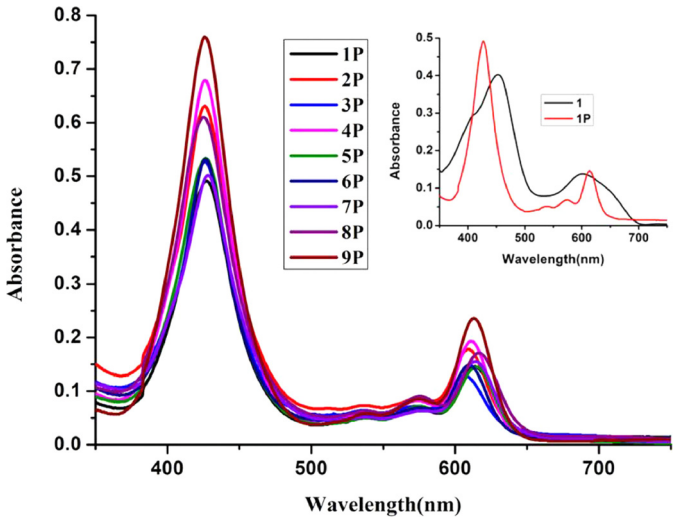



**Fig. 1.** UV-Vis spectra of compounds **1P-9P** in DCM ($\approx 2 \times 10^{-5}$ M). Comparison of UV-Vis spectra of compound **1** and **1P** (inset).

### 3.3. NMR spectra

Synthesised phosphorus corroles were characterised by both $^1$H NMR and $^{31}$P NMR spectroscopy. $^1$H NMR spectra of phosphorus complexes, **1P-9P** was recorded in deuterated chloroform ($CDCl_3$), the spectra depicted four strong sets of doublets due to $\beta$-pyrrolic protons along with another four sets of doublets with minor intensity (see the Supporting Information, Figs. S10–S14). The reason behind such a spectrum might be axial-ligand dissociation [11,12], from hexacoordinated geometry to possibly pentacoordinated geometry [19,45]. $^{31}$P NMR spectra of phosphorus complexes, **1P-9P** showed two peaks, one at $\approx -100$ ppm and another at $\approx -200$ ppm (see the Supporting Information, Figs. S15–S18), which

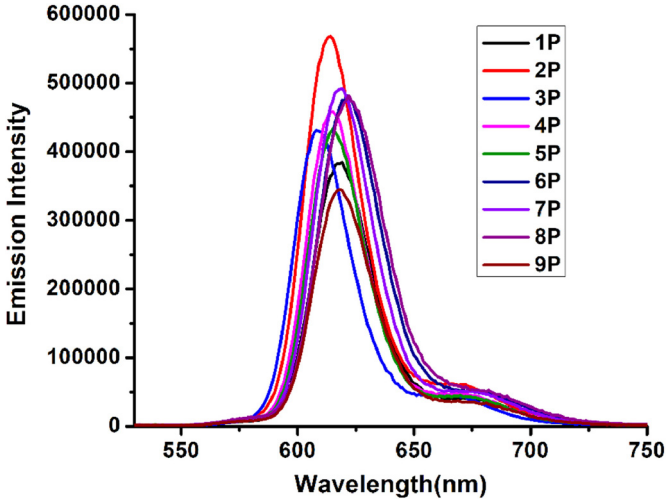



**Fig. 2.** Emission spectra of compounds **1P-9P** in toluene at room temperature.

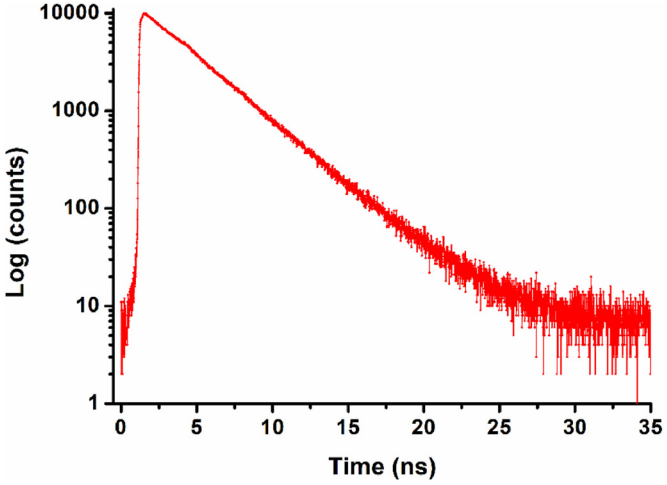



**Fig. 3.** Fluorescence-decay profile of the fluorescence decay of compound **1P** in toluene.

again highlighted the existence of hexacoordinated and pentacoordinated phosphorus center in the solution.

### 3.4. Fluorescence studies

Steady-state and time-resolved fluorescence properties of compounds **1P–9P** were studied. Generally, main group metallocorroles are strongly fluorescent [46] with high quantum yield values in comparison to their corresponding porphyrins derivatives [47–49]. In our research, we observed compounds **1P–9P** displayed a strong fluorescence band around 615 nm (Fig. 2) which are slightly red shifted as compared to already reported $A_3$ type phosphorus corroles [12–15]. These dyes are strongly emissive and the fluorescence quantum yields ($\Phi$= 0.33–0.49) are larger than that of analogous phosphorus porphyrins [12,14]. The fluorescence quantum yields of compounds **1P-9P** are somewhat lower than the previously reported $A_3$ type phosphorus corroles as a consequence of the presence of electron withdrawing nitro groups [15]. Nitro groups are considered to be fluorescence quenchers, these groups quenches fluorescence through a spin orbit coupling mechanism that increases the intersystem-crossing processes [50–52]. The position of nitro group in the nitroaromatics is a very important deciding factor when it comes to fluorescence quenching. When nitro group is present at para-position it does not show much significant quenching while when present at ortho or meta position it quenches the fluorescence much more significantly may be due to steric factors. More near the nitro group is to the macrocycle ring more quenching action is seen. In 2012, Ghosh and Ravikanth discussed nitro group containing $A_3$ type phosphorus corroles, where when nitro group was present at the para position quantum yield was found to be 0.30 while when present at the meta position quantum yield was 0.04. Evidently, fluorescence quantum yield decreased when the nitro group moved closer to the macrocycle ring [12]. In the P(V) corrole complexes 1P-9P, nitro group is present at the para position that is why it does not quench fluorescence to that extent and have comparatively higher quantum yield values ($\Phi$= 0.33–0.49). We have chosen these corroles with the aim of exploring ABA type P(V) corroles which have not been explored yet and in turn focusses on highlighting the effect of electron donating and electron withdrawing groups at the B position. Quantum yields of compounds containing halogen atoms, viz compound **3P-5P**, compound **7P** and **9P** were considerably low, reason being the heavy halogen atom effect which is responsible for enhancement of intersystem crossing rates and in turn decreases the quantum yields. In order to explore photophysical properties of the synthesized corroles more deeply, fluorescence lifetimes ($\tau$) were also determined. Fluorescence decays of compounds **1P–9P** were fitted to a single exponential and singlet-state lifetimes ($\tau$) of compounds **1P–9P** were in the range 3.14–3.91 ns and values are in accordance with the previously reported phosphorus corroles [12,14]. Single exponential fitted illustrative fluorescent decay profile of compound **1P** in toluene solvent has been presented in Fig. 3. Compound **2P**, exhibited the highest quantum yield ($\Phi$=0.49) whereas compound **3P** showed the highest singlet state lifetime ($\tau$) of nearly 3.91 ns.

**Table 1**
Photophysical and electrochemical data of phosphorus corroles **1P-9P**.

| Compound | Fluorescence quantum yield (%)(In toluene) | Life time (ns)(In toluene) | $E^{ox}_{1/2}$(V) | $E_{pc}$(V) |
|---|---|---|---|---|
| **1P** | 33 | 3.27 | 0.79 | −1.20 |
| **2P** | 49 | 3.59 | 0.85 | −1.23 |
| **3P** | 34 | 3.91 | 0.82 | −1.22 |
| **4P** | 38 | 3.41 | 0.77 | −1.24 |
| **5P** | 36 | 3.35 | 0.84 | −1.26 |
| **6P** | 44 | 3.15 | 0.79 | −1.25 |
| **7P** | 43 | 3.36 | 0.83 | −1.22 |
| **8P** | 46 | 3.14 | 0.84 | −1.21 |
| **9P** | 30 | 3.34 | 0.83 | −1.25 |

### 3.5. Electrochemistry

The electrochemical properties of compounds **1P-9P** were studied by using cyclic voltammetry. Scan rate was kept at 100 $mVs^{-1}$ using tetrabutylammonium perchlorate (TBAP) as the supporting electrolyte (in acetonitrile). We have recorded cyclic voltammograms of **1P-9P** corroles in acetonitrile. Fig. 4 shows the representative cyclic voltammogram of compound **1P** and voltammograms of **2P-9P** are in the Supporting Information, Figs. S19–S26. The half-wave potential ($E_{1/2}$) for the anodic part and peak cathodic potential ($E_{pc}$) for the cathodic part of all the synthesized $P^v$ corroles (1P-9P) are compiled in Table 1. All nine $P^v$ corrole complexes have different substituent at $C_{10}$ *meso* position of $P^v$ corrole. In the anodic part, reversible redox couple were obtained for all the phosphorus corroles (1P-9P). $E_{1/2}$ for the first oxidation process were found to be in the range 0.77 V to 0.85 V. Whereas in the cathodic part, an irreversible redox couple was obtained with $E_{pc}$ of the first reduction process in the range −1.20 V to −1.26 V.

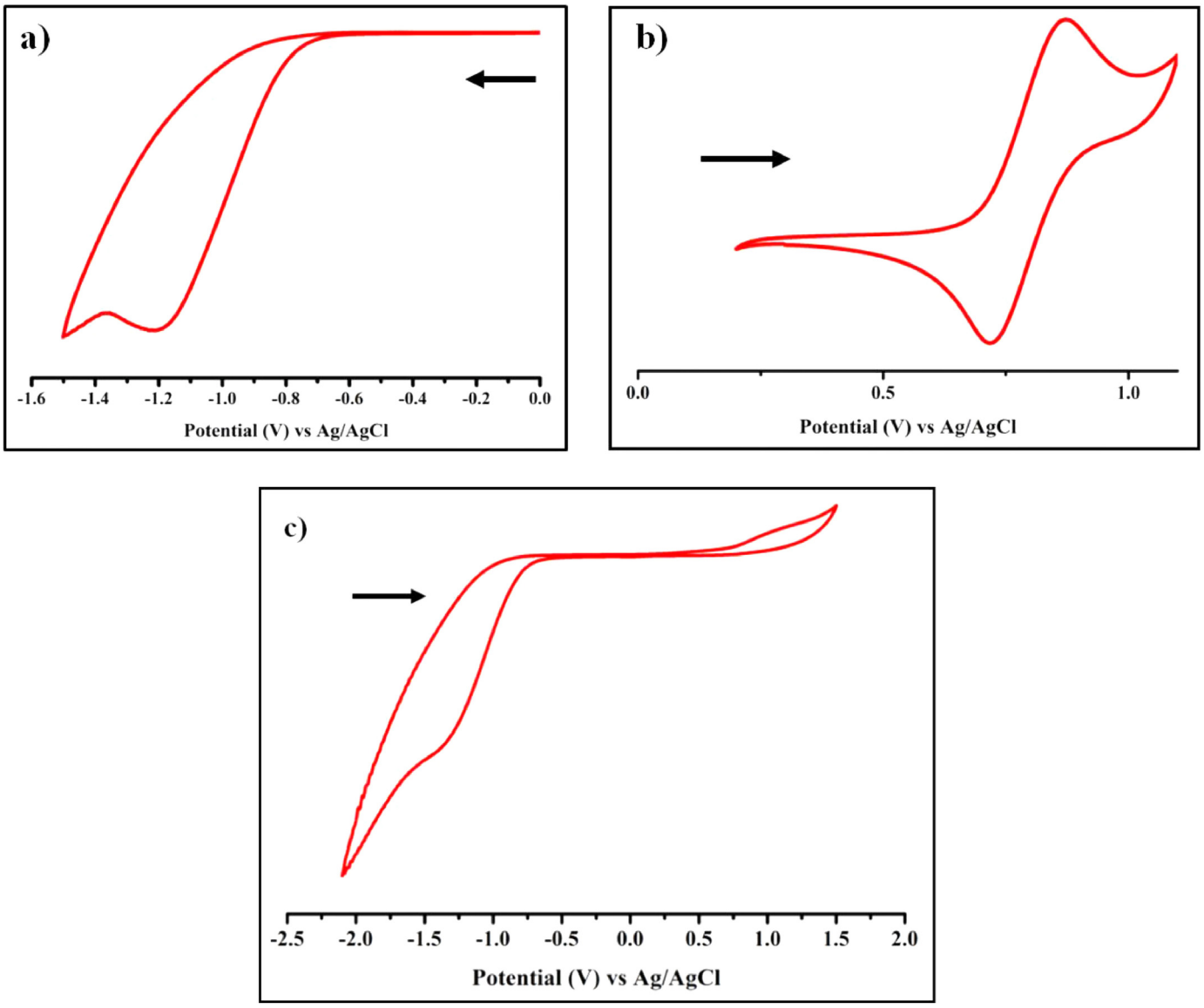



**Fig. 4.** Cyclic voltammogram of compound **1P** scanned from (a) 0 V to −1.6 V; (b) 0 V to 1.1 V; (c) −2.0 V to 2.0 V in acetonitrile using TBAP as supporting electrolyte; scan rate: 100 mV $s^{-1}$.

The half-wave and peak cathodic potential values for the synthesised ABA corroles (1P-9P) were found to be comparable with that of already reported $A_3$ type $P^v$ corroles [12–15].

### 3.6. DFT calculation

The Gouterman four orbital model can be used to explain most of the electronic properties and transitions in tetrapyrrole systems. The four orbitals in focus are the two highest occupied molecular orbitals (HOMO and HOMO-1) and the two lowest unoccupied molecular orbitals (LUMO and LUMO+1). Based on the DFT calculations on unsubstituted phosphorus corrole, it has been shown that the orbitals do not mix generally [53]. The HOMO-1 contains electron density mostly on the pyrrole ring and in HOMO it is somewhat on the *meso* carbon and the nitrogen. LUMO and LUMO+1, on the other hand still possesses most of the $\pi$ character and are energetically much closer. In the synthesized ligands two more orbitals can be observed between these two orbitals. The LUMO+1 and LUMO+2 mostly have the electron density based on the p-nitrophenyl ring, while LUMO+4 again represents similar density map of LUMO+1 of the unsubstituted phosphorus corrole. Variation in the calculated energy gap between the frontier molecular orbitals in the compounds 1P-9P has been displayed in Fig. 5. The presence of these closely placed unoccupied orbitals may be responsible for easy coupling and reduction in fluorescence efficiency reported for dinitro phosphorus corroles compared to their other metallocorroles. The variation in $\Delta E$ for these compounds follows the electron donation pattern of the phenyl ring at $C_{10}$ position. For the Hammet parameter plot against the calculated $\Delta E$ or the HOMO-LUMO energy gap almost gives a straight line (Fig. 6). This is expected as more electron withdrawing groups like fluorine will pull electrons towards the ring thereby decreasing the energy of HOMO. The empty LUMO on the other hand will be affected less thereby making the HOMO-LUMO gap larger in case of pentafluoro (**3P**) substitution. Methoxy groups having electron donating properties at the para position (**8P**) on the other hand should have a reverse effect and have lower HOMO-LUMO gap. At meta position on the other hand methoxy groups show electron withdrawing nature thereby, in 3,4,5 trimethoxy phenyl group (**6P**) at $C_{10}$ the HOMO-

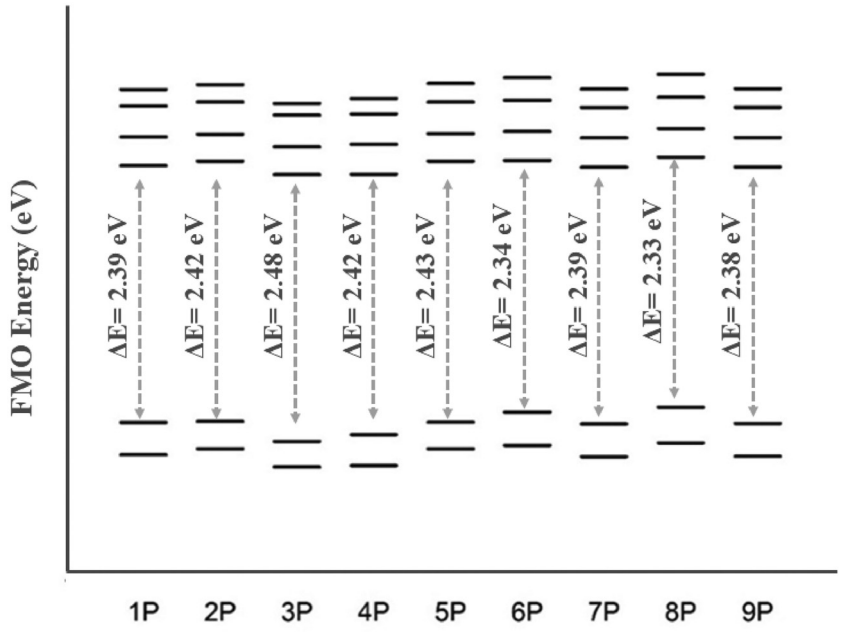



**Fig. 5.** Energy of the frontier molecular orbitals of compounds **1P-9P** of HOMO-1, HOMO, LUMO, LUMO+1, LUMO+2, LUMO+3.

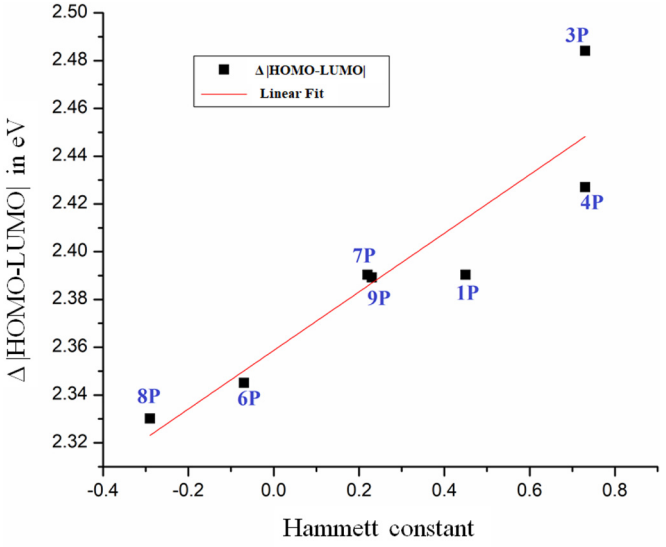



**Fig. 6.** Hammet parameter plot against the HOMO-LUMO energy gap for compounds **1P-9P**.

LUMO gap is slightly higher than **8P**. Frontier molecular orbitals energy values of DFT optimized molecules of **1P-9P** have been depicted in Table 2.

### 3.7. Crystallographic characterization

In order to further enable the detailed insight on structural arrangement, the X-ray structures of compounds **2P, 5P, 5P($F_2$)** and **6P** were obtained by slow diffusion method, where n-heptane was diffused slowly into a solution of these compounds in dichloromethane. The ORTEPs view of compound **2P, 5P, 5P($F_2$)** and **6P** are shown in Fig. 7. Compounds **2P, 5P, 5P($F_2$)** and **6P** crystallized in the triclinic space group P-1. Other crystal data is well tabulated in **Table T2–T4** (supporting information). Compounds **2P** and **5P** crystallized in the form of hexacoordinated complex with $OH^-$ groups as axial ligand, whereas **6P** crystallized as a pentacoordinated complex. Herein we report, the second P(V)=O complex of triarylcorrole (**6P**), characterized by single-crystal X-ray studies. Compound **5P($F_2$)** crystallized in the form of hexacoordinated complex with $F^-$ groups as axial ligand. In compounds **2P, 5P** and **5P($F_2$)**, phosphorus atom (present in the center) was nearly in the mean plane defined by four pyrrolic nitrogen atoms, while pentacoordinated phosphorus ion is 0.55 Å out of the mean corrole plane towards the axial oxygen atom. To our knowledge, phosphorus atom was usually found to lie above the mean corrole plane in pentacoordinated complexes and in the plane in case of hexacoordinated complexes, likewise previously reported metalloporphyrins and Metallocorroles [54–58]. The P–O bond lengths of compounds **2P** and **5P** are not equivalent and the average P–O bond lengths are 1.6565 Å and 1.6575 Å, respectively, they were almost in the same range of the analogous P(V) corrole complexes [59–61]. The P=O bond length (1.497 Å) of compound **6P** is reasonably shorter than P–O lengths observed in compounds **2P** and **5P**. These observations are an important structural insight, that the phosphorus atom is bound to oxygen atom through a double bond in **6P** while in **2P** and **5P** it is bound to the axial oxygen atom via single bond. The average P–F bond length of compound **5P($F_2$)** was 1.6075 Å. The average P–N bond lengths (1.82 Å) of compounds **2P, 5P, 5P($F_2$)** and **6P** were in the same range as those found in other previously reported phosphorus corroles [3,4,62–64]. Furthermore, in compounds **2P, 5P, 5P($F_2$)** and **6P** dihedral angle made by the *meso*-phenyl group (at the $\beta$ position) to the mean corrole plane were 84.28°, 80.31°, 88.16° and 60.98°, respectively. Detailed crystal data and data collection parameters of compounds **2P, 5P, 5P($F_2$)** and **6P** along with their comparison with other reported analogous corroles are displayed in **Table T2–T4** (supporting information).

## 4. Conclusion

A series of nine ABA type phosphorus corrole complexes (**1P-9P**) with different substituents have been synthesised and characterized by UV-visible and NMR spectroscopy. X-ray crystal structures of four complexes were obtained. For all the synthesised corroles photophysical characteristics were studied. Replacing $A_3$ type phosphorus corroles with ABA type led to a slight red shift in the fluorescence band. Although fluorescence quantum yields were

**Table 2**
Frontier molecular orbitals energy values of DFT optimized molecules of **1P-9P**.

| Molecule | Molecular orbitals (eV) | | | | | | ΔE (eV) |
|---|---|---|---|---|---|---|---|
| | HOMO-1 | HOMO | LUMO | LUMO+1 | LUMO+2 | LUMO+3 | |
| **1P** | −5.83117 | −5.52967 | −3.1394 | −2.87082 | −2.58238 | −2.42864 | 2.3903 |
| **2P** | −5.77539 | −5.51987 | −3.09586 | −2.84416 | −2.54592 | −2.38482 | 2.4240 |
| **3P** | −5.94492 | −5.70682 | −3.22267 | −2.96362 | −2.66837 | −2.55843 | 2.4841 |
| **4P** | −5.93267 | −5.64396 | −3.21696 | −2.93858 | −2.65585 | −2.51299 | 2.4270 |
| **5P** | −5.77675 | −5.52804 | −3.09532 | −2.84007 | −2.54673 | −2.37258 | 2.4327 |
| **6P** | −5.74382 | −5.43388 | −3.08879 | −2.81994 | −2.52986 | −2.31924 | 2.3451 |
| **7P** | −5.84968 | −5.54464 | −3.15437 | −2.88035 | −2.59653 | −2.42483 | 2.3903 |
| **8P** | −5.72124 | −5.38899 | −3.05886 | −2.79327 | −2.50075 | −2.28605 | 2.3301 |
| **9P** | −5.84587 | −5.54028 | −3.1511 | −2.87763 | −2.59354 | −2.42129 | 2.3892 |

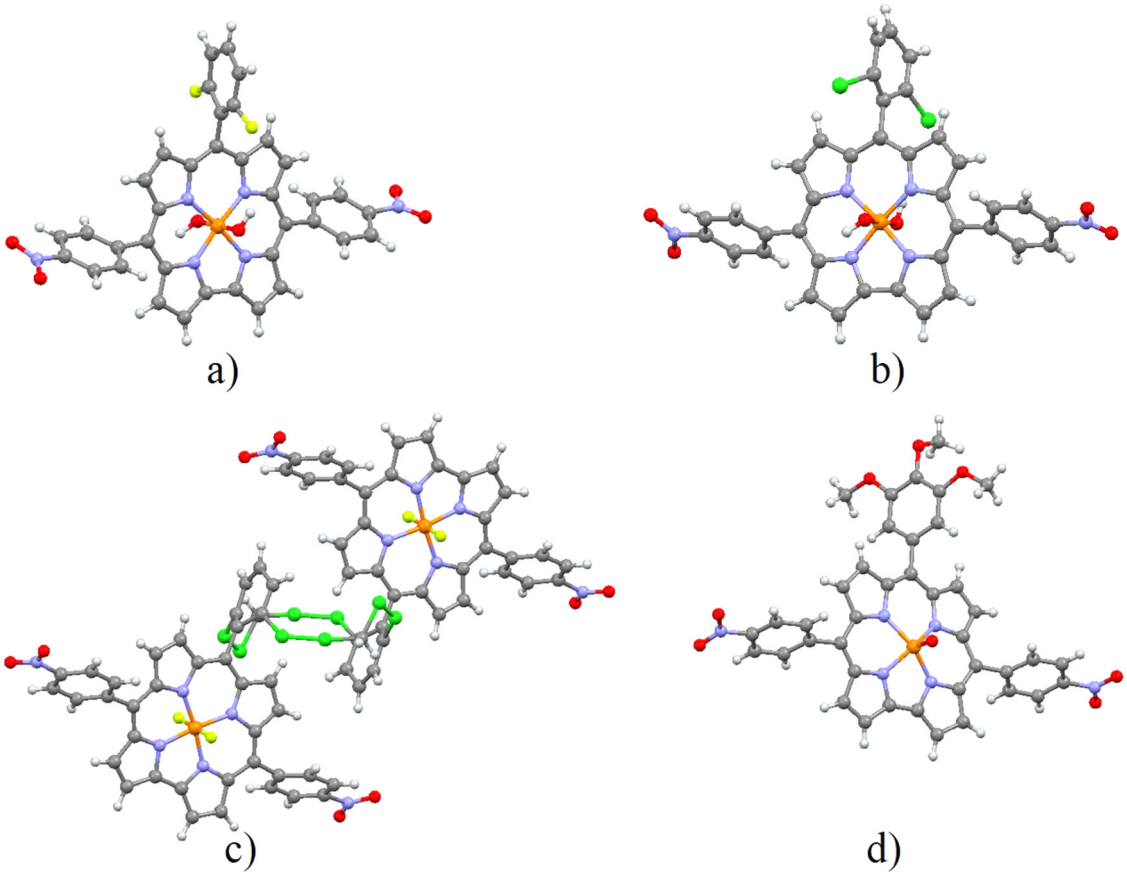



**Fig. 7.** X-ray crystal structures of compounds (a) **2P**; (b) **5P**; (c) **5P(F$_2$)** and (d) **6P**. Thermal ellipsoids are shown at the 50% probability level.

seen to decrease in the case of phosphorus corroles due to the presence of electron withdrawing nitro groups at ortho position which induces fluorescence quenching. But in compounds **1P-9P** negligible quenching was experienced, reason being the presence of nitro group at para position. Quenching action of nitro substituents increases when present in proximity to the macrocyclic ring while it decreases when away from the ring that is why quite high quantum yield values (0.33 to 0.49) have been obtained in our case. It can be concluded that when compared with $A_3$ type, ABA type P(V) corroles (**1P-9P**) do have less fluorescence quantum yields but are still fairly fluorescent to be employed for numerous applications in the time to come. Furthermore, the electrochemical properties were explored using cyclic voltammetry, which revealed $E_{1/2}$ values for the first oxidation process in the range 0.77 V to 0.85 V and $E_{pc}$ of the first reduction process in the range −1.20 V to −1.26 V for **1P-9P** corroles. Measured electrode potentials had nearly same values as that of previously reported $A_3$ type $P^v$ corroles. In addition to this, DFT calculations were carried out which further strengthened the effect of *meso* substituents on the electrochemical and photophysical properties of the synthesized phosphorus corroles. We believe that these exclusive features of phosphorus corroles would be of great importance and would unfold promising applications in the near future.

## 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.

## CRediT authorship contribution statement

**Ritika Kubba:** Writing – original draft, Methodology. **Omprakash Yadav:** Writing – review & editing, Conceptualization.

**Suman Maji:** Methodology. **Natalia Fridman:** Methodology. **Anil Kumar:** Conceptualization, Methodology, Software.

## Data availability

I have attached the supplementary materialssupplementary in

## Acknowledgement

AK and RK are grateful to Delhi Technological University (India) for infrastructural facilities. Authors also thank Professor Zeev Gross, Schulich Faculty of chemistry, Technion-Israel Institute of Technology, Haifa for using instrumental facilities.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.molstruc.2022.133780.

## References

[1] A.W. Johnson, I.T. Kay, Corroles. Part I. Synthesis, J. Chem. Soc. (1965) 1620–1629.
[2] R. Paolesse, S. Licoccia, M. Fanciullo, T. Boschi, Synthesis and characterization of cobalt (III) corroles complexes of *meso*-phenyl-substituted corroles, Inorg. Chim. Acta 203 (1993) 107–114.
[3] L.M. Reith, M. Stiftinger, U. Monkowius, Synthesis and Characterization of a stable bismuth (III) A 3 À corrole, Inorg. Chem. 50 (2011) 6788–6797.
[4] A.M. Albrett, J. Conradie, P.D.W. Boyd, G.R. Clark, A. Ghosh, P.J. Brothers, Corrole as a binucleating ligand : preparation, molecular structure and density functional theory study of diboron corroles, J. Am. Chem. Soc. 130 (2008) 2888–2889, doi:10.1039/b714894a.(6).
[5] K.M. Kadish, C. Erben, Z. Ou, V.A. Adamian, S. Will, E. Vogel, Corroles with group 15 metal ions . synthesis and characterization of octaethylcorroles containing As, Sb, and Bi Ions in +3, +4, and +5 oxidation states, Inorg. Chem 39 (2000) 3312–3319.
[6] K.M. Kadish, S. Will, V.A. Adamian, B. Walther, C. Erben, Z. Ou, N. Guo, E. Vogel, Synthesis and electrochemistry of TIN (IV) octaethylcorroles, (OEC)Sn($C_6H_5$) and (OEC)SnCl, Inorg. Chem 37 (1998) 4573–4577.
[7] R. Paolesse, T. Boschi, S. Licoccia, G. Khoury, K.M. Smith, Phosphorus complex of corrole, Chem. Commun. (1998) 1119–1120.
[8] K.M. Kadish, Z. Ou, V.A. Adamian, R. Guilard, C.P. Gros, C. Erben, W. Stefan, E. Vogel, Corroles with group 15 ions. 2. Synthesis and characterization of octaethylcorroles containing a phosphorus central atom, Inorg. Chem. 39 (2000) 5675–5682, doi:10.1021/ic0010196.
[9] L. Simkhovich, A. Mahammed, I. Goldberg, Z. Gross, Synthesis and characterization of germanium, tin, phosphorus, iron, and rhodium complexes of tris(pentafluorophenyl)corrole, and the utilization of the iron and rhodium corroles as cyclopropanation catalysts, Chem. A Eur. J. 7 (2001) 1041–1055.
[10] G. Pomarico, S. Nardis, M.L. Naitana, M. Grac, H. Vicente, K.M. Kadish, P. Chen, L. Prodi, D. Genovese, R. Paolesse, Aluminum, gallium, germanium, copper, and phosphorus complexes of *meso*-triaryltetrabenzocorrole, Inorg. Chem 52 (2013) 4061–4070.
[11] A. Ghosh, W. Lee, M. Ravikanth, Synthesis, structure and properties of a five-coordinate oxophosphorus (V) *meso*-triphenylcorrole, Eur. J. Inorg. Chem. (2012) 4231–4239, doi:10.1002/ejic.201200651.
[12] A. Ghosh, M. Ravikanth, Synthesis, structure, spectroscopic, and electrochemical properties of highly fluorescent phosphorus(V)– *meso*-triarylcorroles, Chem. A Eur. J. 18 (2012) 6386–6396, doi:10.1002/chem.201103226.
[13] J. Vestfrid, R. Kothari, A. Kostenko, I. Goldberg, B. Tumanskii, Z. Gross, Intriguing physical and chemical properties of phosphorus corroles, Inorg. Chem. 55 (2016) 6061–6067, doi:10.1021/acs.inorgchem.6b00544.
[14] Q. Chen, Z. Xiao, S. Fite, N. Fridman, X. Zhan, O. Keisar, Tuning chemical and physical properties of phosphorus corroles for advanced applications, Chem. A Eur. J. 25 (2019) 11383–11388, doi:10.1002/chem.201902686.
[15] A. Mahammed, K. Chen, J. Vestfrid, J. Zhao, Z. Gross, Phosphorus corrole complexes : from property tuning to applications in photocatalysis and triplet – triplet annihilation upconversion †, Chem. Sci. 10 (2019) 7091–7103, doi:10.1039/c9sc01463b.
[16] X. Liang, J. Mack, L. Zheng, Z. Shen, N. Kobayashi, Phosphorus(V)-corrole: synthesis, spectroscopic properties, theoretical calculations, and potential utility for *in Vivo* applications in living cells, Inorg. Chem. 53 (2014) 2797–2802, doi:10.1021/ic402347w.
[17] J. Pohl, I. Saltsman, A. Mahammed, Z. Gross, Inhibition of green algae growth by corrole-based photosensitizers, J. Appl. Microbiol. 118 (2014) 305–312, doi:10.1111/jam.12690.
[18] C. Di Natale, C.P. Gros, R. Paolesse, Corroles at work: a small macrocycle for great applications Corrado, Chem. Soc. Rev. 51 (2022) 1277–1335, doi:10.1039/d1cs00662b.
[19] R. Kubba, O. Yadav, P. Yadav, N. Fridman, A. Kumar, Penta -hexa coordination behaviour of ABA-P (V) corrole, J. Mol. Struct. 1243 (2021) 130857, doi:10.1016/j.molstruc.2021.130857.
[20] N. Krishnakanth, S. Achary, A.R. Ramya, C. Biswas, S. Santosh, K. Raavi, L. Giribabu, V. Rao, Ultrafast photophysical and nonlinear optical properties of novel free base and axially substituted phosphorus (V) corroles, J. Mol. Liq. 311 (2020) 113308, doi:10.1016/j.molliq.2020.113308.
[21] L. Giribabu, J. Kandhadi, Phosphorus (V) corrole- porphyrin based hetero trimers : synthesis, spectroscopy and photochemistry, J. Fluoresc. 24 (2014) 569–577, doi:10.1007/s10895-013-1328-1.
[22] M.E. Alberto, B.C. De Simone, G. Mazzone, N. Russo, Photophysical properties of nitrated and halogenated phosphorus tritolylcorrole complexes: insights from theory, Molecules 23 (2018) 2779, doi:10.3390/molecules23112779.
[23] S. Lai, L. Wang, B. Wan, A. Lu, H. Wang, H. Liu, Photophysical properties, singlet oxygen generation and DNA binding a ffi nity of Tris (4-pyridyl) corrole and its phosphorous, gallium and tin complexes, J. Photochem. Photobiol. A Chem. 390 (2020) 112283, doi:10.1016/j.jphotochem.2019.112283.
[24] B.S. Achary, A.R. Ramya, J.B. Nanubolu, S. Seetharaman, G.N. Lim, Y. Jang, F. D'Souza, L. Giribabu, Axially substituted phosphorous(V) corrole with polycyclic aromatic hydrocarbons: synthesis, X-ray structure, and photoinduced energy and electron transfer studies, New J. Chem. 42 (2018) 8230–8240, doi:10.1039/C7NJ04363E.
[25] A. Mohajeri, N. Hassani, The interplay between structural perfectness and CO oxidation catalysis on aluminum, phosphorous and silicon complexes of corroles, Phys. Chem. Chem. Phys. 21 (2019) 7661–7674, doi:10.1039/c8cp07372d.
[26] X. Zhan, P. Teplitzky, Y. Diskin-posner, M. Sundararajan, Z. Ullah, Q. Chen, L.J.W. Shimon, I. Saltsman, A. Mahammed, M. Kosa, M. Baik, D.G. Churchill, Z. Gross, Maximizing property tuning of phosphorus corrole photocatalysts through a trifluoromethylation approach, Inorg. Chem. 58 (2019) 6184–6198, doi:10.1021/acs.inorgchem.9b00436.
[27] A. Mahammed, Z. Gross, Metallocorroles as photocatalysts for driving endergonic reactions, exemplified by bromide to bromine conversion, Angew. Chem. Int. Ed. 54 (2015) 12370–12373, doi:10.1002/anie.201503064.
[28] A. Preuß, I. Saltsman, A. Mahammed, M. Pfitzner, I. Goldberg, Z. Gross, Photodynamic inactivation of mold fungi spores by newly developed charged corroles, J. Photochem. Photobiol. B Biol. 133 (2014) 39–46, doi:10.1016/j.jphotobiol.2014.02.013.
[29] Z. Zhang, H. Yu, H. Huang, H. Wang, S. Wu, H. Liu, H. Zhang, The photodynamic activity and toxicity evaluation of 5,10,15- tris(ethoxylcarbonyl)corrole phosphorus(V) *in vivo* and *in vitro*, Eur. J. Med. Chem. (2019) 779–786, doi:10.1016/j.ejmech.2018.12.031.
[30] X. Jiang, R. Liu, H. Liu, C.K. Chang, Corrole-based photodynamic antitumor therapy, J. Chin. Chem. Soc. (2019), doi:10.1002/jccs.201900176.
[31] R.C. Soy, B. Babu, J. Mack, T. Nyokong, Dyes and pigments the photodynamic activities of the gold nanoparticle conjugates of phosphorus (V) and gallium (III) $A_3$ *meso*-triarylcorroles, Dye. Pigment. 194 (2021) 109631, doi:10.1016/j.dyepig.2021.109631.
[32] B.J. Brennan, Y.C. Lam, P.M. Kim, X. Zhang, G.W. Brudvig, Photoelectrochemical cells utilizing tunable corroles, Appl. Mater. Interfaces 7 (2015) 16124–16130, doi:10.1021/acsami.5b05050.
[33] I. Gupta, M. Ravikanth, Synthesis and photophysical studies of covalently linked porphyrin-21-thiaporphyrin dyads, Inorg. Chim. Acta 360 (2007) 1731–1742, doi:10.1016/j.ica.2006.09.016.
[34] E.D. Sternberg, D. Dolphin, Porphyrin-based photosensitizers for use in photodynamic therapy, Tetrahedron 54 (1998) 4151–4202.
[35] O.V. Dolomanov, L.J. Bourhis, R.J. Gildea, J.A.K. Howard, H. Puschmann, OLEX2 : a complete structure solution, refinement and analysis program, J. Appl. Crystallogr. 42 (2009) 339–341, doi:10.1107/S0021889808042726.
[36] L.J. Bourhis, O.V. Dolomanov, R.J. Gildea, A.K. Howard, The anatomy of a comprehensive constrained, restrained refinement program for the modern computing environment – Olex2 dissected, Acta Crystallogr. Sect. A Found. Adv. (2015) 59–75, doi:10.1107/S2053273314022207.
[37] G.M. Sheldrick, Crystal structure refinement with SHELXL, Acta Crystallogr. Sect. C Struct. Chem. (2015) 3–8, doi:10.1107/S2053229614024218.
[38] B. Koszarna, D.T. Gryko, Efficient synthesis of *meso*-substituted corroles in a $H_2O$-MeOH mixture, J. Org. Chem. 71 (2006) 3707–3717, doi:10.1021/jo060007k.
[39] O. Yadav, A. Varshney, A. Kumar, R. Kumar, M. Singh, Spectrochimica acta part A : molecular and Biomolecular Spectroscopy $A_2B$ corroles : fluorescence signaling systems for sensing fl uoride ions, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 202 (2018) 207–213, doi:10.1016/j.saa.2018.05.051.
[40] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery, J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, in: Gaussian 09 Revision D.01, Gaussian, Inc., Wallingford CT, 2010, pp. 1–20. citeulike-article-id:9096580.

[41] R. Dennington, T.A. Keith, J.M. Millam, GaussView 6.0.16, Gaussian Inc, Wallingford, 2016.
[42] C. Lee, W. Yang, R.G. Parr, Development of the colic-salvetti correlation-energy into a functional of the electron density formula chengteh, Phys. Rev. B. 37 (1988) 785–789.
[43] W.J. Hehre, R. Ditchfield, J.A. Pople, Self — consistent molecular orbital methods. XII. further extensions of gaussian — type basis sets for use in molecular orbital studies of organic molecules, J. Chem. Phys. 56 (1972), doi:10.1063/1.1677527.
[44] M.J. Frisch, J.A. Pople, J.S. Binkley, Selfconsistent molecular orbital methods 25 . supplementary functions for Gaussian basis sets, J. Chem. Phys. 80 (1984) 3265–3269, doi:10.1063/1.447079.
[45] T. Chatterjee, W. Lee, M. Ravikanth, Stabilization of hexa-coordinated P(V) corroles by Axial Silyloxy groups, Dalt. Trans. 45 (2016) 7815–7822, doi:10.1039/c6dt00842a.
[46] B. Ventura, A.D. Esposti, B. Koszarna, D.T. Gryko, L. Flamigni, Photophysical characterization of free-base corroles, promising chromophores for light energy conversion and singlet oxygen generation, New J. Chem. 29 (2005) 1559–1566.
[47] M. Tasior, D.T. Gryko, J. Shen, K.M. Kadish, T. Becherer, H. Langhals, B. Ventura, L. Flamigni, Energy- and electron-transfer processes in corrole - perylenebisimide - triphenylamine Array, J. Phys. Chem. C. 112 (2008) 19699–19709.
[48] A. Mahammed, Z. Gross, Aluminum corrolin, a novel chlorophyll analogue, J. Inorg. Biochem. 88 (2002) 305–309.
[49] P.J. Brothers, Boron complexes of porphyrins and related polypyrrole ligands : unexpected chemistry for both boron and the porphyrin, Chem. Commun. (2008) 2090–2102, doi:10.1039/b714894a.
[50] M. Chen, D. Chen, P. Chou, Fluorescent chromophores containing the Nitro Group : relatively unexplored emissive properties, Chempluschem 86 (2021) 11–27, doi:10.1002/cplu.202000592.
[51] Y.M. Poronik, B. Sadowski, K. Szychta, F.H. Quina, I. Vullev, D.T. Gryko, Revisiting the non-fluorescence of nitroaromatics : presumption versus reality, J. Mater. Chem. C 10 (2022) 2870–2904, doi:10.1039/d1tc05423f.
[52] W. Rodrı, F. Corte, L. Gutie, J. Peon, Excited state dynamics and photochemistry of nitroaromatic compounds, ChemComm 57 (2021) 12218–12235, doi:10.1039/d1cc04999b.
[53] A. Ghosh, Electronic structure of corrole derivatives : insights from molecular structures, spectroscopy, electrochemistry, and quantum chemical calculations, Chem. Rev. 117 (2017) 3798–3881, doi:10.1021/acs.chemrev.6b00590.
[54] L. Simkhovich, N. Galili, I. Saltsman, I. Goldberg, Z. Gross, Coordination chemistry of the novel 5, 10, 15-tris (pentafluorophenyl) corrole : synthesis, spectroscopy, and structural characterization of its cobalt (III), rhodium (III), and iron (IV) complexes, Inorg. Chem 39 (2000) 2704–2705.
[55] S. Nardis, R. Paolesse, S. Licoccia, F.R. Fronczek, M. Grac, T.K. Shokhireva, S. Cai, F.A. Walker, T. Vergata, NMR and structural investigations of a nonplanar iron corrolate : modified patterns of spin delocalization and coupling in a slightly saddled chloroiron (III) corrolate radical, Inorg. Chem. 44 (2005) 7030–7046.
[56] L. Wagnert, A. Berg, E. Stavitski, T. Berthold, G. Kothe, I. Goldberg, Applied magnetic resonance exploring the photoexcited triplet states of aluminum and tin corroles by time-resolved Q-band EPR, Appl. Magn. Reson. 30 (2006) 591–604.
[57] L. Simkhovich, P. Iyer, I. Goldberg, Z. Gross, Structure and chemistry of N-substituted corroles and their rhodium(i) and zinc(ii) metal-ion complexes, Chem. A Eur. J. 8 (2002) 2595–2601.
[58] S. Nardis, F. Mandoj, R. Paolesse, F.R. Fronczek, K.M. Smith, L. Prodi, M. Montalti, G. Battistini, Synthesis and functionalization of germanium triphenylcorrolate : the first example of a partially brominated corrole, Eur. J. Inorg. Chem. (2007) 2345–2352, doi:10.1002/ejic.200700184.
[59] C.A. Marrese, C.J. Carrano, Synthesis, characterization, and electrochemistry of (5, 10, 15, 20-tetraphenylporphinato) dichlorophosphorus (V) chloride, Inorg. Chem 22 (1983) 1858–1862.
[60] C.J. Carrano, M. Tsutsui, Unusual metalloporphyrins. Phosphorus complexes of tetraphenylporphine, J. Coord. Chem. 7 (1977) 79–83, doi:10.1080/00958977708073043.
[61] P. Sayer, M. Gouteman, C.R. Connell, Porphyrins, 34. phosphorus complexes of octaethylporphyrin, J. Am. Chem. Soc. (1977) 1082–1087.
[62] K.E. Thomas, I.H. Wasbotten, A. ghosh, copper -octakis (trifluoromethyl) corroles : new paradigms for ligand substituent effects in transition metal complexes, Inorg. Chem. 47 (2008) 10469–10478.
[63] O.R. Bjorn, V. Veryazov, J. Conradie, P.R. Taylor, A. Ghosh, Not Innocent : verdict from ab initio multiconfigurational second-order perturbation theory on the electronic structure of chloroiron corrole, Thr J. Phys. Chem. B. 112 (2008) 14099–14102.
[64] D. Kowalska, X. Liu, U. Tripathy, A. Mahammed, Z. Gross, S. Hirayama, R.P. Steer, Ground- and excited-state dynamics of aluminum and gallium corroles, Inorg. Chem 48 (2009) 2670–2676.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/361109568

# Two/Three-element textile Diversity MIMO Antenna for Biomedical and Wearable Applications

**Article** · January 2021

CITATIONS
0

READS
24

**3 authors**, including:

Anubhav Kumar
Raj Kumar Goel Institute of Technology and Management
**32** PUBLICATIONS **73** CITATIONS
SEE PROFILE

Rakesh Kumar Jain
Shobhit Institute of Engineering and Technology
**112** PUBLICATIONS **279** CITATIONS
SEE PROFILE

**Some of the authors of this publication are also working on these related projects:**




Eight Element 5G Antenna View project

All content following this page was uploaded by Anubhav Kumar on 06 June 2022.

The user has requested enhancement of the downloaded file.

# Two/Three-element textile Diversity MIMO Antenna for Biomedical and Wearable Applications

Anubhav Kumar [1], Asok De [2], R.K Jain [3]
[1 2 3] Department of Electronics and Communication Engineering
[1 3] Shobhit Institute of Engineering and Technology,(Deemed to be University), Meerut, Uttar Pradesh, India
[2] Delhi Technological University (DTU), New Delhi
[1] rajput.anubhav@gmail.com

Abstract – A compact two/three-port textile antenna with pattern diversity performance for wearable and biomedical communication is proposed. The antenna-1 is placed orthogonally with antenna-2 and antenna-3 on textile jeans material with 30x30x1 $mm^3$ dimension. The two symmetrical inverted L-shaped open-stubs between antenna-1 and antenna-2 as well as two L-shaped open-stubs behind the antenna-3 with the diagonal strip are used to perturb and mitigate the surface current between elements. The bandwidth of antenna-1 and antenna-2 is covering 5G and WLAN/ISM bands where antenna-3 covers a partial 5G frequency band. The envelope correlation coefficient (ECC) and channel capacity loss (CCL) are analyzed for Multi-input-Multi-output (MIMO) a antenna performance, which represent the diversity performance of the antenna.

**Keywords - Three-port antenna, Diversity antenna, Textile antenna, Wearable application, 5G communication**

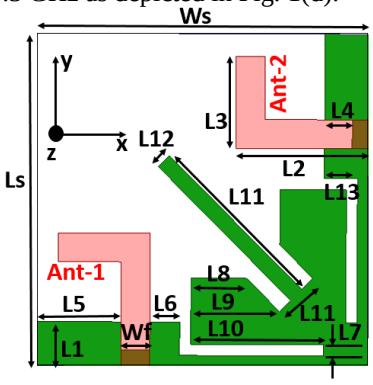
## I. Introduction

The requirements of data rates, channel capacity and bandwidth in communication channel increases with the new technology introduced in recent wireless communication. MIMO antenna can meet these demands more easily than a single antenna therefore the focus of researchers has been increased on MIMO antenna designing. In recent times, there has been an increase in demand for wearable antennae in portable, medical and sports applications. The wearable antenna should be flexible and compact for body communication applications so that it can easily accommodate sensors and batteries in external wearable devices. Therefore, the MIMO antenna must be compact and flexible for on-body communication applications for wireless communication [1-5]. In [5], defected ground based isolation enhancement is introduced in two-port circular polarized antenna. In a MIMO antenna, if the antenna elements are close in limited space then it reduces the overall size of the antenna but increases the correlation coefficient between the antenna elements, which also affects the channel capacity [6]. Pattern diversity, defected ground and open stubs in the ground helps to reduce surface current between antenna elements. Many materials are used to design a wearable antenna but the rigid material does not match the agreement of not being flexible for this application, therefore flexible material can make the antenna more comfortable for on-body communication. A study of the wearable antenna in the literature suggests that design of wearable antennae have been emphasized with flexible and compact antenna [1-5, 7-10]. In [7], an s-shaped radiator with co-axial feed is used to obtain a dual-band on jeans material where open stubs in the ground influence the impedance matching of the antenna. In [8], rectangular radiator antenna is modified to an E-shaped radiator antenna with the partial ground on flexible jeans material for medical and wearable applications. In [9], two rectangular radiators are used as a MIMO antenna for wideband and wearable communication accomplished on jeans substrate where dual I-shaped strip in the ground is used to enhance the isolation. In [10], a switch on the resonator is used for reconfigurable application in jeans substrate for wearable wideband applications, where an open-stub and slot in the ground is used for surface current elimination between ports. In [11], circular polarized and CPW-fed antenna is designed where diagonal decoupling structure and open stubs are used to perturb the current and increase the port isolation. In [12], inverted L-shaped microstrip feed is discussed to design 4-element antenna where modified CSRR is used to enhance the IBW of the antenna. In [13], beak-shaped radiators are introduced for sub-6 GHz bands where T-shaped decoupling with slots are accomplished to perturb the surface current. In [14], CPW-fed two port antenna is discussed where H-shaped EBG are used to perturb and mitigate the surface current. In [15], modified open stub with slits in open stub is introduced in the ground in two-port antennas, which improves the port isolation and diminishes the surface current. In [16], microstrip-feed based two-port flexible antenna is designed for WLAN where strip is used as decoupling structure which mitigate the surface current between ports and enhance the port isolation. Keeping in view the requirement of wearable antennas, in this paper, two and three-port antennas are designed in limited size. The 10 dB bandwidth of antenna-1 and antenna-2 varies from 4.4 GHz to 6.8 GHz, which is covered 5G and WLAN / ISM bands. The 10 dB impedance bandwidth (IBW) of antenna-3 varies from

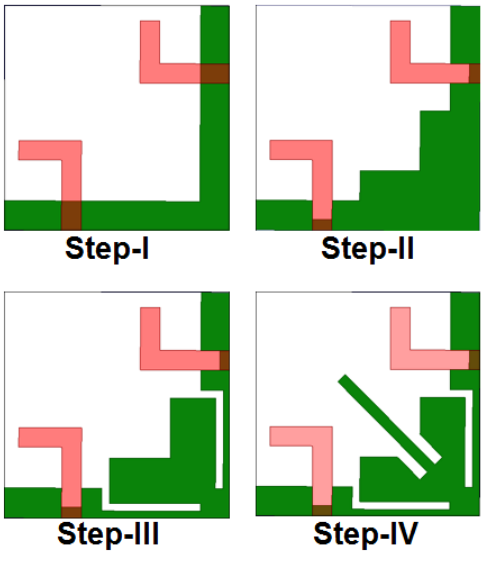
4.55 GHz to 4.8 GHz and can work for a 5G wearable application. Open Stub and diversity placement of antenna elements are used to improve the isolation in present technology where the isolation of antenna-1 with antenna-2 is more than 20 dB and the isolation of antenna-3 with antenna-1 is more than 19.7 dB whereas with antenna-2 is more than 16.7 dB. The ECC and CCL have been computed to analyze the diversity and MIMO characteristics of the antenna, which is under the acceptable limit, and make the antenna helpful for wearable applications with pattern diversity performance.

## II. Design of two-element antenna

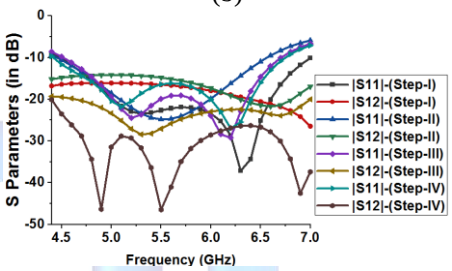
The prototype and design of the two-port MIMO antenna is illustrated in Fig. 1(a) where both antennas are placed orthogonally to each other. The flexible jeans material (permittivity (εr) = 1.7 with loss tangent (tan δ) = 0.025) is used to design the MIMO antenna with a copper tape of 0.057 mm thickness as a conducting element. The dimensions of two-part antenna are as follows in mm: Ws=30, Ls=30, Wf=2.6, L2=12, L4=2.5, L6=2.7, L8=5, L10=12, L12=1.414, L1=4, L13=3, L3=8.3, L5=7.7, L7=1, L9=8, L11=17, thickness (t) = 1 mm. The antenna evolution steps are divided into four parts as represented in Fig. 1(b), where its s-parameters are illustrated in Fig. 1(c). In the first step, two orthogonally placed inverted L-shaped microstrip feed radiator is accomplished with the partial ground. The IBW of the antenna obtains from 4.5 GHz to 7 GHz with port isolation is more than 15 dB. In step two, conducting part of the ground plane is enhanced with the defected ground below the microstrip feed. It enhances the impedance matching and provides the support to make changes in the ground for next step isolation enhancement. In step three, two symmetric L-shaped open stubs are introduced in the ground, which perturb and diminished the surface current. The IBW of the antenna varies from 4.5 GHz to 6.7 GHz with more than 19.5 dB isolation. In the fourth step, a diagonal strip is placed in the ground between antenna elements, which improves the isolation and maximizes it up to 46 dB. The IBW of the antenna varies from 4.4 GHz to 6.75 GHz where isolation varies from 20 dB to 46 dB. The 3D radiation pattern is analyzed to investigate the pattern diversity of the antenna at 5.1 GHz and 6.3 GHz as depicted in Fig. 1(d).



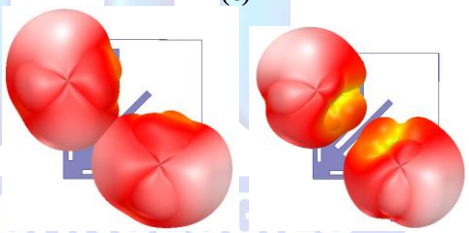

(a)




(b)




(c)



(i) (ii)

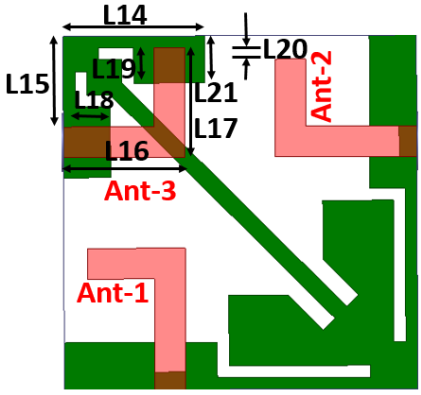
(d)

Figure 1. Textile two-element antenna (a) prototype design (b) design steps (c) s-parameters of design steps (d) 3D radiation pattern at (i) 5.1 GHz (ii) 6.3 GHz

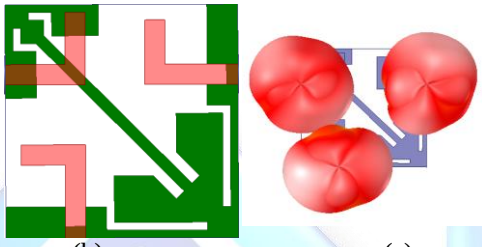
## III. Design of three-element antenna

In this section, designing of antenna-3 and its results are discussed. The Inverted L-shaped radiator is accomplished in the limited space of a two-port antenna, which is discussed in the previous section with pattern diversity performance. The common ground plane is used in all three-port antenna for keeping in the mind of practical applications as depicted in Fig. 2(a) and Fig. 2(b). The two symmetric L-shaped stubs behind antenna-3 are used for impedance matching and reduced the surface current from antenna-1 and antenna-2. The dimension of the antenna-3 is as follows in mm: L14=12, L16=10.3, L18=3, L20=1, L15=7.7, L17=9.3, L19=3, L21=4. The 10 dB impedance bandwidth (10 dB) of antenna-3 is 4.55 GHz to 4.8 GHz applicable for 5G communication. The isolation of antenna-3 from antenna-1 and antenna-2 is 19.7 dB to 22 dB and 16.7 dB to 25 dB as illustrated in Fig. 2(d). The 3D

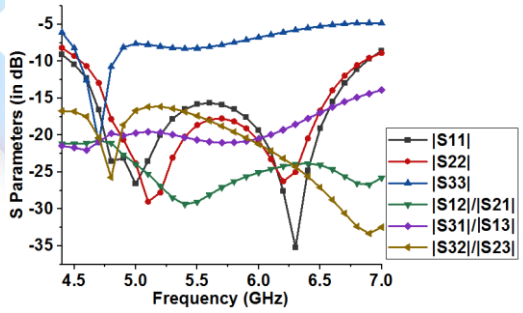
radiation patterns of all three port antenna is depicted in Fig. 2(c), where patterns are designed of antenna-1 and antenna-2 at 5.1 GHz and of antenna-3 at 4.7 GHz.




(a)



(b) (c)




(d)

Figure 2. Textile three-element antenna (a) prototype design with dimensions (b) final design (c) 3D radiation pattern at 5.1 GHz (Ant-1 and Ant-2) and at 4.7 GHz (Ant-3) (d) S-parameters

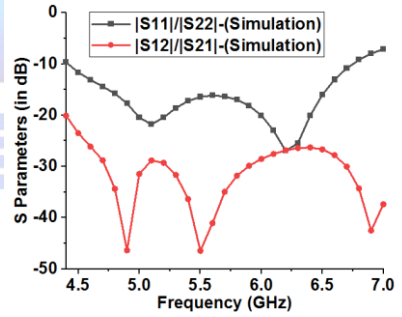
## IV. Simulated Results Analysis

The flexible two/three-port textile antenna is simulated in HFSS 13 and HFSS 17 software. The simulation investigation of the two-port antenna and three-port antenna is discussed in Fig. 3 and Fig. 4. The two-port antenna IBW varies from 4.4 GHz to 4.75 GHz with more than 20 dB isolation as described in Fig. 3(a). The realized gain (dB) varies from 2.4 to 5.4 in the entire operating bandwidth as illustrated in Fig. 3(b). The radiation efficiency (%) of the two-port antenna is above 93 in the operating bandwidth. The diversity and MIMO characteristics of the antenna is analyzed with ECC and CCL, which is calculated from equations 1 and equation 2 with simulated s-parameters [6].

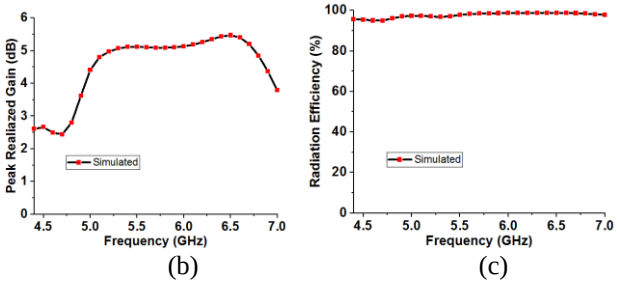
$$ECC = \frac{|S_{11}^{*}S_{12}+S_{21}^{*}S_{22}|^2}{(1-|S_{11}|^2-|S_{21}|^2)(1-|S_{22}|^2-|S_{12}|^2)} \quad (1)$$

$$CCL = -log_2 det(\psi^R) \quad (2)$$

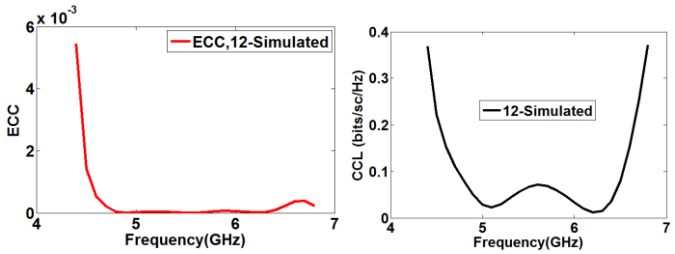
The ECC is analyzed and represents the correlation performance between antenna elements where its acceptable values should be less than 0.5. The CCL is extracted from equation 2, where $\psi^R$ is represented the correlation matrix at receiving end. The acceptable value of CCL is 0.4 bits/sec/Hz [6]. The ECC and CCL of the two-port MIMO antenna are less than 0.006 and 0.4 bits/s/Hz and under suitable limits, as depicted in Fig. 3(d) and Fig. 3(e). The s-parameter of the three-port antenna is represented in Fig. 4(a), where |S11| (dB) of antenna-1 and antenna-2 varies from 4.5 GHz to 7.8 GHz and 4.6 GHz to 7.8 GHz. The |S11| (dB) of antenna-3 varies from 4.55 GHz to 4.8 GHz. The port isolation |S12| is more than 20 dB where the port isolation |S13| is more than 19.7 dB and antenna-2 to antenna-3 is more than 16.7 dB. The gain (realized) of antenna-1 and antenna-2 is more than 2.2 dB where it is more than 1.7 dB for antenna-3. The radiation efficiency of all antenna elements is more than 84 % in operating bandwidth as discussed in Fig. 4(c). The ECC and CCL of the antenna-2 to antenna-3 is less than 0.02 and 0.4 bits/sec/Hz respectively and under the satisfactory limit, as depicted in Fig. 4(d) and Fig. 4(e). The radiation pattern (normalized) of antenna-1 at 5.1 GHz and of antenna-3 at 4.7 GHz is depicted in Fig. 5 and Fig. 6 in a different plane (xz and yz) where a stable radiation pattern is achieved. Table-1 is used to analyze different textile and MIMO antenna analysis where the proposed two/three-port antenna is compact with acceptable MIMO parameters and can be used for portable, wearable, sports, and biomedical applications.




(a)

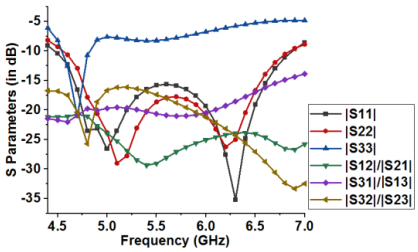



(b) (c)

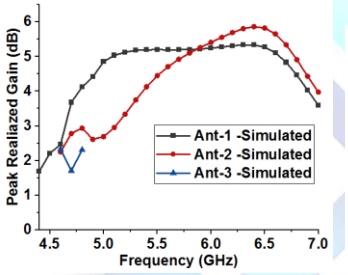



(d) (e)

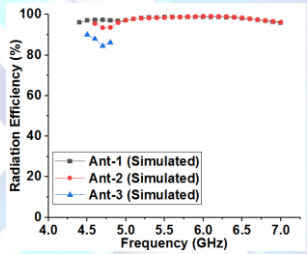
Figure 3. Two-port antenna (a) s-parameters (|S11| and |S12|) (b) realized gain (dB) (c) efficiency in % (radiation) (d) ECC (e) CCL



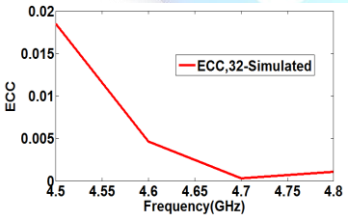

(a)



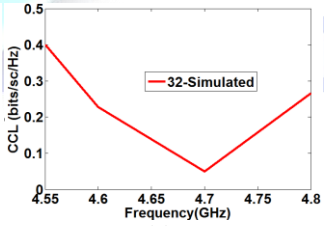

(b)




(c)

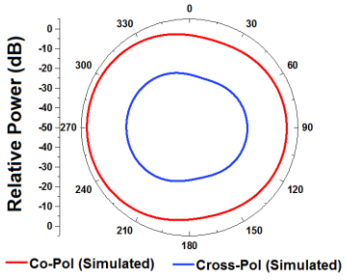



(d)

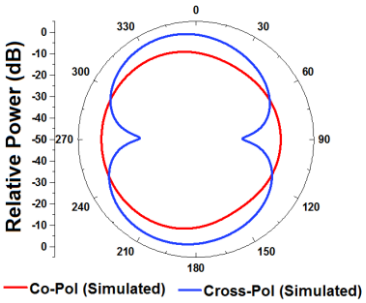



(e)

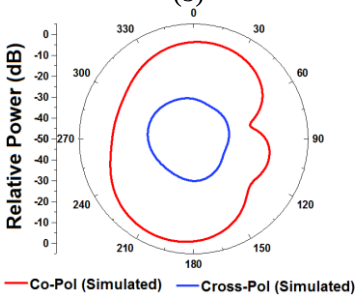
Figure 4. Three-port antenna (a) s-parameters (|S11| and |S12|) (b) realized gain (dB) (c) efficiency in % (radiation) (d) ECC (e) CCL




(a)

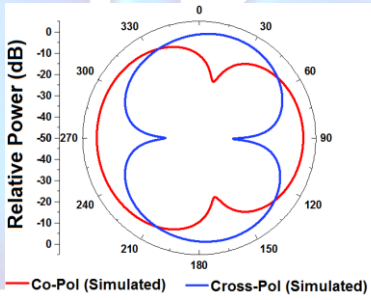



(b)

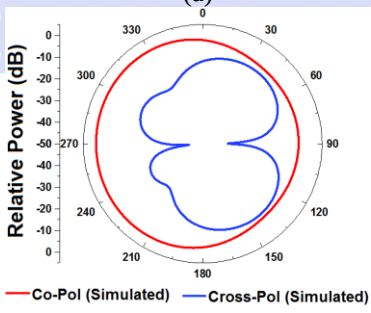



(c)

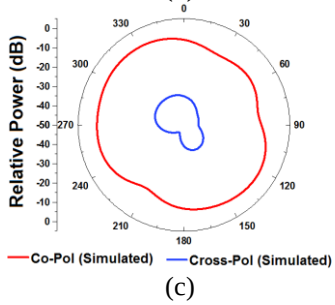
Figure 5. Normalized radiation pattern of Ant-1 at 5.1 GHz (a) in xz, (b) in yz (c) in xy plane




(a)




(b)




(c)

Figure 6. Normalized radiation pattern of Ant-3 at 4.7 GHz (a) in xz, (b) in yz (c) in xy plane

## V. Conclusions

A novel, compact two/three-element textile antenna is presented based on open-stubs and diagonal strip with low-cost flexible jeans material. The antenna-1 and antenna-2 cover 4.6 GHz to 6.8 GHz for 5G and WLAN/ISM bands with isolation of more than 20 dB. The antenna-3 covers 4.55 GHz to 4.8 GHz with isolation more than 19.7 dB from antenna-1 and 16.7 dB from antenna-2. The ECC between antenna-1 and antenna-2 is less than 0.006 and from antenna-2 to antenna-3 is less than 0.02 where CCL is under acceptable limits. The proposed antenna is low-cost, compact size, flexible and good diversity performance can be a good choice for biomedical, sports, portable, and wearable communications.

## References


1. Yalduz, Husnu, et al. "Design and analysis of low profile and low SAR full-textile UWB wearable antenna with metamaterial for WBAN applications." *AEU-International Journal of Electronics and Communications* 126 (2020): 153465.
2. Sundarsingh, Esther F., Malathi Kanagasabai, and Vimal Samsingh Ramalingam. "Completely integrated multilayered weave electro-textile antenna for wearable applications." *International Journal of Microwave and Wireless Technologies* 9.10 (2017): 2029.
3. Roudjane, Mourad, et al. "New generation wearable antenna based on multimaterial fiber for wireless communication and real-time breath detection." *Photonics*. Vol. 5. No. 4. Multidisciplinary Digital Publishing Institute, 2018.
4. El Gharbi, Mariam, et al. "A Review of Flexible Wearable Antenna Sensors: Design, Fabrication Methods, and Applications." *Materials* 13.17 (2020): 3781.
5. Kumar, Anubhav, Asok De, and R. K. Jain. "Circular Polarized Two-Element Textile Antenna with high isolation and Polarization Diversity for Wearable Applications." *International Journal of Microwave and Wireless Technologies* (2022).
6. Kumar, Anubhav. "Compact 4x4 CPW-Fed MIMO Antenna with Wi-Fi and WLAN Notch for UWB Applications." *Radioelectronics and Communications Systems* 64.2 (2021): 92-98.
7. Bhattacharjee, Shankar, et al. "Design of a miniaturized dual-band textile antenna using characteristic modal analysis for on-body applications." *Journal of Electromagnetic Waves and Applications* 32.18 (2018): 2415-2430.
8. Ashyap, Adel YI, et al. "Inverted E-shaped wearable textile antenna for medical applications." *IEEE Access* 6 (2018): 35214-35222.
9. Biswas, Ashim Kumar, and Ujjal Chakraborty. "A compact wide band textile MIMO antenna with very low mutual coupling for wearable applications." *International Journal of RF and Microwave Computer-Aided Engineering* 29.8 (2019): e21769.
10. Biswas, Ashim Kumar, and Ujjal Chakraborty. "Reconfigurable wide band wearable multiple input multiple output antenna with hanging resonator." *Microwave and Optical Technology Letters* 62.3 (2020): 1352-1359.
11. Kumar, Anubhav, Asok De, and Rakesh K. Jain. "Circularly polarized CPW fed MIMO/Diversity antenna for Wi-Fi and WLAN applications." *Frequenz* 76.1-2 (2022): 37-44.
12. Jha, Pankaj, et al. "Modified CSRR Based Dual-Band Four-Element MIMO Antenna for 5G Smartphone Communication." *Progress In Electromagnetics Research Letters* 101 (2021).
13. Kumar, Anubhav, Asok De, and R. K. Jain. "Size Miniaturization and Isolation Enhancement of Two-Element Antenna for Sub-6 GHz Applications." *IETE Journal of Research* (2021): 1-8.
14. Kumar, Anubhav, Asok De, and R. K. Jain. "Novel H-shaped EBG in E-plane for Isolation Enhancement of Compact CPW-fed Two-Port UWB MIMO Antenna." *IETE Journal of Research* (2021): 1-7.
15. Kumar, Anubhav, Asok De, and R. K. Jain. "Modified Open slot based Two-Element Compact Antenna for 5G and wearable applications." *2021 8th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE, 2021.
16. Jha, Pankaj, et al. "Flexible and Textile Two-port compact antenna for WLAN and wearable applications." *2021 8th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE, 2021.


Table.1 Comparison analysis of existing textile and MIMO antenna with proposed antenna

| Reference | Number of elements | Size (mm x mm) | Material | Bandwidth (GHz) | Isolation | ECC |
|---|---|---|---|---|---|---|
| [5] | 2 | 31x31 | Jeans | 3.3 to 4.3 | >24 | <0.2 |
| [7] | 1 | 60×30 | felt | 2.45 and 5.8 | NA | NA |
| [8] | 1 | 30 × 20 | Denim | 2.4 | NA | NA |
| [9] | 2 | 40 × 70 | Jeans | 2.4-8.0 | >22 | <0.01 |

| | | | | | | |
|---|---|---|---|---|---|---|
| [10] | 2 | 35 × 30 | Jeans | 3.11-5.15, 4.81-7.39 | >19.5, 21 | < 0.18 |
| [11] | 2 | 30 × 30 | FR-4 | 4.95 to 5.95 | >20 | <0.005 |
| [12] | 4 | 170 x 80 | FR-4 | 3.40-3.625, 3.90-4.55 | >20.1 | <0.3 |
| [13] | 2 | 40 × 30 | FR-4 | 2.12–3.85 and 4.95–6 | >20 | <0.002 |
| [14] | 2 | 35 × 22 | FR-4 | 3.4–10.4 | >20 | <0.01 |
| [15] | 2 | 30 x 24 | FR-4 | 3.15 to 3.95 | >22 | <0.004 |
| [16] | 2 | 26x24 | Jeans | 4.9-6 | >27 | <0.002 |
| **Proposed** | **2** | **30x30** | **Jeans** | **4.4 to 6.75** | **>20** | **<0.006** |
| **Proposed** | **3** | **30x30** | **Jeans** | **4.6 to 6.8, 4.55 to 4.8** | **>20, 16.7** | **<0.02** |

View publication stats

Endocrine
https://doi.org/10.1007/s12020-022-03125-5

ORIGINAL ARTICLE

# Type-2 Diabetes Mellitus (T2DM): Spatial-temporal Patterns of Incidence, Mortality and Attributable Risk Factors from 1990 to 2019 among 21 World Regions

Mehak Nanda[1] · Rajesh Sharma[1] · Sumaira Mubarik[2] · Aashima Aashima[1] · Kai Zhang[3]

Received: 15 November 2021 / Accepted: 17 June 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

**Purpose** Type-2 diabetes Mellitus (T2DM) is one of the leading causes of death and disability worldwide. This study examines temporal patterns of the global, regional, and national burden of T2DM in the last three decades.

**Data and methods** The estimates of age, sex and location-wise incident cases, deaths, prevalent cases, and disability-adjusted-life-years (DALYs) and risk factors for 21 regions and 204 countries are retrieved from the Global Burden of Disease 2019 study from 1990 to 2019. Socio-demographic index (SDI) is used as the indicator of the development status of countries, and quadratic regression is employed to examine the relationship between country-level age-standardized rates and SDI.

**Results** Globally, incident cases of T2DM more than doubled from 8.4 million[95% uncertainty interval, 7.8–9.1 million] in 1990 to 21.7 million[20.0–23.5 million] in 2019, and deaths more than doubled from 606,407[573,069–637,508] to 1.5 million[1.4–1.6 million] between 1990 and 2019. Global T2DM prevalence increased from 148.4 million[135.5–162.6 million] in 1990 to 437.9 million[402.0–477.0 million] in 2019. In 2019, global age-standardized prevalence rate stood at 5282.8/100,000[4853.6–5752.1], varying from 2174.5/100,000[1924.3–2470.5] in Mongolia to 19876.8/100,000[18211.1–21795.3] in American Samoa. SDI exhibited inverted-U shaped relationship with country-level age-standardised rates. Globally, high body-mass-index (51.9%), ambient particulate matter pollution (13.6%), smoking (9.9%) and secondhand smoke (8.7%) were the major contributing risk factors towards T2DM DALYs in 2019.

**Conclusion** With ubiquitously rising prevalent cases globally, particularly in low and middle-income countries and regions, T2DM requires immediate attention and targeted policy response worldwide centered on lifestyle interventions (e.g., physical activity, smoking, diet, and obesity), air pollution control and cost-effective timely treatment.

**Keywords** Type-2 Diabetes Mellitus · Incidence · Mortality · Disability Adjusted Life Years · Global Burden of Disease

## Introduction

Diabetes mellitus (DM) is one of the most prominent chronic diseases worldwide, and its incidence and prevalence has snowballed in recent years, primarily led by low and medium-income countries (LMICs, 79% of global prevalent cases) [1]. In adults, type-2 diabetes mellitus (T2DM) is the most common DM worldwide, accounting for 90% of DM prevalent cases [1]. The pathogenesis of T2DM involves insulin resistance and $\beta$-cell dysfunction, which leads to increased glucose production in the liver, decreased glucose uptake in muscles and adipose tissue, and insufficient insulin release [2, 3]. Its progression is associated with microvascular complications (e.g., retinopathy, nephropathy, and neuropathy) and macrovascular outcomes

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1007/s12020-022-03125-5.

✉ Rajesh Sharma
rajesh.sharma@dtu.ac.in

[1] University School of Management and Entrepreneurship, Delhi Technological University, Delhi, India

[2] Department of Epidemiology and Biostatistics, School of Health Sciences, Wuhan University, Wuhan 430071, China

[3] Empire Innovation Associate Professor, Department of Environmental Health Sciences, School of Public Health | University at Albany, State University of New York, Albany, USA

Published online: 16 July 2022

(e.g., coronary heart disease, ischaemic stroke, and other vascular diseases) [4, 5]. Due to its complications and long-term treatment requirements, the global estimated cost of DM was $1.3 trillion in 2015 [6].

A United Nation (UN) high-level meeting on non-communicable diseases in 2011 set a target of 2025 to halt the rise in adult prevalence rate of DM to its 2010 levels [7]. As T2DM constitutes 90% of DM prevalent cases, such targets cannot be attained till the rising incidence and prevalence of T2DM are checked. Towards this end, the public health response against T2DM, framing policy and devising interventions require a continuous and updated assessment of T2DM epidemiology and its temporal patterns.

In this study, we examined the global, regional, and national burden of T2DM, drawing upon estimates from the Global Burden of Diseases, Injuries and Risk Factors Study (GBD) 2019 for 204 countries from 1990 to 2019 [8, 9]. We presented and compared country and region-wise estimates of T2DM burden using a battery of epidemiological metrics: incidence, mortality, prevalence, and disability-adjusted life years (DALYs). In addition to widely reported and studied epidemiological metrics (i.e., incidence, mortality, and prevalence), we also reported and examined the DALYs count to gauge both fatal and non-fatal burden of T2DM at the country and regional levels [9, 10]. The development status of a country was measured using socio-demographic index (SDI). The association between age-standardized rates and SDI was examined using quadratic regression. Lastly, we examined the risk-attributable burden of 16 major risk factors in T2DM burden at the global and regional level. A previous study [11] also examined the worldwide prevalence of DM (inclusive of type-1 and type-2) from 1980 to 2014. However, our study specifically focuses on the estimates of T2DM, examining all dimensions such as incidence, deaths and DALYs in addition to prevalence. We believe that an assessment of temporal patterns of T2DM burden and associated risk factors at global, regional, and national levels can help policymakers to track progress (or lack of) against this growing epidemic and devise strategies accordingly.

## Materials and methods

### GBD 2019 Estimates

We retrieved estimates of incidence, mortality, prevalence, and DALYs from the GBD 2019 study, which provided age, sex, year and location-wise estimates for 204 countries for 1990-2019 [8, 9, 12]. To arrive at cause-specific estimates, GBD employs all available data from sources such as vital registration, verbal autopsies, surveillance data, and disease registry data [8, 9]. The GBD estimates are available for 21 GBD-defined geographical regions and 204 countries for 20 non-overlapping age-groups from below-5 to 95-plus. The age-standardized rates are expressed as counts per 100,000 person-years and calculated using GBD world age population standard [9]. GBD employs the cause of death ensemble model (CODEm) developed by Foreman et al. [13] to obtain cause-specific mortality estimates. The GBD estimation framework and calculation of various metrics is detailed elsewhere [8, 9]. The GBD estimates reported in this study are provided with 95 percent uncertainty intervals (UIs), which reflect uncertainties due to difference in data availability, data quality, and modelling; we report these UIs inside square brackets along with mean estimates.

### Risk Factors

The T2DM burden attributable to different risk factors is based on a comparative risk assessment framework and the details of GBD 2019 risk factors methodology are available elsewhere [14]. The population attributable fraction (PAF) of a risk-outcome pair is calculated on the basis of relative risk, population exposure to a risk factor and theoretical minimum risk exposure level of a risk factor [14]. The PAF is then multiplied with age, sex and location-specific DALYs and deaths to calculate the risk-attributable burden. In this study, we examined the percentage contribution of 16 major risk factors in T2DM DALYs, which had non-zero contributions in the T2DM DALYs and deaths. These 16 risk factors are: alcohol use, ambient particulate matter (PM) pollution, diet high in processed meat, diet high in red meat, sugar-sweetened beverages, diet low in fiber, fruits, nuts/seeds, whole grains, high body mass index (BMI), high temperature, household air pollution from solid fuels, low physical activity, low temperature, secondhand smoke, and smoking.

### Statistical Analysis

We calculated the annual average percentage change (AAPC) in the T2DM burden using Joinpoint regression software 4.9.0.0. The AAPC is calculated based on the following regression.

$$\ln(P) = \alpha_0 + \alpha_1 y$$

P is the metric for which AAPC is calculated in the above equation and y is time (years). The AAPC is calculated as per the following formula.

$$\text{AAPC} = \exp(\alpha_1) - 1.$$

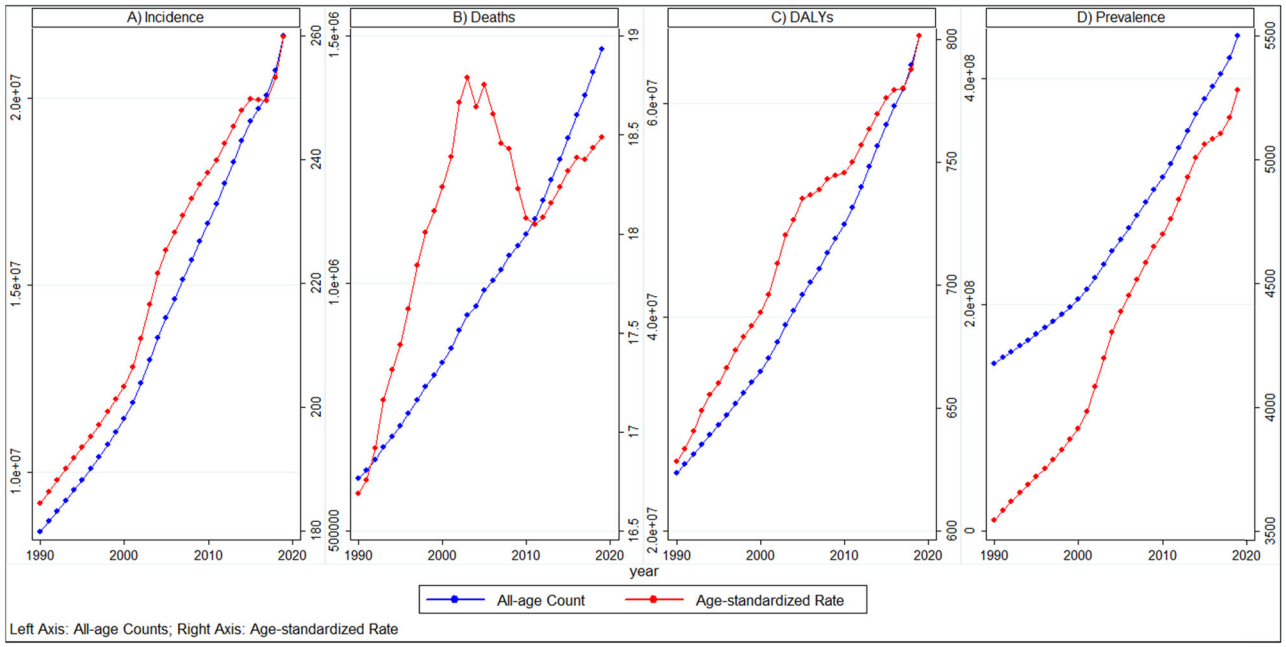



**Fig. 1** Temporal Movement of Age-standardized Rates from 1990 to 2019 **A**) Incidence **B**) Deaths **C**) DALYs **D**) Prevalence. T2DM: Type-2 Diabetes Mellitus; DALYs: Disability-adjusted-life-years. Left axis in all the graphs represents absolute counts and right axis represents age-standardized rates (per 100,000 person-years). Data Source: Global Burden of Disease Study 2019

The relationship between age-standardized rates of T2DM and SDI was explored using quadratic regression as specified below.

$$y_i = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \varepsilon_i$$

In the above equation, $y_i$ denotes country-level age-standardized rate in 2019, $x_i$ denotes the country-level SDI in 2019 and $\varepsilon_i$ denotes the residual term. The SDI is a measure of the development status of a country and this measure is developed by the GBD [12]. SDI summarizes three development measures: lag distributed income per capita, educational attainment, and total fertility rate. All three measures are scaled from 0 (lowest) to 1 (highest), and the geometric mean of these three measures provides the final SDI value of a country. Based on SDI values in 2019, the countries are divided into five quintiles by the GBD: low SDI, low-middle SDI, middle SDI, high-middle SDI and high SDI. All analysis is done in Stata 13.1, Python 3.6 and R 4.1.1.

## Results

### Global Trends, 1990–2019

Globally, T2DM incident cases more than doubled from 8.4 million[95% UI, 7.8–9.1 million] in 1990 to 21.7 million[20.0–23.5 million] in 2019 at AAPC of 3.4% (95% Confidence Interval (CI), 3.3–3.4%), whereas death count increased from 606,407[573,069–637,508] to 1.47 million[1.37–1.57 million] (Fig. 1A–B) at AAPC of 3.0% (95% CI, 3.0–3.1%) between 1990 and 2019. Globally, DALYs due to T2DM increased from 25.5 million[21.7–29.8 million] in 1990 to 66.3 million[55.5–79.0 million] in 2019 at AAPC of 3.3% (95% CI, 3.3–3.4%). Global prevalent cases of T2DM increased from 148.4 million[135.5–162.6 million] in 1990 to 437.9 million[402.0–477.0 million] in 2019. Globally, the age-standardized incidence rate (ASIR) increased from 184.6/100,000[170.9–199.7] in 1990 to 259.9/100,000[240.3–281.4] in 2019 at AAPC of 1.2% (95% CI, 1.2–1.3%), whereas age-standardized mortality rate (ASMR) increased from 16.7/100,000[15.7–17.5] in 1990 to 18.5/100,000[17.2–19.7] in 2019 at AAPC of 0.3% (95% CI, 0.2–0.4%). The age-standardized DALYs rate (ASDALR) increased from 628.3/100,000[537.2–730.9] in 1990 to 801.5/100,000[670.6–954.4] in 2019 at AAPC of 0.8% (95% CI, 0.8–0.9%). The age-standardized prevalence rate (ASPR) increased from 3546.5/100,000[3243.7– 3862.7] to 5282.8/100,000[4853.6–5752.1] from 1990 through 2019 (Fig. 1D).

### Temporal Patterns by Region, 1990–2019

In 2019, out of 21 regions, South Asia was ranked first with 5.1 million[4.6–5.6 million] new cases, registering a growth of 242.6%[234.2–250.8%] between 1990 and 2019

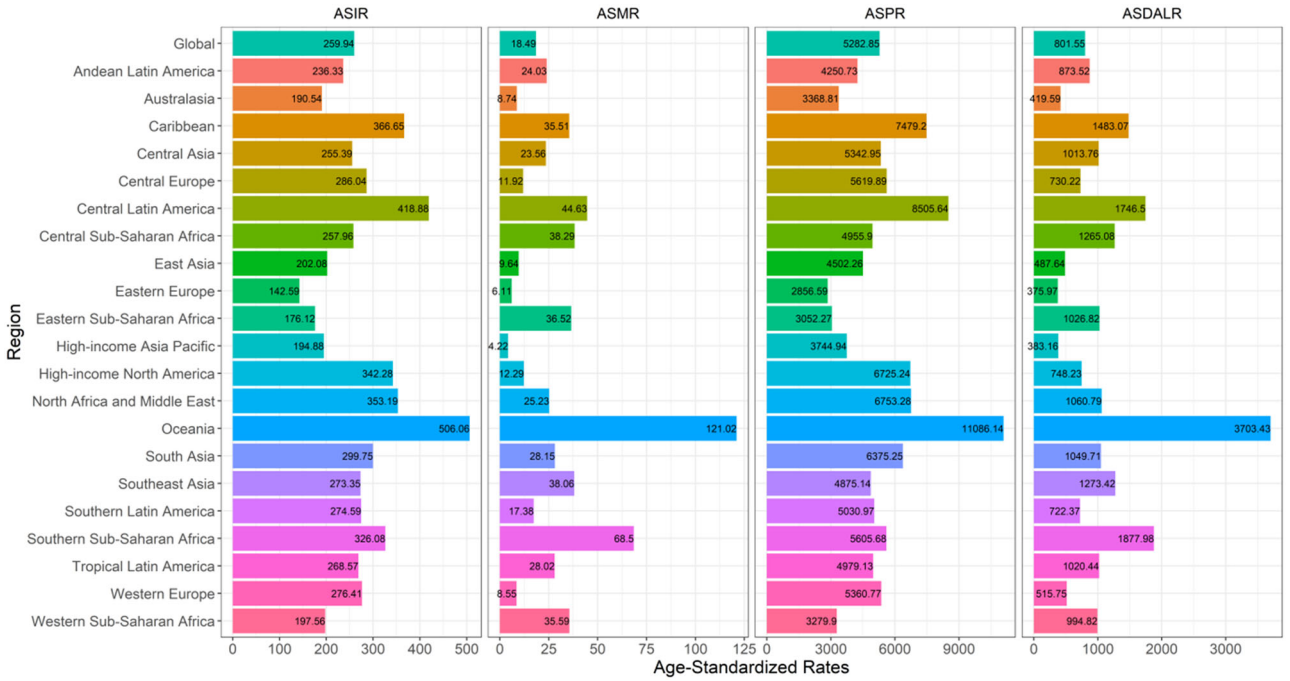



**Fig. 2** Region-wise Age-standardised Rates of T2DM in 2019 **a**) ASIR **b**) ASMR **c**) ASPR **d**) ASDALR. T2DM: Type-2 Diabetes Mellitus. ASIR: Age-standardized incidence rate (per 100,000 person-years); ASMR: Age-standardized mortality rate (per 100,000 person-years); ASPR: Age-standardized prevalence rate (per 100,000 person-years); ASDALR: Age-standardized DALYs rate (per 100,000 person-years). Data Source: Global Burden of Disease Study 2019

(Supplementary Table 1; Supplementary Figure 1). Incident cases doubled or more in 16/21 regions between 1990 and 2019, with North Africa and Middle East recording the highest growth (354.5%[341.0–367.5%]) and Eastern Europe (45.4%[41.1–49.7%]) witnessing the least growth in T2DM incident cases (Supplementary Figure 1).

In terms of death counts, South Asia was ranked first with 337,920 [300,995–379,757] deaths in 2019, with a sharp upsurge of 255.1%[198.8–316.8%] between 1990 and 2019 (Supplementary Table 1; Supplementary Figure 1). In 2019, East Asia was ranked third with 182,874[157,802–208,399] deaths behind South Asia (337,920 [300,995–379,757]) and Southeast Asia (214,764[193,046–236,296]). From 1990 through 2019, T2DM deaths doubled or more in majority of regions, with Central Asia (302.3%[260.5–348.0%]) registering the highest growth and Western Europe (18.9% [11.0–24.6%]) recording the minimum growth (Supplementary Figure 1). In 2019, prevalent cases were the highest in Asian regions, led by South Asia (98.1 million[88.4–108.8 million]); the highest increase between 1990 and 2019, however, was experienced by North Africa and Middle East (373.1%[361.4–385.0%]), followed by Andean Latin America (341.5%[323.9–359.2%]) (Supplementary Table 1; Supplementary Figure 1).

In 2019, the ASIR varied four-folds across regions, spanning from 142.6/100,000[131.0–156.0] in Eastern Europe to 506.1/100,000[472.9–542.4] in Oceania (Fig. 2). The ASMR disparities were much starker than ASIR, with 29-folds variation across regions, spanning from 4.2/100,000[3.7–4.6] in high-income Asia Pacific to 121.0/100,000[100.2–146.5] in Oceania. The ASDALR varied from 376.0/100,000[295.1–468.2] in Eastern Europe to 3703.4/100,000[3060.0–4399.3] in Oceania. The ASPR varied 4-folds between regions, from 2856.6/100,000[2582.5–3157.0] in Eastern Europe to 11086.1/100,000[10196.7–12087.4] in Oceania (Fig. 2).

## T2DM Age-standardized Rates by Sex

Sex-wise age-standardized rates of T2DM at global and regional levels in 2019 are presented in Supplementary Figure 2. Globally, ASIR among males (273.7/100,000) was higher than females (246.1/100,000) and other age-standardized rates were also higher in males than females. On a similar theme, apart from a few regions, males had higher age-standardized rates than females in majority of regions. For instance, in southern SSA and Central Asia, the ASIR and ASPR among females were higher than males; however, ASMR and ASDALR were slightly higher in males than females. By contrast, North Africa and Middle East, Caribbean, and Eastern Europe had higher ASIR and ASPR among males and higher ASMR and ASDALR among females (Supplementary Figure 2).

## T2DM Burden by Country

In 2019, India was the leading country with 4.2 million[3.8–4.6 million] new cases and 258,927[224,420–300,779] deaths, followed by China with 3.7 million[3.4–4.1 million] new cases and 168,388[143,233–194,030] deaths (Supplementary Table 2). T2DM prevalent cases were again the highest in China (90.0 million[82.3–98.5 million]), followed by India (81.8 million[73.6–91.0 million]). With high incidence and mortality, India and China were again the leading countries in terms of DALYs, with 12.0 million[9.8–14.5 million] and 9.6 million[7.6–11.9 million] DALYs, respectively, in 2019 (Supplementary Table 2).

Figure 3 shows geographically the variation between countries in terms of age-standardized rates of T2DM in 2019. The ASIR varied eight-folds across countries, spanning from 104.8/100,000[94.2–116.7] in Mongolia to 819.4/100,000[762.7–882.2] in American Samoa; the ASMR, however, had much larger variation across countries, spanning from 2.0/100,000[1.7–2.1] in Japan to 257.4/100,000[210.3–309.2] in Fiji (Fig. 3). The ASPR varied from 2174.5/100,000[1924.3–2470.5] in Mongolia to 19876.8/100,000[18211.1–21795.3] in American Samoa. The ASDALR varied from 278.2/100,000[220.6–345.7] in France to 6884.3/100,000[5667.7–8214.8] in Fiji.

Supplementary Figure 3 presents the temporal patterns of incident cases and ASIR in seven countries (China, India, Unites States, Brazil, Mexico, Germany, and Indonesia), which had more than 10 million prevalent cases of T2DM in 2019. Among these seven countries, the incident cases increased the most in India (242.8%[233.5–251.8%]), whereas Germany (63.5%[46.7–76.9%]) registered the minimum rise between 1990 and 2019. The ASIR increased the most in India (60.1%[55.9–64.5%]), whereas Mexico (8.8%[4.4–13.8%]) registered the minimum increase between 1990 and 2019.

## T2DM Age-Standardised Rates and SDI

Figure 4 displays the bivariate association between country-specific age-standardized rates and SDI in 2019. Notably, all metrics have an inverted-U-shaped relationship with SDI, implying that as SDI increases, the age-standardized rates worsen (increase) initially, and in the later stages of development, the age-standardized rates decrease. As per quadratic regression, the turning point is observed for SDI value ranging 0.5097 - 0.6529 for different models. However, these models have low statistical power with $R^2 < 0.12$ for all bivariate relationships, such that several countries recorded age-standardized rates much higher or much lower than predicted by the bivariate models.

## Age-specific Incidence and Mortality

In 2019, global T2DM incident cases increased with age and plateaued at age-groups of 50–54 and 55–59 years and declined thereafter; in terms of death counts; however, those between 60–89 years accounted for the bulk of global deaths (Fig. 5). Deaths in males were higher than in females in below-65 age groups, and the pattern reversed thereafter (Fig. 5A-B). Between 1990 and 2019, T2DM incident cases and deaths increased in majority of age groups, whereas the age-specific incidence rate (per 100,000) increased in both males and females up to 70–74 years and declined thereafter (Supplementary Figure 4). In terms of age-specific mortality rate (per 100,000), most age groups witnessed decline in females up to the age group of 70–74 years and a significant increase occurred among those aged 75 years and above. The age-specific mortality rates increased in majority of age groups, with increases in males being higher than females in all age groups (Supplementary Figure 4D).

## Risk Factors

Figure 6 presents the temporal patterns of risk-attributable T2DM burden of 16 risk factors at the global level between 1990 and 2019. It shows that all-age DALYs increased due to majority of risk factors, whereas ASDALR (per 100,000 person-years) increased due to all but four risk factors (smoking, low temperature, household air pollution from solid fuels, and diet low in fiber) in the study period; for these four risk factors, the ASDALR increased (or remained stagnant) till mid-2000 and declined thereafter. For most dominant risk factors such as high BMI and ambient PM pollution, both all-age counts and ASDALR almost doubled between 1990 and 2019.

Different SDI categories have varying temporal patterns of risk-attributable DALYs and ASDALR (Supplementary Figures 5a-e). For instance, the ASDALR due to ambient PM pollution decreased in high SDI quintiles and increased in all other SDI quintiles between 1990 and 2019. Notably, all-age DALYs and ASDALR due to high BMI and low physical activity increased in all SDI quintiles. The risk-attributable burden of T2DM due to majority of dietary risk factors worsened in majority of SDI quintiles.

In order to gauge risk-attributable burden in 21 world regions, we examine the region-wise percent contribution of 16 risk factors in T2DM DALYs in 2019 (Supplementary Figure 6). Globally, high BMI (51.9%), ambient PM pollution (13.6%), smoking (9.9%), and secondhand smoke (8.7%) were the major contributing risk factors. High BMI was the most significant risk factor in majority of the regions. Alcohol use has negative contribution (−0.60%) in T2DM DALYs globally as well as in several regions, which

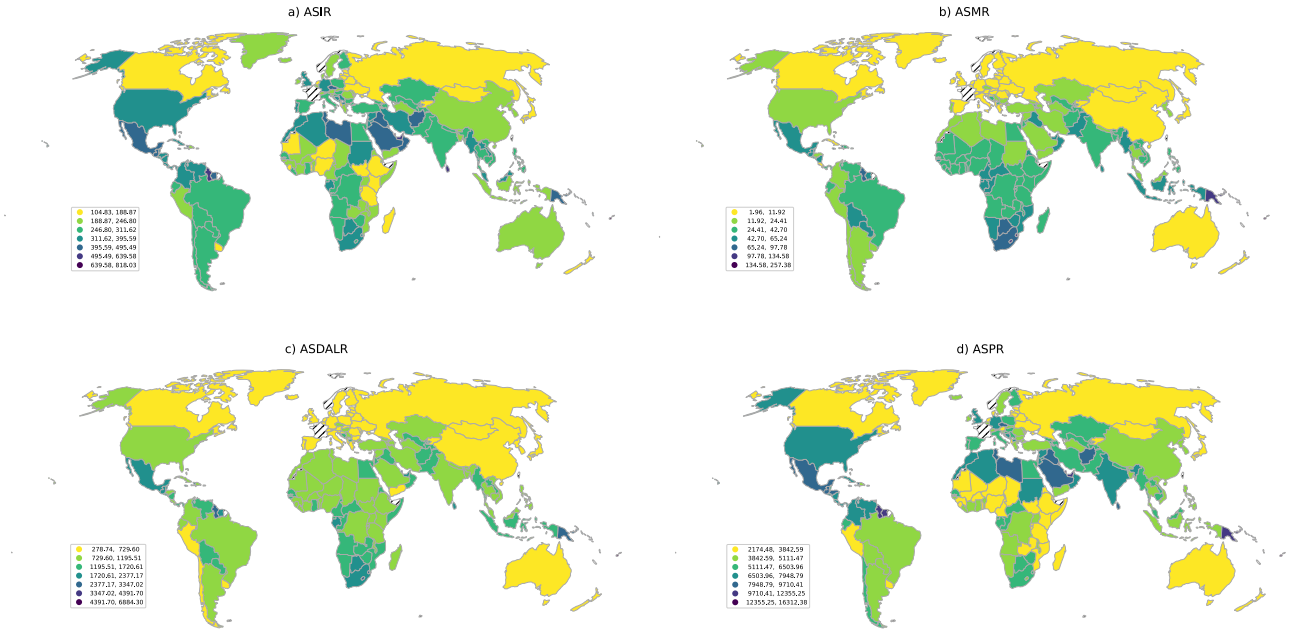



**Fig. 3** Country-wise age-standardized rates of T2DM in 2019 **a**) ASIR **b**) ASMR **c**) ASDALR **d**) ASPR. T2DM: Type-2 Diabetes Mellitus; ASIR: Age-standardized incidence rate (per 100,000 person-years); ASMR: Age-standardized mortality rate (per 100,000 person-years); ASPR: Age-standardized prevalence rate (per 100,000 person-years); ASDALR: Age-standardized DALYs rate (per 100,000 person-years). Data Source: Global Burden of Disease Study 2019

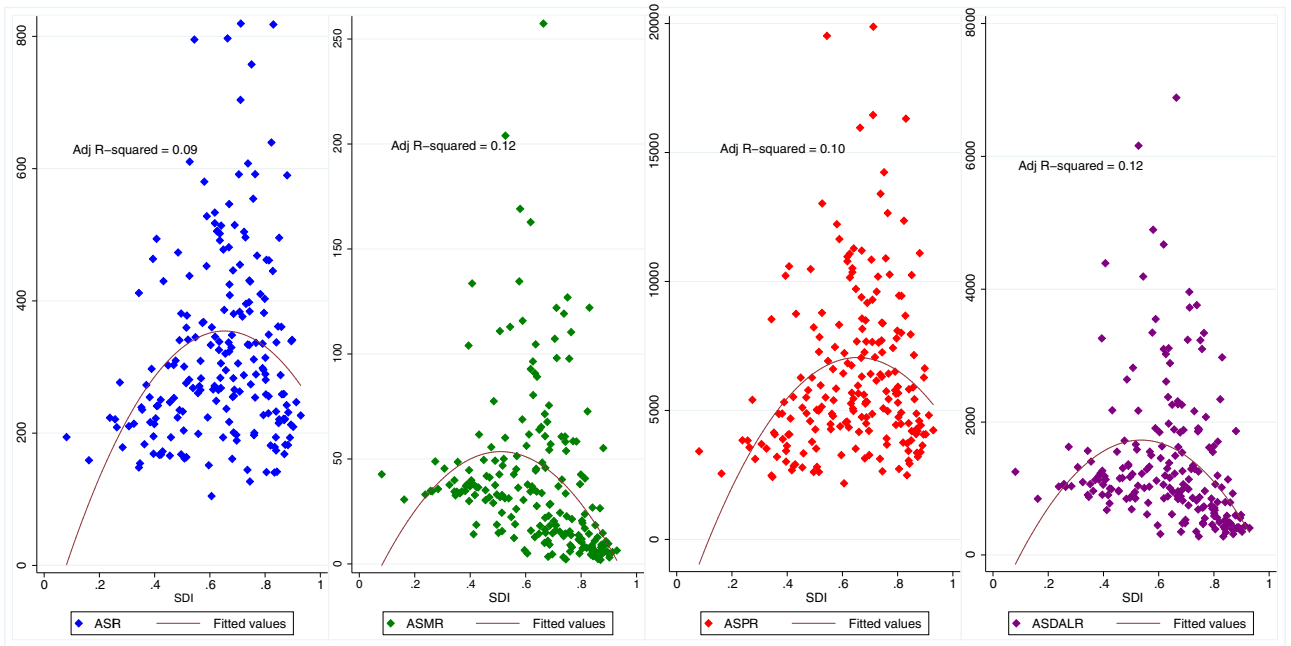



**Fig. 4** Bivariate association between Age-standardised Rates of T2DM and SDI in 2019. T2DM: Type-2 Diabetes mellitus; SDI: Socio-demographic index. ASIR: Age-standardized incidence rate (per 100,000 person-years); ASMR: Age-standardized mortality rate (per 100,000 person-years); ASPR: Age-standardized prevalence rate (per 100,000 person-years); ASDALR: Age-standardized DALYs rate (per 100,000 person-years). Data Source: Global Burden of Disease Study 2019. The fitting line is based on quadratic fit regression

suggests that it might have some protective effect against T2DM. In Australasia, apart from high BMI (61.8%), diet high in red meat (14.8%) and processed meat (13.6%) were the leading risk factors. In high-income North America, high BMI (68.8%), diet high in processed meat (15.8%) and red meat (11.5%), and smoking (12.8%) were the leading risk factors. Diet low in fruits contributed ~10% DALYs in a few regions of Asia and Africa, such as South Asia (10.8%) and Southern SSA (9.4%). Household air pollution from solid fuels was one of the significant risk factors in

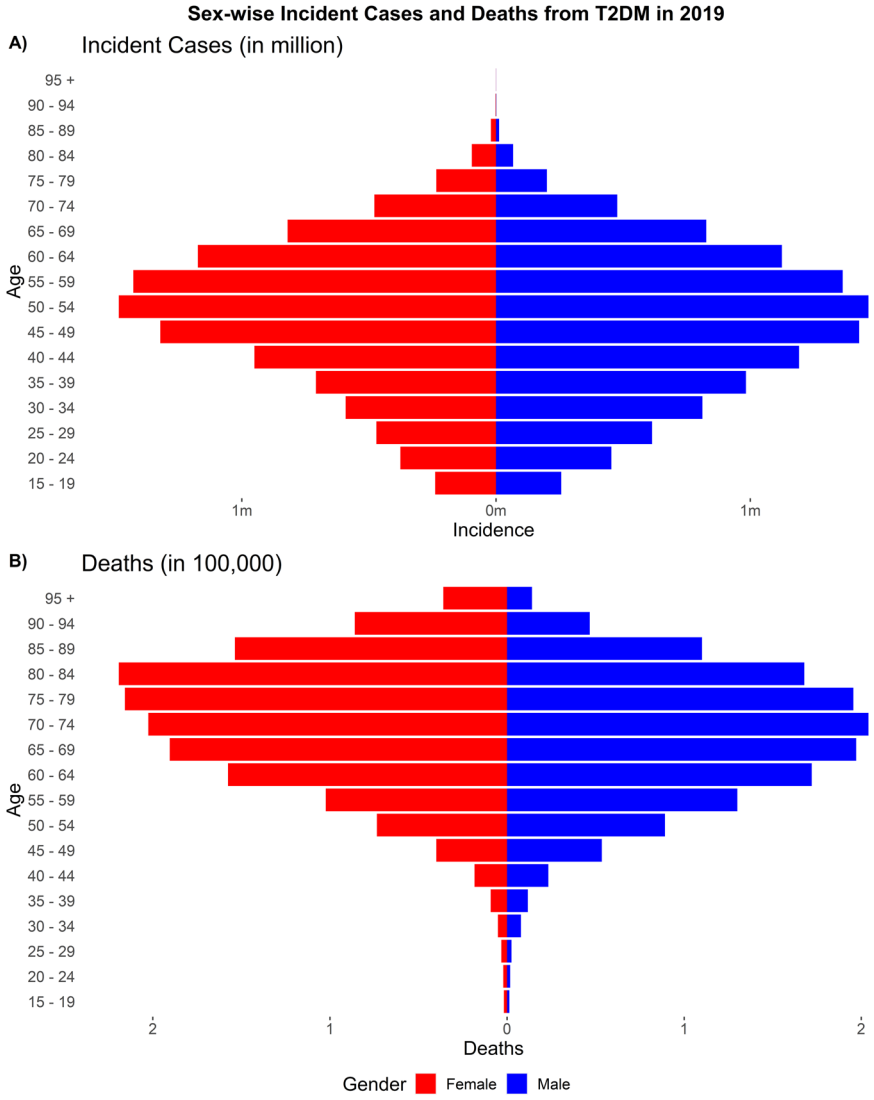



**Fig. 5** Global Age and Sex-wise T2DM Burden in 2019 **A**) Incidence **B**) Deaths. T2DM: Type-2 Diabetes Mellitus. Data Source: Global Burden of Disease Study 2019

regions of SSA: Eastern SSA (21.4%), Central SSA (17.8%), and Western SSA (16.0%).

## Discussion

This study examined the temporal patterns of T2DM burden in 21 regions and 204 countries. We found that incident cases doubled or more in 16/21 regions between 1990 and 2019. Country-level development measured using SDI had an inverted-U shaped relationship with age-standardised rates with a turning point of SDI ranging between 0.5097 and 0.6529. It implies that as countries develop, the age-standardized rates of T2DM worsen initially and towards later stages of development, the age-standardised rates fall, implying possibly better access to care. High BMI, smoking, ambient PM pollution and a host of dietary risk factors are the most prominent risk factors of T2DM at the global and regional levels.

We found higher age-standardized rates in males than females across majority of regions. High age-standardized rates of DM in males were also observed in previous studies [1, 11, 15]. A study of 51,920 males and 43,137 females found that males develop T2DM at lower BMI than females [16]. A later study suggested that more than BMI, higher visceral fat in males than females is a better predictor of higher T2DM burden among males [15]. Moreover, visceral adiposity is stronger risk factor than BMI towards glucose intolerance, insulin resistance, metabolic perturbations and hyperinsulinemia [17–19]. Several fundamental aspects of energy and metabolic homeostasis are also differently regulated in males and females [20, 21]. Sex differences are also present in the occurrence and progression of diabetic complications; therefore, diabetes care and management require more sex-specific approaches [20].

Asia has become the epicenter of the T2DM epidemic, with China and India becoming the global leaders

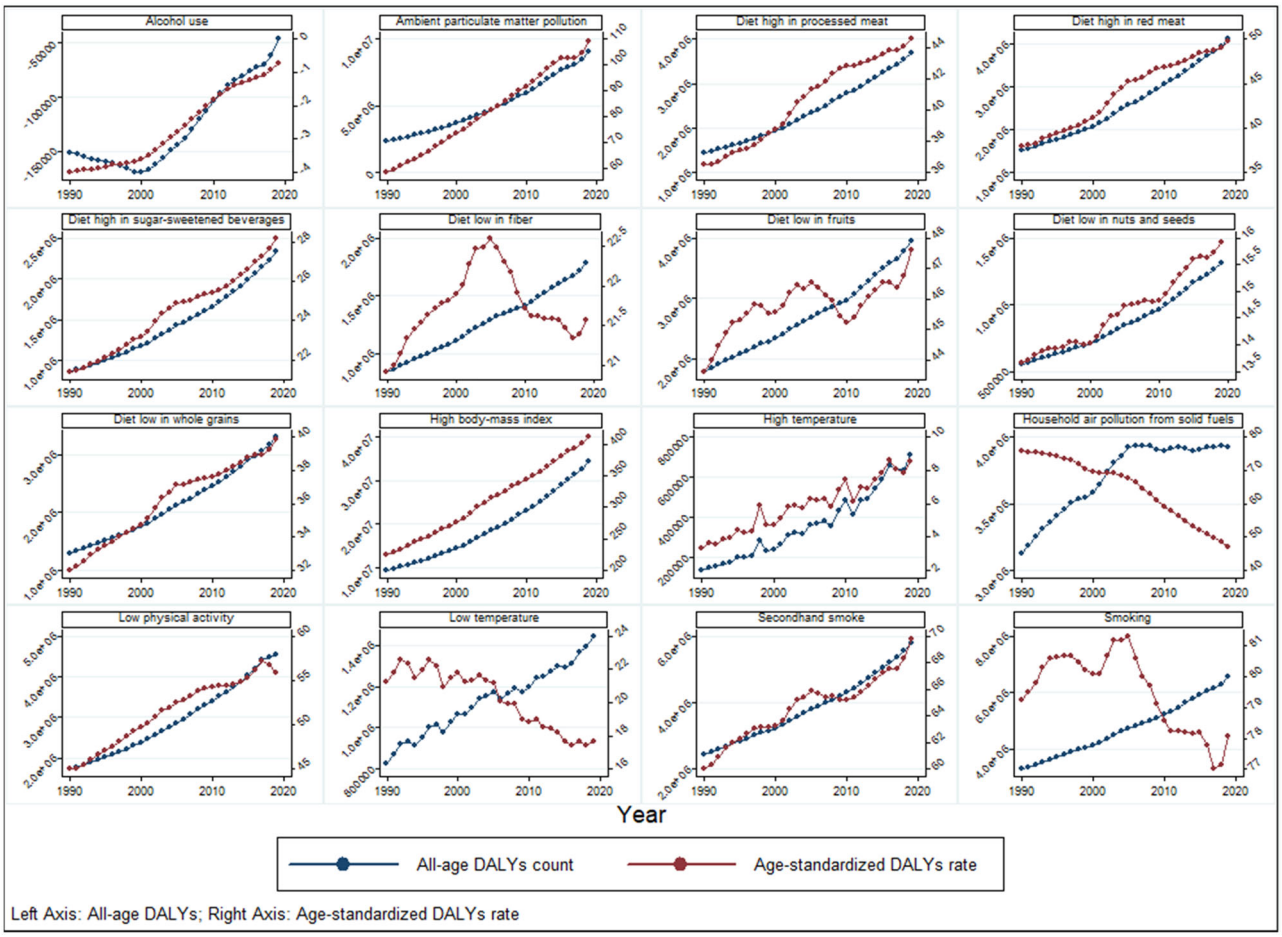



**Fig. 6** Global Temporal Patterns of Percentage Contribution of 16 risk factors in T2DM DALYs, 1990-2019. T2DM: Type-2 Diabetes Mellitus; DALYs: Disability-adjusted-life-years; Age-standardized DALYs (per 100,000 person-years). Data Source: Global Burden of Disease Study 2019

accounting for 36.7% of incident cases and 29.0% of global deaths in 2019. T2DM deaths more than doubled between 1990 and 2019 in the majority of regions led by Central Asia (all-age deaths 302.3% [260.5–348.0%]) and ASMR 163.8% [138.1–192.5%]), which again epitomizes the growing and unabated T2DM threat in Asia. In recent decades, the Asian regions have been struggling with a twin epidemic of obesity and T2DM, owing to economic development and urbanization, which has resulted in a sharp transition of lifestyles characterized by increased intake of sugar, animal products, and reduced consumption of cereals [22, 23].

North Africa and Middle East, with a rise of 373.1% [361.4–385.0%] in prevalent cases between 1990 and 2019, has become another hotspot of T2DM burden besides Asian regions. All-age incident cases also grew the fastest in a few regions of Africa and Asia led by North Africa and Middle East (354.5% [341.1–367.5%]). The ASMR reduced in a few high-income regions; for instance, high-income Asia Pacific and Western Europe registered a decline of 45.8% [42.0 to 51.6%] and 34.6% [32.0 to 37.8%], respectively, between 1990 and 2019; in contrast, it grew by 163.8% [138.1 to 192.6%] in Central Asia and 60.6% [44.1 to 77.7%] in Southern SSA. The increased burden in North Africa and Middle East has been attributed to similar risk factors as those in Asia (e.g., high BMI, lack of physical exercise, ambient air pollution) [24]. By contrast, in a few SSA countries, undernutrition in early life might lead to higher risk of hyperglycemia and T2DM later in adult life [25].

T2DM global prevalent cases among those between 15 and 19 years increased from 0.95 million in 1990 to 1.9 million in 2019. A rise in the T2DM burden in adolescents is a worrisome development documented before [26, 27]. Although the causes of this rise are poorly understood [27], limited evidence has attributed this increased burden to genetic and epigenetic factors, with obesity identified as the most significant risk factor [28]. The 15–64 age groups accounted for 85.1% of new cases, 57.8% of DALYs, and 63.4% of T2DM prevalent cases. Such a large proportion of T2DM burden in the working age group with whole-body

effects with disease progression (e.g., eyes, kidneys, heart, and liver) implies a loss of productive hours and hence, lost human capital in developing countries and regions [29–31]. DM not only results in the loss of human potential and economic productivity, but a DM patient incurs medical expenses 3-times more than a patient without DM [32]. A study [31] estimated that the direct economic cost to a T2DM patient varied from $242 (at 2011 PPP) in Mexico to $11,917 in the United States, while the indirect economic cost ranged from $45 (2011 PPP) in Pakistan to $16,914 in the Bahamas.

The widespread increase in T2DM burden globally is attributed to a complex interplay of modifiable (e.g., obesity, diet, smoking, physical activity) and non-modifiable risk factors (e.g., genetic and epigenetic) [33, 34]. Epidemiological studies have identified obesity as the single most significant risk factor of T2DM [35, 36]. We also found that 51.9% of global DALYs of T2DM are attributable to high BMI, with its risk attributable DALYs increasing in all SDI quintiles between 1990 and 2019. As per WHO estimates, globally, 38.9% of adults (males: 38.5%; females: 39.2%) had BMI ≥ 25 kg/m$^2$ in 2016 up from 21.5% (males 20.0% vs females 22.7%) in 1975 [37], and by 2030, up to 57.8% of the world's adult population is expected to be either overweight or obese [38]. Therefore, if these obesity trends persist, the T2DM burden is expected to escalate further in the future.

Apart from high BMI, ambient PM pollution (13.6%), smoking (9.9%), and secondhand smoke (8.7%) were the other notable risk factors, with household air pollution due to solid fuels being one of the leading risk factors in a few SSA regions. We also found that DALYs attributable to ambient PM pollution and smoking are increasing in majority of SDI quintiles. PM pollution increases T2DM risk [39–41], potentially through increased insulin resistance [42], alterations in gut microbiota [43] and β-cell dysfunction [44]. Active smoking [45–47] and passive smoking [48–51] are the other prominent risk factors. The concerning fact is that active and passive smoking of spouse [51] and among those at the workplace [52] is also associated with heightened T2DM risk. GBD estimates show that alcohol use might have protective effect against T2DM risk in few regions as its contribution is negative in T2DM DALYs. Previous research have also found that light-to-moderate alcohol consumption might reduce T2DM risk whereas heavy or binge drinking increases the T2DM risk [53–56] or there is inconclusive evidence on link between heavy alcohol consumption and T2DM risk [57].

### Strengths and Limitations of the study

The main strength of the study is that we employed data from one of the most reliable data sources, GBD 2019 study, which provided comprehensive and comparable estimates of T2DM burden for 204 countries from 1990 to 2019. No other study provides consistent and comparable estimates of T2DM burden for three decades while also providing estimates of the risk-attributable burden. GBD utilizes all input data from sources such as vital registration, verbal autopsy and disease registries to arrive at the mortality and incidence estimates. Moreover, the disease burden (e.g., deaths, incidence, DALYs) and risk-attributable burden of T2DM is based on rich data sources and internally consistent methods (e.g., CODEm, CodeCorrect, descriptive epidemiological meta-regression tool) to estimate the disease burden, relative risk, population exposure levels and PAFs. This study, however, also suffers from a few limitations. First, the accuracy of GBD estimates depends upon data availability from national data sources, such as vital statistics. Due to data unavailability or due to lack of quality data in several LMICs, GBD arrives at estimates using spatio-temporal modelling efforts; therefore, the model estimates in several LMICs are expected to be biased [58]. Moreover, the estimates are provided with wide UIs for several LMICs, which yield few meaningful policy implications.

## Conclusion

Worldwide prevalence of T2DM ballooned to 437.9 million in 2019, from 148.4 million in 1990. Amidst the rising global burden of T2DM, three key themes emerged from this study. First, all regions have witnessed significant increases in T2DM incident and prevalent cases in the last three decades, mainly fuelled by changing lifestyles (e.g., dietary changes, sedentary lifestyles) and deteriorating air pollution. Second, developing countries and resource-constrained economies in Asia and Africa accounted for the bulk of the global burden, which is also experienced mainly by the working-age group. These developing countries were expected to ride on the wave of demographic transition (i.e., demographic dividend) due to large proportion of the population in the working-age group. T2DM, therefore, presents significant threats in reaping demographic dividend as 85.1% of incident cases and 63.4% of T2DM prevalent cases occurred in the working-age group in 2019, which results in loss of human potential and lost opportunity of economic development. Third, the most significant risk factors of T2DM are high BMI, PM pollution and a host of dietary risk factors; therefore, urgent attention must be paid to curtail the obesity epidemic and prevent deterioration of environmental conditions as long-term exposure to air pollution may contribute to T2DM risk even in non-smokers and physically-active people [59]. Lastly, due to worsening obesity prevalence, behavioural and environmental risk factors and increasing T2DM

prevalence worldwide, the targets set by UN-high level meeting [7] to halt diabetes prevalence by 2025 to its 2010 levels are difficult to achieve until critical interventions are put in place.

## Data availability

All data is procured from GBD Results Tool http://ghdx.healthdata.org/gbd-results-tool.

**Acknowledgements** We thank Institute of Health Metrics and Evaluation for providing GBD 2019 estimates in the public domain.

**Funding** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## Compliance with ethical standards

**Conflict of interest** The authors declare no competing interests.

**Consent for publication** The authors of the study have consent and responsibility for submission to the journal.

**Ethical approval and consent to participate** The research was conducted using data available in the public domain and did not include any human participants or animals. Therefore, no ethical approvals were required.

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

## References

1. IDF Atlas Ninth Edition (2019). Available at https://www.diabetesatlas.org/upload/resources/material/20200302_133351_IDFATLAS9e-final-web.pdf. Accessed 13.3.2021
2. G.M. Reaven, Banting lecture 1988. Role of insulin resistance in human disease. Diabetes **37**, 1595–1607 (1988)
3. C. Weyer, C. Bogardus, D.M. Mott, R.E. Pratley, The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. J. Clin. Invest **104**, 787–794 (*1999*)
4. The Emerging Risk Factors Collaboration, Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet **375**, 2215–22 (2010)
5. Y. Zheng, S.H. Ley, F.B. Hu, Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat. Rev. Endocrinol. **14**(2), 88 (2018)
6. C. Bommer, E. Heesemann, V. Sagalova, J. Manne-Goehler, R. Atun, T. Bärnighausen et al. The global economic burden of diabetes in adults aged 20–79 years: a cost-of-illness study. Lancet Diabetes Endocrinol. **5**(6), 423–30 (2017)
7. WHO. Global action plan for the prevention and control of noncommunicable diseases 2013–20 (2013). Available at https://www.who.int/publications/i/item/9789241506236. Accessed April 16, 2022
8. T. Vos, S.S. Lim, C. Abbafati et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. Lancet **396**, 1204–22 (2020)
9. H. Wang, K.M. Abbas, M. Abbasifard et al. Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950–2019: a comprehensive demographic analysis for the global burden of disease study 2019. Lancet **396**, 1160–203 (2020)
10. C.J. Murray, Quantifying the burden of disease: the technical basis for disability-adjusted life years. Bull. World health Org. **72**(3), 429–445 (1994)
11. B. Zhou, Y. Lu, K. Hajifathalian et al. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4· 4 million participants. Lancet **387**(10027), 1513–30 (2016)
12. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle: Institute for Health Metrics and Evaluation (IHME). Source: Institute for Health Metrics Evaluation. Used with permission. All rights reserved. Available from http://ghdx.healthdata.org/gbd-results-tool. Accessed: Jan–Feb 2021
13. K.J. Foreman, R. Lozano, A.D. Lopez, C.J. Murray, Modeling causes of death: an integrated approach using CODEm. Popul Health Metr. **10**(1), 1 (2012)
14. C.J. Murray, A.Y. Aravkin, P. Zheng et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet **396**(10258), 1223–49 (2020)
15. A. Nordström, J. Hadrévi, T. Olsson, P.W. Franks, P. Nordström, Higher prevalence of type 2 diabetes in men than in women is associated with differences in visceral fat mass. J. Clin. Endocrinol. Metab. **101**(10 Oct), 3740–6 (2016)
16. J. Logue, J.J. Walker, H.M. Colhoun, G.P. Leese, R.S. Lindsay, J.A. McKnight, A.D. Morris, D.W. Pearson, J.R. Petrie, S. Philip, S.H. Wild, Do men develop type 2 diabetes at lower body mass indices than women? Diabetologia **54**(12), 3003–6 (2011)
17. F.E. von Eyben, E. Mouritsen, J. Holm et al. Intra-abdominal obesity and metabolic risk factors: a study of young adults. Int J. Obes. Relat. Metab. Disord. **27**, 941–949 (2003)
18. A. Chandra, I.J. Neeland, J.D. Berry et al. The relationship of body mass and fat distribution with incident hypertension: observations from the Dallas Heart Study. J. Am. Coll. Cardiol. **64**, 997–1002 (2014)
19. P. Wiklund, F. Toss, L. Weinehall et al. Abdominal and gynoid fat mass are associated with cardiovascular risk factors in men and women. J. Clin. Endocrinol. Metab. **93**, 4360–4366 (2008)
20. B. Tramunt, S. Smati, N. Grandgeorge, F. Lenfant, J.F. Arnal, A. Montagner, P. Gourdy, Sex differences in metabolic regulation and diabetes susceptibility. Diabetologia **63**(3 Mar), 453–61 (2020)
21. F. Mauvais-Jarvis, Sex differences in metabolic homeostasis, diabetes, and obesity. Biol. Sex. Differ. **6**, 14 (2015)
22. J.C. Chen et al. Diabetes in Asia: epidemiology, risk factors, and pathophysiology. JAMA **301**, 2129–2140 (2009)
23. S.H. Ley, O. Hamdy, V. Mohan, F.B. Hu, Prevention and management of type 2 diabetes: dietary components and nutritional strategies. Lancet **383**(9933), 1999–2007 (2014)
24. A. Majeed, A.A. El-Sayed, T. Khoja, R. Alshamsan, C. Millett, S. Rawaf, Diabetes in the Middle-East and North Africa: an update. Diabetes Res Clin. Pr. **103**(2), 218–22 (2014)
25. Y. Li, Y. He, L. Qi, V.W. Jaddoe, E.J. Feskens, X. Yang et al. Exposure to the Chinese famine in early life and the risk of hyperglycemia and type 2 diabetes in adulthood. Diabetes **59**(10), 2400–6 (2010)
26. L. Chen, D.J. Magliano, P.Z. Zimmet, The worldwide epidemiology of type 2 diabetes mellitus—present and future perspectives. Nat. Rev. Endocrinol. **8**(4), 228–36 (2012)
27. D. Dabelea, E.J. Mayer-Davis, S. Saydah, G. Imperatore, B. Linder, J. Divers et al. Prevalence of type 1 and type 2 diabetes

among children and adolescents from 2001 to 2009. JAMA **311**(17), 1778–86 (2014)
28. T. Reinehr, Type 2 diabetes mellitus in children and adolescents. World J. diabetes **4**(6), 270 (2013)
29. J. Holmes, E. Gear, J. Bottomley, S. Gillam, M. Murphy, R. Williams, Do people with type 2 diabetes and their carers lose income? (T2ARDIS-4). Health Policy **64**(3), 291–6 (2003)
30. X. Liu, C. Zhu, Will knowing diabetes affect labor income? Evidence from a natural experiment. Econ. Lett. **124**, 74–8 (2014)
31. T. Seuring, O. Archangelidi, M. Suhrcke, The economic costs of type 2 diabetes: a global systematic review. Pharmacoeconomics **33**, 811–31 (2015)
32. R.J. Rubin, W.M. Altman, D.N. Mendelson, Health care expenditures for people with diabetes mellitus, 1992. J. Clin. Endocrinol. Metab. **78**(4), 809A-F (1994)
33. F.B. Hu et al. Diet, Lifestyle, and the risk of type 2 diebetes mellitus in women. N. Eng. J. Med **345**, 790–97 (2001)
34. E.S. Schellenberg, D.M. Dryden, B. Vandermeer, C. Ha, C. Korownyk, Lifestyle interventions for patients with and at risk for type 2 diabetes: a systematic review and meta-analysis. Ann. Intern Med **159**(8), 543–51 (2013)
35. C. Merlotti, A. Morabito, A.E. Pontiroli, Prevention of type 2 diabetes; a systematic review and meta-analysis of different intervention strategies. Diabetes Obes. Metab. **16**, 719–27 (2014)
36. Diabetes Prevention Program Research Group, Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the Diabetes Prevention Program Outcomes Study. Lancet Diabetes Endocrinol. **3**, 866–75 (2015)
37. Global Health Observatory data repository. Prevalence of overweight among adults, BMI ≥ 25, age-standardized. Available at http://apps.who.int/gho/data/view.main.CTRY2430A?lang=en. Accessed 26 Feb 2020
38. T. Kelly, W. Yang, C.S. Chen, K. Reynolds, J. He, Global burden of obesity in 2005 and projections to 2030. Int J. Obes. **32**(9), 1431–7 (2008)
39. E.V. Balti, J.B. Echouffo-Tcheugui, Y.Y. Yako, A.P. Kengne, Air pollution and risk of type 2 diabetes mellitus: a systematic review and meta-analysis. Diabetes Res Clin. Pr. **106**(2), 161–72 (2014)
40. G. Weinmayr, F. Hennig, K. Fuks, M. Nonnemacher, H. Jakobs, S. Möhlenkamp, R. Erbel, K.H. Jöckel, B. Hoffmann, S. Moebus, Long-term exposure to fine particulate matter and incidence of type 2 diabetes mellitus in a cohort study: effects of total and traffic-specific air pollution. Environ. Health **14**(1), 1–8 (2015)
41. C. Liu, C. Yang, Y. Zhao, Z. Ma, J. Bi, Y. Liu, X. Meng, Y. Wang, J. Cai, H. Kan, R. Chen, Associations between long-term exposure to ambient particulate air pollution and type 2 diabetes prevalence, blood glucose and glycosylated hemoglobin levels in China. Environ. Int. **92**, 416–21 (2016)
42. S. Zhang, S. Mwiberi, R. Pickford, S. Breitner, C. Huth, W. Koenig, W. Rathmann, C. Herder, M. Roden, J. Cyrys, A. Peters, Longitudinal associations between ambient air pollution and insulin sensitivity: results from the KORA cohort study. Lancet Planet. Health **5**(1), e39–49 (2021)
43. T. Liu, X. Chen, Y. Xu, W. Wu, W. Tang, Z. Chen, G. Ji, J. Peng, Q. Jiang, J. Xiao, X. Li, Gut microbiota partially mediates the effects of fine particulate matter on type 2 diabetes: evidence from a population-based epidemiological study. Environ. Int. **130**, 104882 (2019)
44. T.L. Alderete, R. Habre, C.M. Toledo-Corral, K. Berhane, Z. Chen, F.W. Lurmann, M.J. Weigensberg, M.I. Goran, F.D. Gilliland, Longitudinal associations between ambient air pollution with insulin sensitivity, β-cell function, and adiposity in Los Angeles Latino children. Diabetes **66**(7), 1789–96 (2017)
45. C. Willi, P. Bodenmann, W.A. Ghali, P.D. Faris, J. Cornuz, Active smoking and the risk of type 2 diabetes: a systematic review and meta-analysis. Jama **298**(22), 2654–64 (2007)
46. L. Zhang, G.C. Curhan, F.B. Hu, E.B. Rimm, J.P. Forman, Association between passive and active smoking and incident type 2 diabetes in women. Diabetes care **34**(4), 892–7 (2011)
47. A. Pan, Y. Wang, M. Talaei, F.B. Hu, T. Wu, Relation of active, passive, and quitting smoking with incident type 2 diabetes: a systematic review and meta-analysis. Lancet Diabetes Endocrinol. **3**(12), 958–67 (2015)
48. T.K. Houston, S.D. Person, M.J. Pletcher, K. Liu, C. Iribarren, C.I. Kiefe, Active and passive smoking and development of glucose intolerance among young adults in a prospective cohort: CARDIA study. Bmj **332**(7549), 1064–9 (2006)
49. B. Zhu, X. Wu, X. Wang, Q. Zheng, G. Sun, The association between passive smoking and type 2 diabetes: a meta-analysis. Asia Pac. J. Public Health **26**(3), 226–37 (2014)
50. X. Wei, E. Meng, S. Yu, A meta-analysis of passive smoking and risk of developing Type 2 Diabetes Mellitus. Diabetes Res Clin. Pr. **107**(1), 9–14 (2015)
51. S. Oba, A. Goto, T. Mizoue, M. Inoue, N. Sawada, M. Noda, S. Tsugane, Passive smoking and type 2 diabetes among never-smoking women: The Japan Public Health Center-based Prospective Study. J. Diabetes Investig. **11**(5), 1352–8 (2020)
52. Y. Hayashino, S. Fukuhara, T. Okamura, H. Yamato, H. Tanaka, T. Tanaka, T. Kadowaki, H. Ueshima, A prospective study of passive smoking and risk of diabetes in a cohort of workers: the High-Risk and Population Strategy for Occupational Health Promotion (HIPOP-OHP) study. Diabetes care **31**(4), 732–4 (2008)
53. S.G. Wannamethee, C.A. Camargo, J.E. Manson, W.C. Willett, E.B. Rimm, Alcohol drinking patterns and risk of type 2 diabetes mellitus among younger women. Arch. Intern. Med. **163**(11 Jun), 1329–36 (2003)
54. S. Carlsson, N. Hammar, V. Grill, Alcohol consumption and type 2 diabetes. Diabetologia **48**(6 Jun), 1051–4 (2005)
55. D.O. Baliunas, B.J. Taylor, H. Irving, M. Roerecke, J. Patra, S. Mohapatra, J. Rehm, Alcohol as a risk factor for type 2 diabetes: a systematic review and meta-analysis. Diabetes care **32**(11), 2123–32 (2009)
56. A. Pietraszek, S. Gregersen, K. Hermansen, Alcohol and type 2 diabetes. A review. Nutr., Metab. Cardiovascular Dis. **20**(5 Jun), 366–75 (2010)
57. X.H. Li, F.F. Yu, Y.H. Zhou, J. He, Association between alcohol consumption and the risk of incident type 2 diabetes: a systematic review and dose-response meta-analysis. Am. J. Clin. Nutr. **103**(3 Mar), 818–29 (2016)
58. P. Zimmet, K.G. Alberti, D.J. Magliano, P.H. Bennett, Diabetes mellitus statistics on prevalence and mortality: facts and fallacies. Nat. Rev. Endocrinol. **12**(10), 616 (2016)
59. Z.J. Andersen, O. Raaschou-Nielsen, M. Ketzel, S.S. Jensen, M. Hvidberg, S. Loft, A. Tjønneland, K. Overvad, M. Sørensen, Diabetes incidence and long-term exposure to air pollution: a cohort study. Diabetes care **35**(1), 92–8 (2012)

# Understanding Percentage Impedance (%Z) of Transformer with Examples

**DR. RAJESH KUMAR ARORA**[1]

**ABSTRACT**

*The transformer, being a key element in the transmission and distribution of electrical energy, understanding its specified parameters is of utmost importance for power engineers. The most important design parameter of the transformer is the percentage impedance % Z (Leakage Impedance).*

*This paper provides information about the significance of percentage impedance of transformer & discusses the methods of measurement of percentage impedance of transformer.*

***Keywords*** **:** *Power Transformer, Percentage Impedance (%Z), Leakage Flux, Resistance and Reactance of Transformer*

## 1. INTRODUCTION

The transformer is an electric machine which works on the principle of Faraday's Law of Electromagnetic Induction which states that the magnitude of voltage is directly proportional to the rate of change of flux, and is used to transfer electrical power over the same frequency, but the voltage level can be altered that means voltage can be increased or decreased according to the requirement. Actual picture of power transformer is shown in Figure 1.

The transformer has two main parts; core and windings. The transformer core is designed to provide a path for the magnetic field to flow around, which is necessary for the induction of the voltage between the two windings. The core acts as a support to the windings in the transformer. It also provides a low reluctance path to the flow of magnetic flux. Transformer windings have mainly two types; Primary winding and Secondary winding. The Primary winding has $N_1$ number of turns, while secondary winding has $N_2$ number of turns. Primary winding is used to take power from the source and secondary winding is used to transfer changed voltage to the load (Refer Figure 2).

In transformers, the transformation of voltage from primary to secondary is not perfect, i.e. there is some loss of energy. This energy is lost due to **leakage flux**

**Fig. 1** : Power Transformer

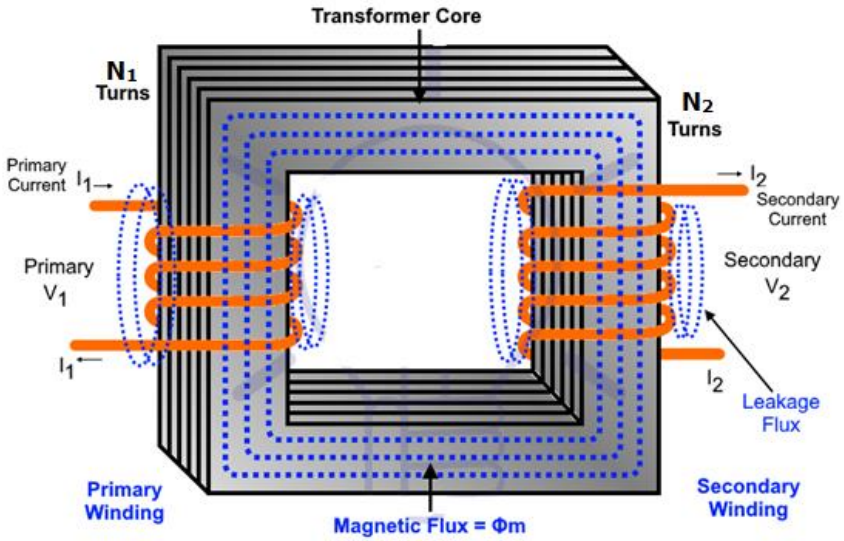



**Fig. 2** : Basic Transformer

1. Delhi Transco Ltd (DTL), Delhi, India.

www.IndianJournals.com Members copy, Not for Commercial Sale Downloaded from IP - 14.139.82.29 on Monday, August 01, 2022 at 12:09:02 PM

(when all the flux from the primary winding do not link to the secondary winding) and the impedance offered by the copper wires (copper loss) in the windings of the transformer.

## 1.1 Resistance of Transformer

Generally, both primary and secondary windings of electrical power transformer are made of copper. Copper is a very good conductor of current but not a super conductor. So both windings will have some resistance. This internal resistance of both primary and secondary windings is collectively known as **resistance of transformer** (Refer Rigure 3).

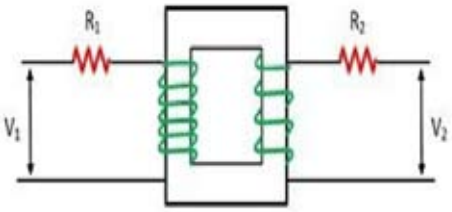



**Fig. 3 :** Winding Resistance of Transformer

## 1.2 Leakage Flux & Reactance of Transformer

All the flux in transformer will not be able to link with both the primary and secondary windings. A small portion of flux will link either winding but not both. This portion of flux is called leakage flux. Due to this **leakage flux in transformer** (Refer Figure 4), there will be a self-reactance in the concerned winding. This self-reactance of transformer is alternatively known as **leakage reactance of transformer**.

## 1.3 Impedance (Z) of Transformer

As we said, both primary and secondary windings will have resistance and leakage reactance. These resistance and reactance will be in combination, is nothing but **impedance of transformer**. Due to this **impedance of transformer**, there will be voltage drops in both primary and secondary transformer windings. (Refer Figure 5)

If $R_1$ and $R_2$ and $X_1$ and $X_2$ are primary and secondary resistance and leakage reactance of transformer respectively, then $Z_1$ and $Z_2$ impedance of primary and secondary windings are respectively,

$$Z_1 = R_1 + jX_1$$

$$Z_2 = R_2 + jX_2$$

Voltage drops in the windings occur due to impedance of transformer. Impedance is combination of resistance and leakage reactance of transformer. If we apply voltage $V_1$ across primary of transformer, there will be a component $I_1X_1$ to balance primary self induced emf due to primary leakage reactance **(Here, $X_1$ is primary leakage reactance).** Now if we also consider voltage drop due to primary resistance of transformer, then voltage equation of a transformer can easily be written as,

$$V_1 = E_1 + I_1R_1 + jI_1X_1 \quad \text{... Equation (1)}$$

Similarly for secondary leakage reactance, the voltage equation of secondary side is,

$$V_2 = E_2 - I_2R_2 - jI_2X_2 \quad \text{.....Equation (2)}$$

The transformer impedance is expressed as some percentage of the rated voltage at full load current. In other words, transformer impedance is the percentage of rated voltage necessary to circulate full load current when the secondary of the transformer is short circuited.

The voltage drop due to impedance of winding is given as:

$$V_D = I_{FLA}Z \quad \text{..... Equation (3)}$$

Also the percentage impedance of the transformer is given as:

$$\%Z = V_D/E*100 \quad \text{..... Equation (4)}$$

Where **$I_{FLA}$ is the full load current**

Z is the impedance of the windings and is given by

$$Z = (R^2+X^2)^{1/2} \quad \text{.....Equation (5)}$$

R is the resistance of transformer

X is the leakage reactance

E is the open terminal voltage.

The name plate or specification plate which specifies the value of percentage impedance of power transformer is shown in Figure 6.

## 2.0 IMPORTANCE OF IMPEDANCE IN TRANSFORMER

Impedance is the current limiting characteristic of a transformer and is expressed in percentage. It is used for determining the interrupting capacity of a circuit breaker or fuse employed to protect the primary winding of a transformer.

The impedance (or resistance to current flow) is important and used to calculate the maximum short circuit current which is needed for sizing, circuit breakers and fuses. This percentage represents the amount of normal rated primary voltage which must be applied to the transformer to produce full rated load current when the secondary winding is short circuited. The maximum short

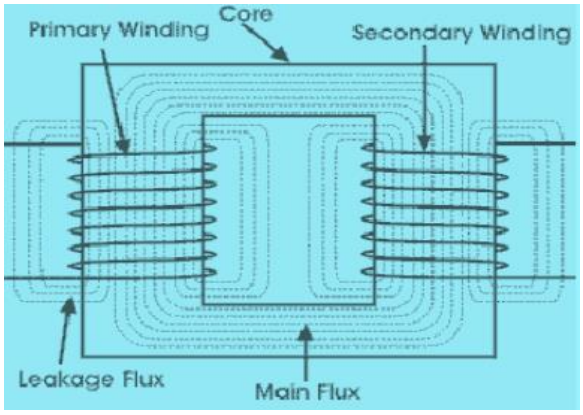



**Fig. 4** : Leakage Flux in Transformer

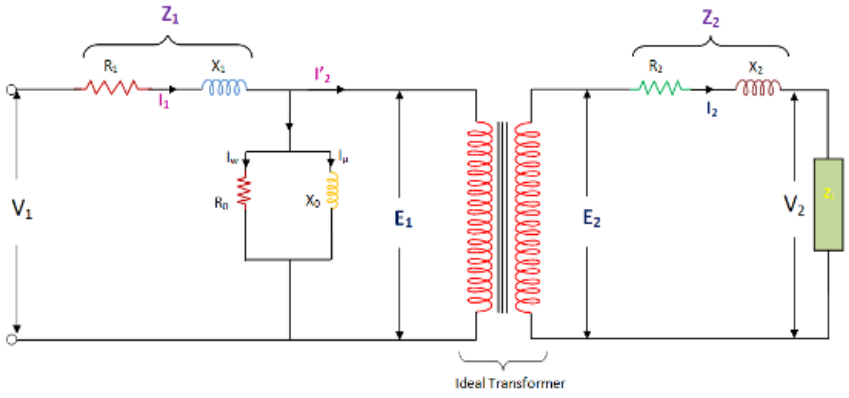



**Fig. 5** : Impedance (Z) of Transformer

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From - 14.139.251.98 on dated 1-Aug

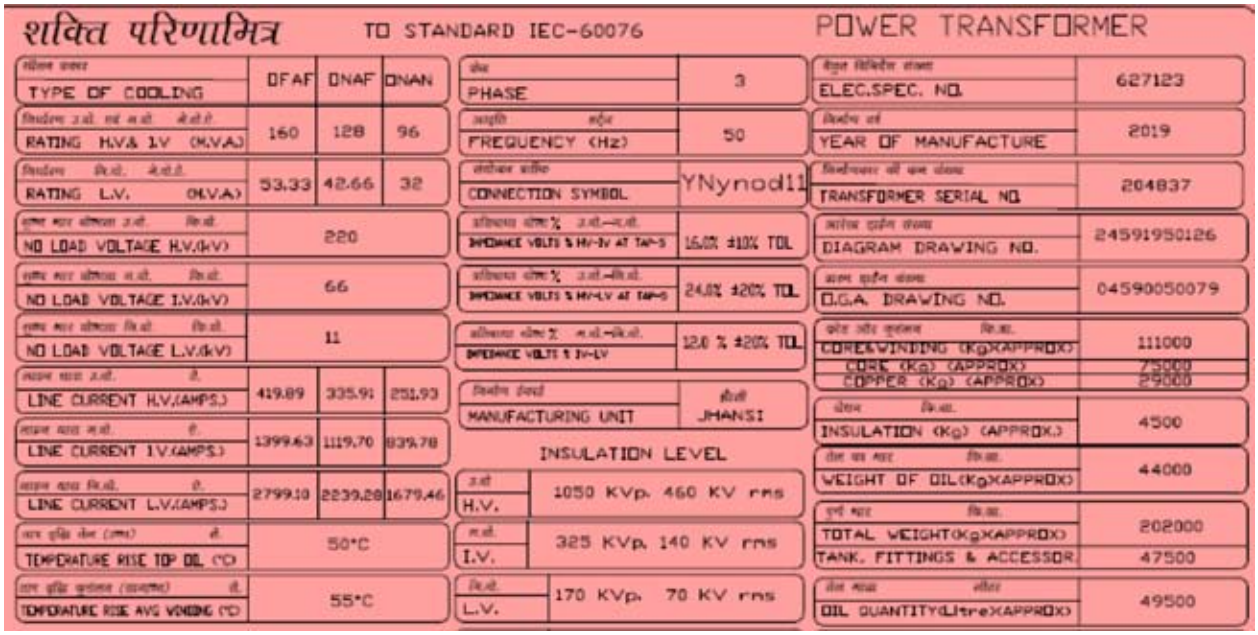



**Fig. 6** : Specification Plate of Transformer

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 157.251.98 on Dated 1-Aug-2022

circuit current that can be obtained from the output of the transformer is limited by the impedance of the transformer.

## 2.1 High and Low Impedance Transformers

High impedance transformers have a lower fault or short circuit current and have no need for high SCC rated breakers. However, they will have a higher voltage drop or regulation.

Low Impedance transformers on the other hand have a lower voltage drop or regulation but have a higher fault or short circuit current and will need higher SCC rated breakers.

In simple words, the percentage impedance of a transformer has a major effect on **system fault levels**. It determines the maximum value of current that will flow under fault conditions.

It is easy to calculate the maximum current that a transformer can deliver under symmetrical fault conditions.

By way of **example**, consider a 2 MVA transformer with an impedance of 5%. The maximum fault level available on the secondary side is:

**2 MVA x 100/5 = 40 MVA**

and from this data, the equivalent primary and secondary fault currents can be calculated.

**A less percentage impedance has both positive and negative effects.**

1. If Z% of the transformer is less, the short circuit current will be more which will produce **more stress in insulation**. This is a **negative factor**.
2. On the other hand, it will reduce the voltage drop in the transformer winding. This will facilitate **better voltage regulation**. This is a **positive factor**.

Therefore the percentage impedance of the transformer has to be precisely selected to maintain a proper balance between fault level and voltage regulation.

## 2.2 Short Circuit Current & Percentage Impedance

As we have discussed in the above section, when we apply rated voltage at the primary winding of a transformer whose secondary winding is short-circuited, **short circuit current** will flow on transformer windings.

The value of the short circuit current is,

$$I_{sc} = I_{rated} \times 100/Z\%$$

The value of percentage impedance is the same for both windings as it is the percentage of rated voltage. However, the value of the rated current will be different for primary and secondary windings. Accordingly, value of short circuit current will also be different for primary and secondary windings.

## 2.3 Percentage Impedance & Parallel Operation

For parallel connection of transformers, primary windings of the transformers are connected to source bus-bars and secondary windings are connected to the load bus-bars. Various conditions that must be fulfilled for the successful parallel operation of transformers are

- Same voltage ratio and turns ratio (both primary and secondary voltage rating is same)
- Same percentage impedance and X/R ratio.
- Identical position of tap changer
- Same KVA ratings
- Same phase angle shift (vector group are same)
- Same frequency rating
- Same polarity
- Same phase sequence.

When transformers which have the same kVA rating but with different percent impedances are connected in parallel, the current divides in inverse proportions to the impedances and larger current flows through the smaller impedance. Thus, the lower percent impedance transformer

can be overloaded when subjected to heavy loading while the other higher percent impedance transformer will be lightly loaded.

The paralleling of transformers is simple unless KVA, percentage impedance, and turn ratios are different. We will look at different scenarios by changing the above mentioned convenient conditions. We are using the following parameters for the examples given below:

1. KVA'= Rating of Transformer 1
2. KVA" = Rating of Transformer 2
3. $KVA_1$ = Load on Transformer 1
4. $KVA_2$ = Load on Transformer 2
5. %Z1 = Impedance of Transformer 1
6. %Z2 = Impedance of Transformer 2
7. K = ratio of KVA of T1 & T2 i.e. $KVA_1$/ $KVA_2$

**(A) Equal Impedances-Equal Voltage Ratios -Same KVA**

If the percentage impedances of each transformer are same, there will be equal division of current and load sharing on each transformer.

*Example:*

Two transformers of **250 KVA, 5%** impedance connected in parallel catering **400** KVA load.

The KVA can be calculated as:

$$\frac{KVA'/\%Z_1}{\frac{KVA'}{\%Z_1}+\frac{KVA''}{\%Z2}}\ X\ KVA_L$$

**$KVA_1$ = $KVA_2$ = 200 KVA.**

**(B) Equal Impedances-Equal Voltage Ratios-Different KVA**

In this type of situation each transformer in parallel carry its rated load. The circulating currents will not be present since the turn ratios are same.

*Example:*

Two transformers of **250 KVA, 400 KVA, 5%** impedance connected in parallel catering 400 KVA load.

$KVA_1$ = 153.84 KVA

$KVA_2$ = 246.15 KVA

**(C) Unequal Impedances-Equal Voltage Ratios-Same KVA**

The current is inversely proportional to the impedance, thus a large amount of current flows through smaller impedance. In this situation, the lower impedance transformer is heavily loaded while the higher impedance transformer is lightly loaded.

*Example:*

Two transformers of **250 KVA, 4% and 5%** impedances respectively connected in parallel catering 400 KVA load.

$KVA_1$=222.22 KVA

$KVA_2$=177.78 KVA

**D) Unequal Impedances-Equal Voltage Ratios-Different KVA**

This situation is not common. But, can be taken into account with care as the load current carried by the combination of transformers would be less than the rated KVA.

*Example:*

Two transformers of **250 KVA with 4% and 400 KVA with 5%** impedances connected in parallel catering 400 KVA load.

$KVA_1$ = 175.43 KVA

$KVA_2$ = 224.56 KVA

## 3.0 MEASUREMENT OF %Z OF TRANSFORMER

The percentage impedance of a transformer is the volt drop on full load due to the winding resistance and leakage reactance expressed as a percentage of the rated voltage.

It is also the percentage of the normal terminal voltage required to circulate full-load current under short circuit conditions.

In other words, the percentage impedance of a transformer is the percentage of the rated voltage applied at one side (primary winding) to circulate rated current on transformer keeping its other side (secondary winding) under short circuit conditions. (Refer Figure 7)

If we apply rated voltage at the primary winding of a transformer keeping its **secondary winding short-circuited**, then the amount of current at both windings will be extremely high as compared to the rated current.

This current is called **short circuit current** and its magnitude is very high due to zero impedance offered by the load (secondary winding is short-circuited).

Now, if we reduce the applied voltage on the transformer primary i.e. we apply a percentage of rated voltage in transformer primary, current on both windings will also reduce.

At a particular percentage of rated voltage, the rated current will flow on transformer windings. This percentage of rated voltage at one side of the transformer which circulates rated current on transformer windings keeping its other side winding short-circuited is called **percentage impedance of the transformer**.

### 3.1 Example Calculation

For example, if a 2,000/240-volt transformer has a measured impedance voltage of 100 volts on the high voltage windings, its impedance (Z), expressed as a percent, is:

**Z% = (Impedance Voltage / Rated Voltage) x 100**

**Percentage Z = (100/2000)*100 = 5 %**

This means there would be a 100-volt drop in the high-voltage winding at full load due to losses in the windings and core. Only 1 or 2% of the losses are due to the core; about 98% is due to the winding impedance.

### 3.2 Example Calculation Using Low Voltage AC

Percentage impedance is usually calculated using low voltage AC for all the taps in the transformer by the formula given below:

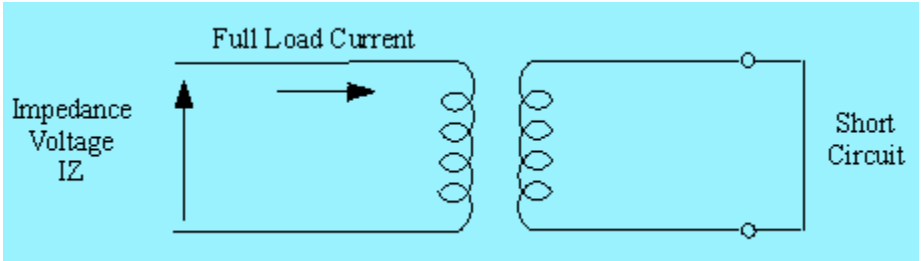



**Fig. 7** : Percentage Impedance Measurement in Transformer

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded by IP 14.1.252.28 on Wed Aug 10 2022

% Impedance = $(V_T/V_R) \times (I_R/I_A) \times 100\%$
$V_T$ is test voltage & $V_R$ is rated Voltage
$I_A$ is Actual Current & $I_R$ is rated Current
Example of % Z calculation for 100 MVA, 220 /33 kV power transformer is given in **Table 1.**

### 3.3 Value & Permitted Tolerance in Z% of Transformer

The value of percentage impedance for different rating as per IS 2026 is given below in Table 2.

The percentage impedance of a transformer is specified at the time of ordering. But it must be noted that **IEC 60076 / IS 2026** permits some **tolerance in percentage impedance** at the manufacturer's end (Refer Table 3)

**Table 1** : % Z Calculation Using Low Voltage & Short Circuit Method ( Lv Side Is Shorted)

| Tap No | Rated Current Amps | Rated Voltage kV | Measured (Actual) Current Amps | Test (Applied) Voltage kV | % Z (IMPEDANCE) |
|---|---|---|---|---|---|
| 1 | 399.9 | 231 | 5.67 | 0.43 | 13.13 |
| 2 | 404.71 | 228.25 | 5.58 | 0.431 | 13.70 |
| 3 | 409.65 | 225.5 | 5.55 | 0.427 | 13.98 |
| 4 | 414.71 | 222.75 | 5.45 | 0.427 | 14.59 |
| 5 | 419.89 | 220 | 5.42 | 0.428 | 15.07 |
| 6 | 425.21 | 217.25 | 5.56 | 0.43 | 15.14 |
| 7 | 430.66 | 214.5 | 5.52 | 0.43 | 15.64 |
| 8 | 436.25 | 211.75 | 5.51 | 0.429 | 16.04 |
| 9 | 441.99 | 209 | 5.57 | 0.43 | 16.33 |
| 10 | 447.88 | 206.25 | 5.48 | 0.43 | 17.04 |
| 11 | 453.94 | 203.5 | 5.44 | 0.43 | 17.63 |
| 12 | 460.16 | 200.75 | 5.45 | 0.43 | 18.09 |
| 13 | 466.55 | 198 | 5.49 | 0.429 | 18.41 |
| 14 | 473.12 | 195.25 | 5.42 | 0.429 | 19.18 |
| 15 | 479.88 | 192.5 | 5.41 | 0.427 | 19.68 |
| 16 | 486.83 | 189.75 | 5.43 | 0.428 | 20.22 |
| 17 | 493.99 | 187 | 5.32 | 0.428 | 21.25 |

**Table 2 :** % Z of Power Transformer as per IS 2026

| Short-circuit impedance at rated current | |
|---|---|
| **Rated power** kVA | **Minimum short-circuit impedance** % |
| Up to 630 | 4,0 |
| 631 to 1 250 | 5,0 |
| 1 251 to 2 500 | 6,0 |
| 2 501 to 6 300 | 7,0 |
| 6 301 to 25 000 | 8,0 |
| 25 001 to 40 000 | 10,0 |
| 40 001 to 63 000 | 11,0 |
| 63 001 to 100 000 | 12,5 |
| Above 100 000 | >12,5 |
| NOTE 1 Values for rated power greater than 100 000 kVA are generally subjected to agreement between manufacturer and purchaser. | |
| NOTE 2 In case of single-phase units connected to form a three-phase bank, the value of rated power applies to three-phase bank rating. | |

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 14.139.251.98 on dated 1-Aug-2022

**Table 3** : Tolerances as per IS 2026

| Sl No. | Item | Tolerance |
|---|---|---|
| (1) | (2) | (3) |
| i) | a) Total losses | +10 percent of the total losses |
| | b) Component losses (*see* Note 1) | +15 percent of each component loss, provided that the tolerance for total losses is not exceeded |
| ii) | a) Voltage ratio at no load on principal tapping for a specified first pair of windings | The lower of the following values:<br>a) ±0.5 percent of declared ratio<br>b) ±1/10 of the actual percentage impedance on the principal tapping |
| | b) Voltage ratio on other tappings, same pair | To be agreed, but not less than the lesser of the values given in (a) and (b) above |
| | c) Voltage ratio for further pairs | To be agreed, but not less than the lesser of the values given in (a) and (b) |
| iii) | Short-circuit impedance for:<br>a) separate-winding transformer with two windings, or<br>b) a specified first pair of separate windings in a multi-winding transformer | |
| | 1) principal tapping | When the impedance value is ≥10 percent<br>±7.5 percent of the declared value<br>When the impedance value is <10 percent<br>±10 percent of the declared value |
| | 2) any other tapping of the pair | When the impedance value is ≥10 percent<br>±10 percent of the declared value<br>When the impedance value is <10 percent<br>±15 percent of the declared value |
| iv) | Short-circuit impedance for:<br>a) an auto-connected pair of winding, or<br>b) a specified second pair of separate windings in a multi-winding transformer | |
| | 1) principal tapping | ±10 percent of the declared value |
| | 2) any other tapping of the pair | ±15 percent of the declared value for that tapping |
| | 3) further pairs of windings | To be agreed, but ≥15 percent |
| v) | No-load current | +30 percent of the declared value |

NOTES

1 The loss tolerances of multi-winding transformers apply to every pair of windings unless the guarantee states that they apply to a given load condition.

2 For certain auto-transformers and booster transformers the smallness of their impedance justifies more liberal tolerance. Transformers having large tapping ranges, particularly if the range is asymmetrical, may also require special consideration. On the other hand, for example, when a transformer is to be combined with previously existing units, it may be justified to specify and agree on narrower impedance tolerances. Matters of special tolerances shall be brought to attention at the tender stage, and revised tolerances agreed upon between manufacturer and purchaser.

3 'Declared value' should be understood as meaning the value declared by the manufacturer.

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 14.139.251.98 on dated 1-Aug-2022

**Example**: If we order a transformer with 8% percentage impedance with tolerance of +-10%, its actual Z% after manufacturing may be any value between 7.2% (-10% of 8) to 8.8% (+10% of 8), unless it is specifically agreed with the manufacturer at the time of ordering.

Tolerance in percentage impedance must be considered for power system calculations and accordingly system fault level & voltage regulation must be finalized.

## 4.0 PER UNIT (OR %) VS ABSOLUTE VALUE OF IMPEDANCE

Transformers used in commercial and industrial applications typically show impedances marked in percentages.

While the technical definition of this percentage or "per unit" impedance is that this value is a percentage of the base impedance of the transformer, there is a more descriptive way of thinking of it.

The per-unit impedance describes that percentage of the rated voltage required to produce full load current while the transformer output is shorted. The lower the impedance, the lower the voltage required to produce full load current. Lower impedance transformers allow higher fault currents to flow at a particular voltage.

Measuring impedance in units of percentage greatly simplifies the calculation of currents and voltages in a power system. While absolute impedance (measured in ohms) could certainly also be used, doing so greatly complicates the calculations.

**4.1 Example 1 :** Consider a transformer with a 5% impedance. A voltage is applied to the primary with the secondary winding shorted (faulted).

At 5% input voltage 100% Full Load Current (FLA) produced

How much current flows at higher voltages? Ignoring non-linear effects such as core saturation, if 5% voltage produces 100% FLA, then:

- At 10% input voltage 200% FLA is produced
- At 50% input voltage 1000% FLA is produced
- At 100% input voltage 2000% FLA is produced

Note that at nominal full voltage on the primary winding, the fault currents, both primary and secondary, increase to 20 times FLA.

**Hence one advantage of publishing per unit impedances** becomes evident. Virtually by inspection, a first approximation of this transformer's available fault current is to multiply the **FLA x 1/%Z.**

$$I_{sc} = I_{FLA}/\%Z$$

It is pertinent to mention here that while percent impedance is independent of the transformer voltage (or tap), absolute impedance is not.

In other words, separate absolute impedances would need to be calculated not only for every voltage winding, but also for every tap setting of every winding.

Consider an ideal transformer (lossless, zero impedance) connected to a source V1 and providing current to a load $Z_L$. (Refer Figure 8)

Figure 8 shows input and output voltages along with the impedance of the load. This provides enough information to calculate the secondary current.

$I_2$ = V/Z (load)=200 A

We know that $V_1 \cdot I_1 = V_2 \cdot I_2$

$I_1 = 40$ A

Since impedance is defined as the ratio of voltage divided by current, using the known voltage and the derived current, the impedance looking through the transformer to the load can be found. (Refer Figure 9)

**Z1 = V1/I1 = 1000/40 = 25 ohm**

When performing a power systems analysis 25 ohms is the actual circuit impedance that limits the 1000 V to only 40 amperes.

However, It is clear that this value could not be published on the transformer since its value is dependent on the load and one other factor i.e. turns ratio squared (1000/200)* (1000/200)= (5*5)=25.0.

Note that this multiplier, 25 is equal to the turns ratio squared:

It is evident that the secondary load impedance observed at the primary terminals of the transformer changes with the ratio of the primary to secondary voltage. That means, changing taps settings will change the apparent impedance looking through the transformer.

Because of this complexity, it is clear why power systems engineers began to favour per unit calculations instead of using absolute measurements.

**4.2 Example 2** : Consider a 1000 kVA, 5 % Z, 1000/200 Volt transformer.

The constant of proportionality used to convert per unit impedance to absolute impedance is the base impedance (= rated KVA , 1000).

$$Z_{ohms} = Z_{pu} * Z_{base}$$

Beginning with the 200V secondary, the base impedance is calculated as:

$Z_{base} = V^2_{base}/S_{base} = 200^2/1000000 = \mathbf{0.40\Omega}$

With base impedance calculated, solve for actual impedance.

$\mathbf{Zohms = Z_{pu} * Z_{base} = (0.05*0.4) = 0.002\ \Omega}$

Looking into the 200V transformer from the secondary side, that transformer would appear to place 0.002 Ω impedance in series with the source.

To show how this value is not constant, we recall the statement that absolute impedance changes with voltage. Solving **base impedance** again but replacing 200V with 1000V:

$Z_{base} = V^2_{base}/S_{base} = 1000^2/1000000 = \mathbf{1\Omega}$

With the new primary base impedance, the absolute impedance is calculated:

$Z_{ohms} = Z_{pu} * Z_{base} = 0.05*1 = \mathbf{0.05\ \Omega}$

This means that measured on the 1000V circuit, the transformer appears to place **0.05 Ω impedance in series with the load, not 0.002 Ω.**

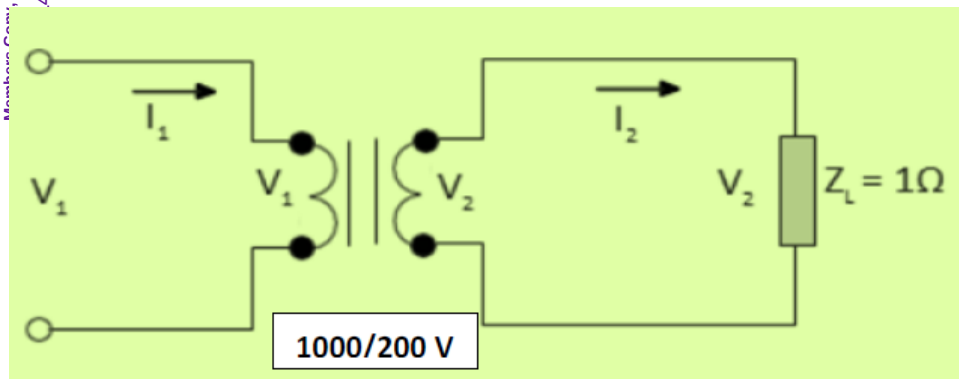



**Fig. 8** : 1000/200 V Single Phase Transformer

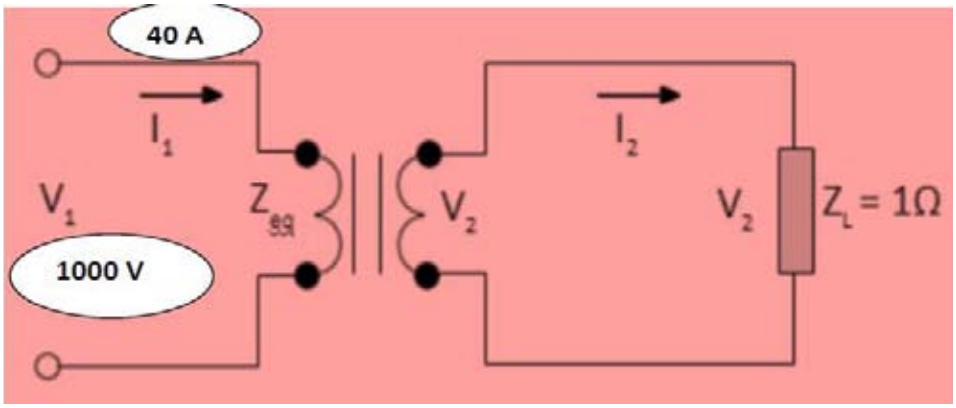



**Fig. 9** : Current in HV Side of 1000/200 V Transformer

## 5.0 CONCLUSION

It is mandatory for electrical engineer to understand the significance and measurement method of percentage impedance of the transformers. The same has been discussed in the article which will definitely help the electrical engineers especially young engineers to understand the percentage impedance of transformer and its significance in the field. The crux of the article is reproduced below:

1. All the flux in transformer will not be able to link with both the primary and

www.IndianJournals.com Members Copy, Not for Commercial Sale 4.139.253.48 downloaded 12-Aug-202

secondary windings. A small portion of flux will link either winding but not both. This portion of flux is called leakage flux. Due to this **leakage flux in transformer,** there will be a self-reactance in the concerned winding. This self-reactance of transformer is alternatively known as **leakage reactance of transformer**.

2. The percentage impedance of a transformer has a major effect on system fault levels. It determines the maximum value of current that will flow under fault conditions. It also plays an extremely vital role in voltage drop calculation. *High impedance transformers have a lower fault or short circuit current and have no need for high SCC rated breakers. However, they will have a higher voltage drop or regulation. Low Impedance transformers on the other hand have a lower voltage drop or regulation but have a higher fault or short circuit current and will need higher SCC rated breakers.*
3. It is worth to understand that percentage impedance is a function of geometry and construction of transformer. Simply % Z is the result of leakage flux i.e. to get low % Z, leakage flux should be minimum/ low.

## REFERENCES


[1] IS 2026 Specification of Power Transformer

[2] IEC 60076 Power Transformer

[3] Manual on "Transformer" Publication No 317, C.B.I.P. New Delhi.

[4] IEEE C57.12.90. Test Code for Liquid -Immersed Distribution , Power and Regulating Transformer

[5] Manual and Test Reports of Transformer Published by OEM


## BIOGRAPHICAL DETAIL OF THE AUTHOR

Dr. **Rajesh Kumar Arora** obtained the B. Tech. & Master of Engineering (ME) degrees in Electrical Engineering from Delhi College of Engineering, University of Delhi, India in 1999 and 2003 respectively. He completed his PhD in grounding system design from UPES, Dehradun. He is also certified Energy Manager and Auditor and has worked in 400kV and 220kV Substation for more than 14 years in Delhi Transco Limited (DTL). He has also worked as Deputy Director (Transmission and Distribution) in Delhi Electricity Regulatory Commission (DERC) for 03 years and 06 months. He has also given his contribution in the OS department of DTL for more than 2 years and rendered his services in the SLDC of Delhi Transco Limited (DTL) also. Presently he is working in D&E (Design and Engineering) department of DTL. His research interests include high voltage technology, grounding system, protection system, computer application and power distribution automation.

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 14.139.251.98 on dated 1-Aug-2022

The electric bill won't
give you a fright if you
remember to turn off the light

Journal of the Asia Pacific Economy

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rjap20

# What causes outward foreign direct investment (OFDI) from India into least developed countries (LDCs)?

Apoorva Jain & Sonal Thukral

**To cite this article:** Apoorva Jain & Sonal Thukral (2022): What causes outward foreign direct investment (OFDI) from India into least developed countries (LDCs)?, Journal of the Asia Pacific Economy, DOI: 10.1080/13547860.2022.2097619

**To link to this article:** https://doi.org/10.1080/13547860.2022.2097619

Published online: 19 Jul 2022.

Submit your article to this journal

Article views: 7

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rjap20

JOURNAL OF THE ASIA PACIFIC ECONOMY
https://doi.org/10.1080/13547860.2022.2097619

# What causes outward foreign direct investment (OFDI) from India into least developed countries (LDCs)?

Apoorva Jain and Sonal Thukral

Delhi School of Management, Delhi Technological University, Delhi, India

**ABSTRACT**

Although Outward Foreign Direct Investment (OFDI) from India into Least Developed Countries (LDCs) has been increasing amid various initiatives and policy reforms undertaken by the Government of India viz., Duty-Free Tariff Preference Scheme, Asia Africa Growth Corridor etc., India still does not feature amongst the top ten investor economies in LDCs. This makes it imperative to understand what causes OFDI from India into LDCs? Do the host country (LDCs) market-seeking and resource-seeking factors help in pushing OFDI from India into LDCs? For this purpose, the study employs Tobit regression model to empirically test the panel data of 13 LDCs during the period 2008–2018. The results indicate that OFDI from India in LDCs is both market-seeking as well as resource-seeking. It is also observed that OFDI in Asian LDCs is more than African LDCs. However, Bilateral Investment Treaties (BITs) between India and LDCs do not play a significant role in attracting FDI from India into LDCs. Finally, the study proposes important implications for policy makers in India *viz.* boosting its investment in African LDCs, encouraging private sector enterprises and capturing the untapped potential in the infrastructural projects in LDCs. The study lays down several implications for policy makers in LDCs also, *viz.* focusing on skill development of human resources, boosting its FDI in sectors like real estate, telecommunications, etc. and ensuring appropriate enforcement of BITs.

**KEYWORDS**

Outward foreign direct investment (OFDI); India; least developed countries (LDCs); panel tobit regression model

## 1. Introduction

Although, Least Developed Countries (LDCs) account for 13% of the world's population, their share in world Gross Domestic Product (GDP) is less than 1.3 per cent and the contribution to world trade is even less at 0.9 per cent. Not just this, LDCs average real GDP per capita is estimated at 1.7 per cent of that of developed nations (United Nations Department of Economic and Social Affairs 2018). The degree of underdevelopment in LDCs calls for integrating their economies with that of the world economy in order to boost their trade and investment. The predominant mode of LDCs entry into Global Value Chains (GVCs) is Foreign Direct Investment (FDI)

CONTACT Sonal Thukral sonalthukral@dtu.ac.in; sthukral1@gmail.com
© 2022 Informa UK Limited, trading as Taylor & Francis Group

(UNCTAD 2018b). FDI flows represent a relatively high share in LDCs total investments. During the first decade of twenty-first century (2001–2010), FDI flows contributed significantly towards the capital formation in LDCs (UNCTAD 2011). During 2013–2017, 21 per cent of the total capital flows to LDCs were represented by FDI (UNCTAD 2018a). In 2011–13, the average annual FDI flows to LDCs were three-fold the average flows in 2001–03. The rate of increase was higher than that of average annual FDI flows world-wide (UNOHRLLS 2006). According to UNCTAD data, during 2000–18, the inward FDI flows in LDCs has witnessed an upward trend with a CAGR of 9.66% (Figure 1). FDI undoubtedly plays a significant role in helping LDCs attain their sustainable development goals. By investing in poor countries, MNEs help eradicating poverty by generating employment (Winters, McCulloch, and McKay 2004), improves standard of living by offering quality products (Jain and Vachani 2006) and enhances people's access to education in LDCs (Driffield and Taylor 2000).

LDCs represent greater investment opportunities since the markets in developed nations are already saturated, hence MNCs seek for potential customers in underdeveloped nations (Musteen, Rhyne, and Zheng 2013). The importance of FDI has been rising in LDCs since a number of Fortune 500 companies are expressing greater interest in marking their presence in LDCs (UNCTAD 2011). Also, to help LDCs in integrating their economies with that of world economy, various international support measures in the area of (a) international trade, (b) development cooperation and (c) support for participation in United Nations systems, have been implemented (United Nations Department of Economic and Social Affairs 2018). Not just this, realizing the significance of FDI in boosting their economic development, LDCs themselves are undertaking greater initiatives in order to attract foreign investors. These measures ranges from setting up investment promotion agencies, entering into regional and bilateral investment agreements, relaxing their policy measures with respect to inward FDI as well as offering various fiscal incentives to the investors abroad (United Nations Secretariat 2015).

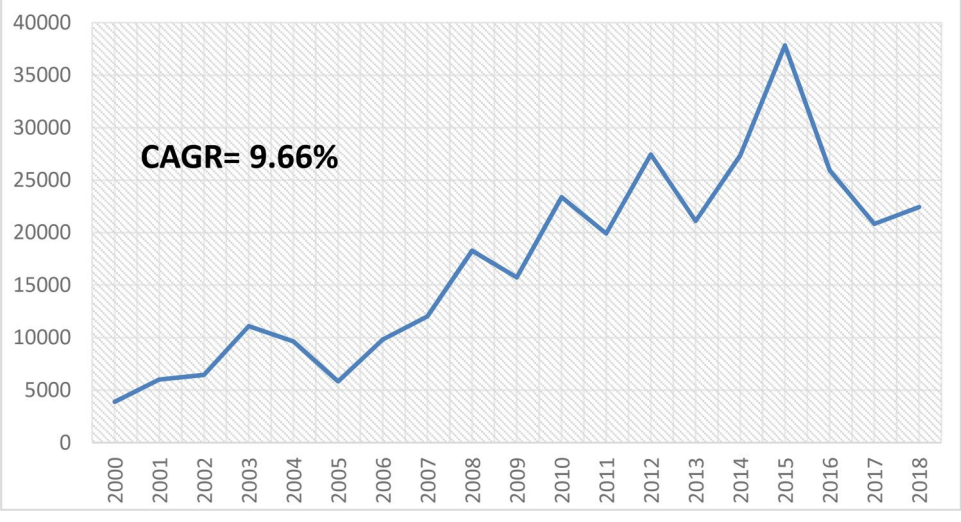



Figure 1. Inward FDI in LDCs during 2000–2018.Source: Authors' compilation from UNCTAD data.

Besides, during 2011 to 2021, significant reform efforts have been undertaken by LDCs in the area of investment policy. At least one policy measure on investment has been enacted by 66% of all LDCs in comparison to 54% of other developing countries. There is a wide range of investment policy measures that LDCs have adopted in order to attract FDI. Their focus has been on revising their investment laws, wherein 28% of LDCs introduced, replaced, or reformed their investment-specific legislation, compared to 15% of other developing countries (UNCTAD 2022). Thus, it is observed that LDCs are undertaking active measures to attract inward investment in comparison to other economies in the world. These measures range from enhancing investment treatment, opening new sectors and activities for FDI, creating investment promotion agencies, and enhancing protection guarantees. Over one-third of these policy measures are entry-related measures that focus on ease of investor entry in these LDCs (UNCTAD 2022). Thus, given the significant efforts undertaken by the LDCs in attracting FDI compared to other developing economies, it provides various opportunities for the firms to invest in LDCs in comparison to the rest of the world.

It is observed that emerging economies like China, Malaysia, India and South Africa are showing greater interests in boosting their trade and investment in Least Developed Countries (LDCs). Particularly, Indian firms have begun playing a significant role in facilitating FDI in African LDCs, mainly due to the presence of abundant natural resources in Africa (UNCTAD 2011). LDCs represent greater scope for India since Indian firms have started investing in the extraction of crude oil and natural gas (UNCTAD 2011). India was the leader among developing economies in terms of greenfield FDI in LDCs in 2012–13 (UNOHRLLS 2020a). Also, during 2015–19, India was among the top three developing and transition economies in terms of its average share in greenfield projects in LDCs (UNCTAD 2020). According to the RBI data, during 2008–18, even though OFDI from India into LDCs holds a small proportion in comparison to the rest of the world, however, the growth rate is eye-catching. The CAGR of OFDI from India into LDCs is 3.14% while for the rest of the world the CAGR is 4.19% (Figure 2). The growth rate indicates that the next destination which India seems to be eyeing is LDCs.

Indian MNCs are motivated to invest in LDCs since developed markets are saturating over time. Thus, it becomes imperative for these companies to identify and capture other markets in order to maintain their growth. Indian firms will be able to seize the first-mover advantage in LDCs by capturing the untapped potential in these economies. Further, by investing in LDCs, Indian firms will be able to diversify their international investment and thus will reduce their exposure to some particular markets. Consequently, it will provide immense opportunities for Indian MNCs to ultimately increase their profitability by undertaking OFDI in LDCs. Thus, given the greater scope and significant location advantages associated with LDCs, various initiatives and policies have been undertaken by the Government of India in order to boost its bilateral ties with LDCs. The Government of India implemented the Duty-Free Tariff Preference (DFTP) scheme in 2008 to grant duty free market access to all LDCs. The government has also undertaken several strategic initiatives in order to strengthen its relationship with African countries such as 'Focus Africa' programme, Asia-Africa Growth Corridor and India Africa Forum Summit.

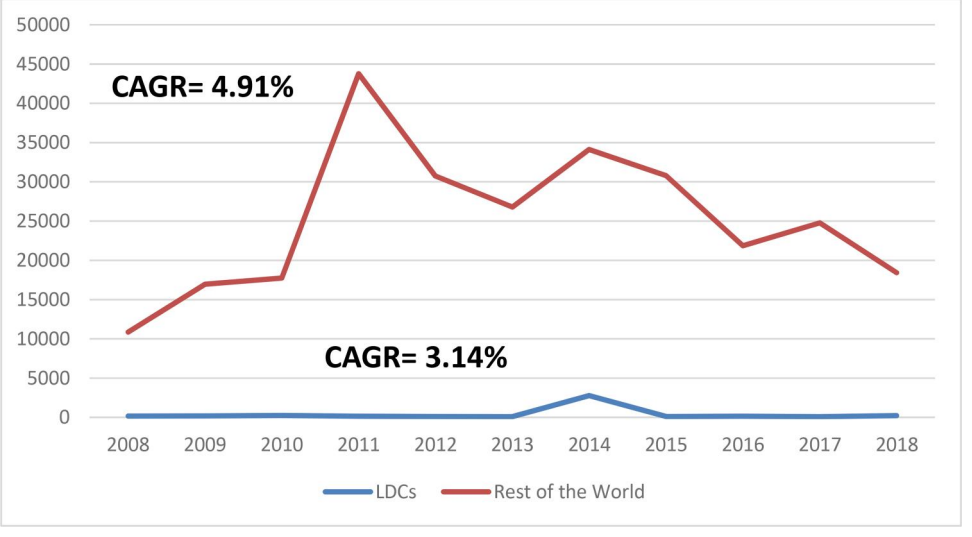



Figure 2. OFDI from India into LDCs vs Rest of the World (2008–18).Source: Authors' Compilation from RBI data.

Notwithstanding several initiatives and reforms undertaken by Government of India, India still does not feature among the top ten investor economies in LDCs. Having said that, in the light of various policy measures been undertaken, it becomes imperative to study the factors determining the outward foreign direct investment from India in LDCs. Also, despite the significant role that foreign direct investment plays in fostering the growth and development of LDCs, not much efforts have been made in understanding the impact of various market-seeking as well as resources-seeking factors in attracting FDI in LDCs. Although certain attempts have been made to study the foreign direct investment in LDCs (Malgwi *et al.* 2006; Sauvant 2015), and several attempts have been made in studying outward foreign direct investment from India (Tripathi and Thukral 2016; Pradhan 2017; Tripathi and Thukral 2018a), till date, to the best of our knowledge, there hasn't been any attempt made to study OFDI flows from India into LDCs. The aim of the present study is thus to examine the effect of host country (LDCs) market-seeking as well as resource-seeking factors on OFDI flows from India during 2008–2018. Since our dependent variable *OFDI* is a censored variable, we employ Tobit regression model to empirically test the panel data of 13 LDCs. The results from the study present important implications in terms of what can be done by policy makers in India, LDCs, and what they can do together in order to boost investment relationship between India and LDCs.

## 2. FDI trends and patterns in least developed countries (LDCs)

The United Nations define Least Developed Countries (LDCs) as '*low-income countries confronting severe structural impediments to sustainable growth. They are highly vulnerable to economic and environmental shocks and have low levels of human assets.*' Committee for Development Policy, a subsidiary body of United Nations, reviews the list of Least Developed Countries (LDCs) every three years and according to its latest

review in March 2018, 47 countries qualified to be designated with the status of least developed countries. In its recent review, the committee adopted three criteria- per capita income, economic vulnerability and human assets to prepare the list of LDCs (UNCTAD 2019a).

In order to achieve the annual GDP growth rate of at least 7% in LDCs as put forward by Brussels Programme of Action (BPoA), attracting FDI has become one of the key strategies of LDCs. During the decade 2001–10, FDI inflows in LDCs exceeded foreign portfolio investment and all other investments combined, making it the most important external private capital flows for LDCs. In 2008, FDI flows accounted for 28 per cent of the LDCs Gross Fixed Capital Formation (UNCTAD 2011). The proportion of FDI inflows in LDCs is small in comparison to developed and developing nations, however, for some LDCs particularly Liberia, Cambodia and Sierra Leone, FDI represented a relatively large proportion of their nation's GDP in 2016 (UNCTAD 2018b). During 2013–2017, 21 per cent of the total capital flows to LDCs were represented by foreign direct investment (UNCTAD 2018a). Although, LDCs reported six-year low FDI inflows in 2017, however, they rebounded to US$24 billion in 2018, nearly 15 per cent increase from 2017. In 2018, Bangladesh and Myanmar stood at first position in terms of FDI inflows amongst all LDCs, followed by Ethiopia, Cambodia and Mozambique (see Figure 3) (UNCTAD 2019b). Albeit the recovery in 2018, FDI flows in 2019 again tumbled by 5.7% to US$21.1 billion (see Figure 4). Much of the decline in FDI during this period could be attributed to natural resources price shocks, since majority of LDCs are oil dependent countries which help them attract investments from abroad (UNCTAD 2020).

LDCs are dominated by African and Asian LDCs, wherein 72% of the total inward FDI in LDCs is accounted by African LDCs, 27% by Asian LDCs and 1% by others (Figure 5). Although, Africa is an important business context, little attention has been paid to the firms in Africa in the mainstream management literature (Barnard, Cuervo-Cazurra, and Manning 2017). While FDI flows in African LDCs mostly targeted extraction industries such as oil and mineral resources, for Asian LDCs it was

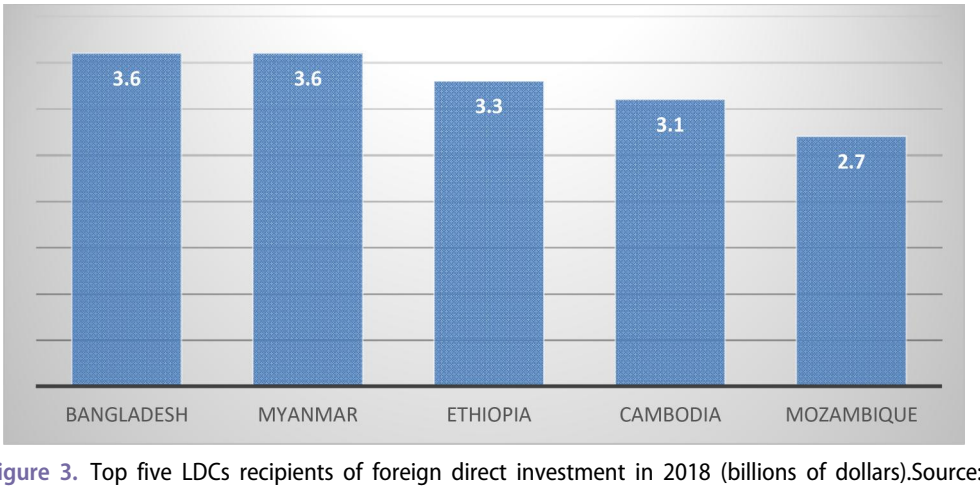



Figure 3. Top five LDCs recipients of foreign direct investment in 2018 (billions of dollars).Source: (UNCTAD 2019b).

Figure 4. FDI Inflows in LDCs, 2000–2019 (Billions of Dollars and Per Cent).Source: (UNCTAD 2020).

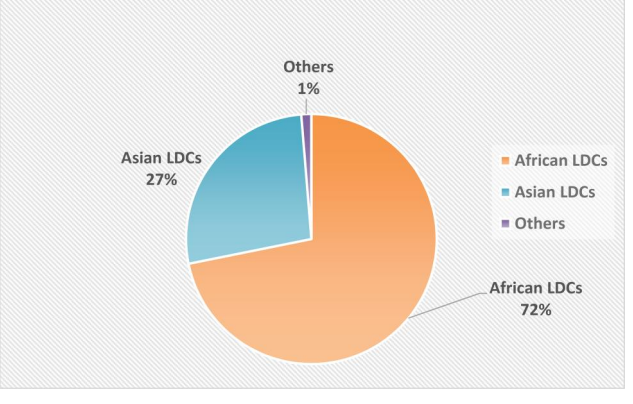



Figure 5. Region wise distribution of inward FDI in LDCs (2000–18).Source: Authors' Compilation from UNCTAD data.

the service industries such as electricity and telecommunications which received majority FDI flows. In African LDCs, manufacturing sector has not been able to attract much of FDI flows, mainly because of lack of skilled workers (UNCTAD 2011). In 2017, while FDI flows for Asian and Oceanian LDCs rose, they declined by around 31% for Africa LDCs. FDI inflows in Asian and Oceanian LDCs are mostly driven by strong investment in real estate and telecommunication development. It is observed that investment in large scale infrastructure projects is significant for LDCs in order to attract FDI flows (UNCTAD 2018a). In 2018, although FDI flows increased in both African as well as Asian LDCs, however, they touched a record high in Asian LDCs. Again, increased FDI flows in Asian LDCs were mostly driven by the investment in power generation as well as real estate projects. Also, as expected, FDI in Africa was mostly targeted in extraction and processing of natural resources (UNCTAD 2019b).

## 3. Relationship between India and least developed countries (LDCs)

Over the years, India has nurtured its relationship with least developed countries by setting up long-lasting partnerships and bilateral relations with them. India has been

expressing greater interest when it comes to extending its support and integrating with LDCs. India, in fact was actively engaged in promoting the idea of creating a separate category of LDCs in the 2nd UNCTAD session held in New Delhi in 1968. India has contributed in shaping the long-term sustainability in LDCs by extending its assistance in the form of technical expertise, infrastructure development and institutional strength (Ministry of External Affairs 2011). Indian market is considered as one of the attractive destinations for LDCs exports, enabling them to expand their exports in India by nine-fold during 2003–2011. During the same period, Indian market stood at the fourth position for LDCs exports (United Nations 2012). By an official notification on 13 August 2008, The Government of India implemented the *Duty-Free Tariff Preference* (DFTP) scheme to grant duty free market access to all LDCs. This means LDCs get the benefit of unilateral tariff preferences for all the products which are originated in LDCs and are imported into India. Subsequently, the Government of India amended the scheme in April 2014 wherein the coverage of tariff lines for duty free access to Indian market was increased further (International Trade Centre 2015). As a support measure for LDCs, recently in May 2020 India decided to contribute towards the areas like broadband infrastructure and digital skills in order to promote the capacity building in LDCs. These measures were adopted owing to the fact that access to high speed broadband is still a distant dream for majority of the population in least developed countries (The Economic Times 2020).

Moreover, Government of India has been taking several strategic initiatives in order to strengthen its relationship with African countries that constitutes major proportion of LDCs. For instance, Government of India launched 'Focus Africa' programme in 2002 in order to boost its trade and investment relations with Africa. The aim of the programme is to tap the tremendous trade potential in sub-Saharan African region. The programme focused on promoting the trade of specific commodities, e.g. cotton yarns, drugs, machinery and instruments, and telecom and information technology between India and African countries. Under this programme, Indian Government extended financial assistance to multinationals and trade agencies in order to boost their trade and investment with African countries (The Economic Times 2002). Also, India launched Asia-Africa Growth Corridor in 2017 in order to boost economic cooperation between the governments of India, Japan and multiple African countries. The aim of India is to develop quality infrastructural facilities in African countries in the areas of health, agriculture, skill enhancement and disaster management. It was aimed at enhancing the interconnectedness between Asia and Africa in order to develop a free and open Indo-Pacific region (Research and Information System for Developing Countries 2017). To further enhance the ties between India and Africa, India-Africa Forum Summit is held once in every three years, whose objective is to strengthen the economic cooperation between India and Africa with respect to peace and security, energy access, education and skill enhancement and sustainable development (Beri 2015).

All the initiatives undertaken by the Indian Government in order to boost its bilateral ties with LDCs indicate the growing interests of the Indian MNEs to capture the untapped potential in LDCs. Thus, what factors influence the investment decisions of Indian MNEs in LDCs is worth catching the attention.

## 4. Theoretical background: Eclectic paradigm

John H. Dunning's Eclectic Paradigm or OLI (Ownership, Location, Internalisation) framework was first propounded in 1979. In his Eclectic Paradigm, John H. Dunning postulates that a multinational corporation will engage in overseas investment only when it has obtained three types of advantages- Ownership Advantages (O), Location Advantages (L) and Internalization Advantages (I). A firm acquires *ownership advantages* once it has developed production techniques, skills and strong brand image which helps the multinational in competing with local firms abroad. *Location advantages* are with respect to host country which include the endowment of rich natural resources, tax incentives, skilful labour, etc. which makes the host country an attractive destination for investment by foreign investors. *Internalisation advantages* are the ones wherein a multinational firm attains benefits in self production rather than outsourcing it to the other party. As a result, a firm will prefer to set up its production facilities in a host country in order to keep its operations in-house only (Dunning 1982). Thus, it was said that ownership (O) and internalisation (I) advantages are push factors, the attainment of which pushes the multinational firm to engage in foreign direct investment, while, location (L) advantages are pull factors, the procurement of which makes the host country attractive to pull investment from investors abroad (Anh and Ngoc 2016).

For the host country, L advantages are important since a multinational corporation will not invest in a host country unless it has some locational advantages attached to it. With respect to locational advantages, the foreign direct investment can have four motives, or it can be divided into four types of investment seeking; market-seeking, resource-seeking, asset-seeking and efficiency seeking. Market-seeking FDI is one where the investing firms benefits from serving the host market. FDI will be market-seeking if the host country market size is large, there is growing consumer market, the economic standards and income level of the people living in host nation is good and the host country has stable economic and political environment. FDI is resource-seeking when investing firm seeks to access the endowment of natural resources, skilled human resources and infrastructural facilities in the host nation. FDI is strategic asset-seeking wherein a multinational firm wants to seek access to host nation's technological capabilities, local knowledge and innovation capabilities. The access to these capabilities provides a firm with the information advantages over other competitors in the host country. Finally, FDI is efficiency-seeking when it helps the investing firm to minimize the overall cost of production, ultimately enhancing the efficiency of the firm. The objective here is to rationalise the production process by developing an efficient production network (Dunning and Narula 1994). As we have already discussed that FDI in LDCs is mainly targeted towards abundant natural resources of African LDCs and rising infrastructural facilities of Asian LDCs. Hence, at this stage, we believe that it is market-seeking as well as resource-seeking factors that play a major role in attracting FDI to LDCs. With this conjecture, we build this paper upon the market-seeking as well as resource-seeking locational advantages of the OLI framework which shall help push OFDI from Indian into LDCs.

## 5. Literature review

Foreign Direct Investment (FDI) can act as a catalyst in promoting capacity building in least developed countries. Such investment will not only help in bringing the much-needed technology, income, employment, capital and managerial know-how in LDCs but the linkages and spill over benefits from such investment will also help in strengthening the domestic firms and human resources in LDCs. This ultimately stimulates their growth and development and help in improving their standards of living (Kumari and Sharma 2017). Traditionally, natural resources were the main source of FDI in LDCs, however the trend has been changing gradually. During 2003–2010, the share of greenfield FDI projects in primary, manufacturing and services was 55%, 28% and 17% respectively, however, these shares were 21%, 28% and 51% respectively during 2011–13. From the home country perspective too, LDCs represent greater investment opportunities since the markets in developed nations are already saturated and hence MNCs seek for potential customers in underdeveloped nations (Musteen, Rhyne, and Zheng 2013). It thus becomes imperative to understand the role of various resource-seeking as well as market-seeking factors in LDCs that help them in attracting FDI from India. The remaining section of literature review discusses these factors in detail and establishes their relationship with OFDI from India.

### *5.1. Resource seeking factors and FDI in LDCs*

Endowment of rich resources in host country is one of the important determinants in attracting FDI to their economies (Asiedu 2006; Ezeoha and Cattaneo 2012). In a study it was evident that FDI in developing economies is more resource seeking than market seeking (Wadhwa and Reddy 2001). Generally, developed countries obtain natural resources by investing in LDCs where the extraction of such resources is cheaper and hence bolster their profit (Jorgenson, Austin, and Dick 2009). Currently, 33 out of 47 Least Developed Countries (LDCs) belong to Africa (UNOHRLLS 2020b). Africa being endowed with rich mineral resources becomes an attractive destination for foreign direct investment, despite the existence of inefficient and weak institutions (Ezeoha and Cattaneo 2012). It was also evident that although emerging economies don't prefer to invest in countries having weak institutions, however this negative effect diminishes if the host country is endowed with enormous natural resources. This means they are willing to invest in 'resource curse' economies, that is the countries who are suffering from weak institutions but are endowed with substantial natural resources (Aleksynska and Havrylchyk 2013). However, it is also observed that the availability of resources stimulate resource FDI but at the same time crowds out non-resource FDI. As a result, natural resource endowed countries suffer from lower aggregate amounts of FDI (Poelhekke and Van der Ploeg 2010). It is clearly evident that LDCs are endowed with abundant natural resources, particularly African LDCs. More than 80% of the FDI inflows in LDCs were accounted by resource rich African nations (UNCTAD 2011). According to The World Investment Report 2019, one of oil rich LDC- Chad, doubled its FDI inflows to US$662 million in 2018, majority of being concentrated in new oil projects. The same report indicates that the

new oil and gas projects in Mozambique and Myanmar will help both the countries in attracting significant levels of FDI in the coming years (UNCTAD 2019b). Thus, we propose,

Hypothesis 1a: The Outward Foreign Direct Investment (OFDI) from India to LDCs increases as the endowment of natural resources in LDCs increases

Not just natural resources but the availability of educated labour force also helps in attracting FDI in a country (Asiedu 2006; Campos and Kinoshita 2003). Earlier, when the education system in LDCs was governed by European Colonizers, the indigenous literary practices in schooling were neglected and limited knowledge sharing was there. As a result, in LDCs, the education system was not able to generate graduates who are qualified enough to seek jobs at the global marketplaces (Jarvis 2006; Ubah 1980). Skill deficit in the human capital of LDCs is the major constraint in achieving the competitiveness at the international marketplaces. The reason being the limited investment in human resource training and lower levels of secondary and tertiary education enrolment (Adhikari 2019). While entering the foreign markets, MNEs tend to invest a significant proportion of their capital in R&D activities in order to generate new technologies. Hence, they prefer educated human resources in the host country that is able to handle and work well with this new technology (Aziz and Mishra 2016; Hakro and Omezzine 2011). Multinationals these days not only look for cost cutting but also seek for highly qualified and skilled human resources in order to enhance operational efficiency (Lewin, Massini, and Peeters 2009; Mohiuddin and Su 2013). It is observed that in order to develop competent public servants, it is important that LDCs should have effective human resource development programs focusing on employee training (Tessema, Winrow, and Teclezion 2012). In LDCs, specifically in Africa, government initiative to invest in education is likely to deliver positive macroeconomic results, ultimately leading to strong economic development (Basten and Cuaresma 2014). Thus we propose

Hypothesis 1b: The Outward Foreign Direct Investment (OFDI) from India to LDCs increases as the quality of human resources in LDCs improves

Also, good infrastructure in a country helps in attracting FDI from abroad (Asiedu 2006; Campos and Kinoshita 2003). Infrastructure is considered as one of the most important resource seeking factor that helps in attracting more FDI. As the infrastructure of the country develops, the productivity of the investment improves and hence FDI flows rises (Loree and Guisinger 1995; Wheeler and Mody 1992). Infrastructural development in an economy in seen in several areas, one of which is the number of internet users. As the number of internet users in a country increases, its ability to attract FDI also increases. This is because, internet can play a significant role in reducing the distance between the countries and hence makes it favourable for the countries to communicate (Choi 2003; Kahouli and Maktouf 2015). Among developing countries, LDCs are categorized at the bottom of the development scale since they have very challenging environment (Cuervo-Cazurra and Genc 2008). According to UNCTAD statistics, African LDCs exhibit road density of just 15% of that of developing nations (UNCTAD 2015). Also, the electricity coverage in more than 15 LDCs is less than 25% (UNCTAD 2017). While shifting their production base to

LDCs, firms face challenges, particularly those from developed economies, since the quality of infrastructure in terms of transportation networks, communication networks and distribution channels with which they work in their home country is not present in LDCs (Prahalad and Hammond 2002; Prahalad and Lieberthal 1998). However, LDCs represent huge growth potential and can act as an important source of revenue for MNEs (Prahalad 2004). In 2018, seeing the growth potential of LDCs, MNEs from the United States and China made large scale investment in power generation infrastructure of LDCs. According to The World Investment report of 2019, the recent development of industrial parks in Ethiopia will help country to attract FDI inflows exceeding US$5 billion in the coming years (UNCTAD 2019b). Hence, we propose,

Hypothesis 1c: The Outward Foreign Direct Investment (OFDI) from India to LDCs increases as the quality of infrastructure of LDCs improves

### 5.2. Market seeking factors and FDI in LDCs

Least Developed Countries (LDCs) are characterized with low levels of development but large consumer base, hence offering greater potential for MNCs to explore (Musteen, Rhyne, and Zheng 2013). Greater market size helps in attracting FDI from countries abroad (Milner and Pentecost 1996; Nunnenkamp 2002; Root and Ahmed 1978). The market size of an economy gives an idea of its economic standards, growth and development; and hence any improvement in these indicators help a nation in attracting more FDI (Rjoub *et al.* 2017; Montanari, Giraldi, and Galina 2019). While investing abroad, multinationals are not just searching for production site but also a growing market for their products. Thus, as the market size increases, economies are able to attract more investment from multinationals (Blomström, Lipsey, and Blomstrom 1991; Belkhodja, Mohiuddin, and Karuranga 2017; Na and Lightfoot 2006). Multinationals always prefer a well-built economy where the possibility of capturing a greater market share is high (Janicki and Wunnava 2004). Also, it is believed that higher market size of an economy is an indicator of higher per capita income and higher levels of development, which in turn attract foreign investors to invest in such economies (Kumari and Sharma 2017). However, since developed economies are usually characterized with ageing population and slow economic growth, it stimulates multinationals to look for developing economies for investment opportunities (Belkhodja *et al.*, 2016). Further, LDCs represent greater investment opportunities since the markets in developed nations are already saturated and hence MNCs seek for potential customers in underdeveloped nations (Musteen, Rhyne, and Zheng 2013). Hence, we propose,

Hypothesis 2a: The Outward Foreign Direct Investment (OFDI) from India to LDCs increases as the market size of LDCs increases

Another characteristic determining the growth of an economy is its income level. It is observed that as the Gross National Income (GNI) of an economy increases, its ability to attract inward FDI also increases. This happens because the high levels of GNI signals the growth of an economy and thus attracts more investors (Turnbull,

Sun, and Anwar 2016). Gross National Income (GNI) per capita is one the defining criteria for LDCs. According to the current definition, if for any country GNI per capita falls below US$1025 then that country is identified as LDC by United Nations (UNCTAD 2019a). GNI being the important criteria in defining LDCs, it thus makes it imperative to study the impact of LDCs national income on its inward FDI. A very limited literature in this area indicates that MNEs from emerging economies prefer to invest in those LDCs where the average income of the citizens is high. Thus, LDCs having higher GNI per capita are able to attract more FDI from investors abroad (Cuervo-Cazurra and Genc 2008). LDCs exhibiting rapid economic growth and higher per capita income attract more FDI since it signals government initiative in adopting more rational policies in their economy (Lim 1983). According to a review conducted in 2018, 17 out of 47 LDCs had a GNI per capita above the graduation threshold (United Nations Department of Economic and Social Affairs 2018). The statistics indicate that several LDCs are experiencing growth in their levels of gross national income, ultimately leading to their economic growth. Thus, we propose,

Hypothesis 2b: The Outward Foreign Direct Investment (OFDI) from India to LDCs increases as the national income of LDCs increases

Also, trade openness helps economies in attracting investment from foreign investors (Asiedu 2006; Campos and Kinoshita 2003). Trade and investments are complimentary to each other. Thus, as the trade openness increases, the integration between the economies also increases and hence helps in attracting more FDI (Deichmann 2001; Janicki and Wunnava 2004). As the economy opens its trade, the relationship between the countries strengthen and hence provides investors with greater investment opportunities (Tintin 2013). Opening up of the trade aids the investor in better understanding the host country local market conditions and hence trade openness helps the host country in attracting FDI (Aziz and Mishra 2016; Elfakhani and Matar 2007). For instance, in 2017, Haiti (one of the LDCs) reported record high FDI inflows in its export-oriented textile industry due its duty free access to the US market (UNCTAD 2019b). Also, as the bilateral relations between the countries enhances, the FDI flows between them also increases (Laabas and Abdmoulah 2009). Preferential Trade Agreements (PTA) help in attracting FDI by strengthening the relationships between the countries and hence removing all the obstacles in fully exploiting the benefits of preferential treatment (Aziz and Mishra 2016; Cardamone and Scoppola 2012). These agreements promote investment between the countries by improving the overall provisions in the areas of dispute settlement, competition policy, investment and intellectual property rights (Aziz and Mishra 2016; Medvedev 2012). Back in 2001, during the Third UN Conference on the Least Developed Countries, 29 Bilateral Investment Treaties (BIT) were signed by nine LDCs with both developed and developing nations around the globe, in order to pave the way for rising FDI inflows and ultimately economic growth (UNCTAD 2001). They are showing active participation in signing Double Taxation Treaties (DTT) and Bilateral Investment Treaties (BIT) which is indeed creating a conducive investment environment in LDCs. By 2010, 188 DTTs and 455 BITs were concluded by LDCs (UNCTAD 2011). Also, according to The Least Developed Countries Report of 2018, all LDCs all the signatories to Bilateral Investment Treaties (BIT) which lead to a

significant growth of 27.6% in their export volumes in between 2000 and 2016 (UNCTAD 2018b). Hence, we propose,

Hypothesis 2c: The Outward Foreign Direct Investment (OFDI) from India to LDCs increases as the Trade Openness in LDCs increases

Hypothesis 2d: The Outward Foreign Direct Investment (OFDI) from India to LDCs increases as LDCs sign more Bilateral Investment Treaties (BIT) with India

## 6. Data and methodology

### 6.1. Data sources and variables

The present study aims at analysing the impact of various market-seeking as well as resource-seeking factors on Outward Foreign Direct Investment (OFDI) by Indian multinationals in Least Developed Countries (LDCs). The study empirically tests the panel data of 13 LDCs- Afghanistan, Bangladesh, Bhutan, Malawi, Mali, Mauritania, Mozambique, Nepal, Niger, Rwanda, Senegal, Sierra Leone and Tanzania from 2008 to 2018. The LDCs in our data set belong to African and Asian region. In our sample, 7 countries are classified as both low income countries as well as LDCs, while 6 are classified as only LDCs.[1] Thus, the complete sample represent LDCs. For the purpose of identifying LDCs, we follow the most recent review by united Nations in March 2018 that identified 47 LDCs (UNCTAD 2019a). The list includes 31 African LDCs, 8 Asian LDCs, 7 Island LDCs and 1 American LDC. Our sample of 13 LDCs include 9 African and 4 Asian LDCs, hence not restricting it to any specific region and thereby rendering more generalizability. We have restricted our sample size to 13 LDCs and our sample period to 2008–2018 due to limited availability of data for the underlying variables.

We intend to conduct our study based upon the sample of LDCs for various important reasons. First, LDCs represent huge growth potential and can act as an important source of revenue for MNEs. Seeing this growth potential, MNEs from various regions are showing greater interests in making large scale investment in LDCs. In recent years, certain LDCs themselves have started making investment in their infrastructure projects to attract FDI from investors abroad (UNCTAD 2019b). Hence, it becomes important to understand the stance of India in tapping on the rising growth potential in LDCs. Second, although certain attempts have been made to study the foreign direct investment in LDCs (Malgwi *et al.* 2006; Sauvant 2015) and several attempts have been made in studying outward foreign direct investment from India (Tripathi and Thukral 2018b; Pradhan 2017), till date, to the best of our knowledge, there hasn't been any attempt made to study the impact of various market-seeking as well as resource-seeking factors on OFDI by India in case of LDCs. Third, over the years there has been various international and regional measures taken (discussed in section 3) to boost trade and investment in LDCs, hence in the light of these policy measures it becomes imperative to understand the investment relationship between India and LDCs.

The variables and the data sources for all the underlying variables studies in the paper are listed below.

#### 6.2.1. Dependent variable

*6.2.1.1. Outward foreign direct investment (OFDI).* The variable measures OFDI from India in Least Developed Countries (LDCs). Data corresponding to OFDI is collected from monthly Reserve Bank of India (RBI) reports on overseas investment. The report provides the data on monthly overseas investment by Indian multinationals in countries abroad. From that dataset we created a panel of 13 LDCs which are receiving investment by Indian multinationals for a period of 11 years from 2008 to 2018.

#### 6.2.2. Independent variables

The study examines the impact of various market-seeking and resource-seeking factors in LDCs that help attract investment in these countries. The study employs three resource- seeking variables viz., natural resource endowment, human resource development and infrastructure development and four market seeking variables viz., market size, GNI per capita, trade openness and Bilateral Investment Treaties (BIT). The description of the variables and their data sources is listed below in Table 1.

#### 6.2.3. Control variables

In all our regression models, we control for two variables. First, inflation, which measures the annual percentage change in the basket of goods and services in an economy. It is measured by consumer price index. The data corresponding to this variable is collected from World Bank website (www.worldbank.org). Second, exchange rate, which determines the conversion rate at which a nation's currency is exchanged for another. The data corresponding to this variable is collected from World Bank website (www.worldbank.org).

### 6.2. Methodology

The study employs Tobit regression model to estimate the effects of host country (LDCs) market-seeking as well as resource-seeking factors on OFDI flows from India. Tobit models are suitable in cases where the regression model includes a censored or limited dependent variable so that its value is observable only at a specific range. Thus, when dependent variable follows a mixed distribution wherein it takes a value zero (0) when the event does not happen while follows a continuous distribution for all the values greater than zero, it makes Ordinary Least Square (OLS) a biased and unsuitable method in such cases. Under OLS, all the observations where latent variable ($y_i^*$) $\leq 0$ are ignored and hence, the distribution of error terms becomes a truncated normal distribution (Maddala 1983) resulting in biased estimators. Hence, Tobit regression model becomes a more suitable approach. The Tobit model supposes that there is latent variable $y_i^*$ which linearly depends upon the set of independent variables. The observable variable $y_i$ becomes equal to the latent variable $y_i^*$ in all the cases where latent variable is above zero, and zero otherwise. In the panel Tobit model, the latent variable $y_i^*$ is specified as in equation 1,

**Table 1.** Description of independent variables.

| Variable name | Variable description | Abbreviation | Source |
|---|---|---|---|
| *Education of Human Resources (log)* | Total number of people enrolled in secondary education in a country | Logeduc | www.worldbank.org. . . |
| *Natural Resource Endowments (log)* | Total natural resources rents which includes oil rents, natural gas rents, coal rents, mineral rents and forest rents that a country receives | Lognatural | www.worldbank.org. . . |
| *Infrastructure Development Index* | Index created using three components- Air Transport Freight, Access to Electricity and Mobile Cellular Subscription | Infra | www.worldbank.org. . . , (Index is built by authors using Principal Component Analysis) |
| *Market Size (log)* | Total population | logpop | www.worldbank.org. . . |
| *GNI per capita (log)* | GNI per capita | logGNI | www.worldbank.org. . . |
| *Trade Openness* | Ratio of the sum of imports and exports to GDP | tradeop | www.worldbank.org. . . |
| *Bilateral Investment Treaties* | Binary variable; 1 for all those years where BIT is signed between India and LDCs and 0 otherwise | BIT | investmentpolicy.unctad.org |
| *Country Dummy (1 = African LDCs, 0 = Asian LDCs)* | Country Dummy | countrydummy | – |
| *Country Dummy*Education of Human Resources (Interaction Term)* | Interaction Term | countrydummy_educ | – |
| *Country Dummy*Natural Resource Endowments (Interaction Term)* | Interaction Term | countrydummy_natural | – |
| *Country Dummy*Infrastructure Development Index (Interaction Term)* | Interaction Term | countrydummy_infra | – |
| *Country Dummy*Market Size (Interaction Term)* | Interaction Term | countrydummy_pop | – |
| *Country Dummy*GNI per capita (Interaction Term)* | Interaction Term | countrydummy_GNI | – |
| *Country Dummy*Trade Openness (Interaction Term)* | Interaction Term | countrydummy_tradeop | – |
| *Country Dummy*Bilateral Investment Treaties (Interaction Term)* | Interaction Term | countrydummy_BIT | – |

$$y_{it}* = \alpha_i + x'_{it}\beta + \mu_{it}, \quad i = 1,2, \ldots, N, \; t = 1,2, \ldots, T$$
$$\mu_{it} = v_i + \epsilon_{it}, \quad \left[v_i \sim N(0,\sigma_v{}^2)\right], \; \left[\epsilon_{it} \sim N(0, \sigma_\epsilon{}^2)\right] \tag{1}$$

Herein, $y_{it}*$ can be observed if $y_{it}* > 0$ and cannot be observed where $y_{it}* \leq 0$. The observed variable $y_{it}$ is specified as in equation 2,

$$y_{it} = max(0, y_{it}*) \quad Or \quad y_{it} = y_{it}* \; if \; y_{it}* > 0 = 0 \; if \; y_{it}* \leq 0 \tag{2}$$

In line with past researchers, we adopt random effects Tobit model due to the limitation of inconsistent estimators associated with fixed effects model. The random effect logarithmic likelihood function of □, $\sigma_u{}^2$, $\sigma_\alpha{}^2$ is specified in equation 3 (Cameron and Trivedi 2005),

$$f(y_i|\ X_i,\ \beta,\ \sigma_u^2,\ \sigma_\alpha^2) = \int f(y_i|\ X_i,\ \alpha_i,\ \beta,\ \sigma_u^2)\ \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} antilog\left(\frac{-\alpha_i}{2\sigma_\alpha^2}\right)^2 d\alpha_i \qquad (3)$$

Our dependent variable, i.e. Outward Foreign Direct Investment (OFDI) is a censored variable as there are instances when there is no overseas investment by Indian multinationals in LDCs for the period 2008–2018. Hence, our dependent variable takes a value zero for all those periods when there is no OFDI while follows a continuous distribution otherwise. As a result, we observe censoring from left. Hence, the study employs Tobit regression model to efficiently account for the problem of left censoring. To observe the impact of various resource-seeking and market-seeking factors on OFDI from India in LDCs, we present the following models:

- **Resource-seeking models**

**Model 1**

$$logOFDI_{it} = \beta_0 + \beta_1 logeduc_{it} + \beta_2 lognatural_{it} + \beta_3 Infra_{it} + \beta_4 countrydummy_{it} + \beta_5 Inflation_{it} + \beta_6 logexchange_{it} + \mu_{it}$$

**Model 2**

$$logOFDI_{it} = \beta_0 + \beta_1 logeduc_{it} + \beta_2 lognatural_{it} + \beta_3 Infra_{it} + \beta_4 countrydummy_{it} + \beta_5 countrydummy_{edu}c_{it} + \beta_6 Inflationit + \beta_7 logexchange_{it} + \mu_{it}$$

**Model 3**

$$logOFDI_{it} = \beta_0 + \beta_1 logeduc_{it} + \beta_2 lognatural_{it} + \beta_3 Infra_{it} + \beta_4 countrydummy_{it} + \beta_5 countrydummy_{natura}l_{it} + \beta_6 Inflationit + \beta_7 logexchange_{it} + \mu_{it}$$

**Model 4**

$$logOFDI_{it} = \beta_0 + \beta_1 logeduc_{it} + \beta_2 lognatural_{it} + \beta_3 Infra_{it} + \beta_4 countrydummy_{it} + \beta_5 countrydummy_{infr}a_{it} + \beta_6 Inflationit + \beta_7 logexchange_{it} + \mu_{it}$$

**Model 5**

$$logOFDI_{it} = \beta_0 + \beta_1 logeduc_{it} + \beta_2 lognatural_{it} + \beta_3 Infra_{it} + \beta_4 countrydummy_{it} + \beta_5 countrydummy_{edu}c_{it} + \beta_6 countrydummy_{natura}l_{it} + \beta_7 countrydummy_{infr}a_{it} + \beta_8 Inflationit + \beta_9 logexchange_{it} + \mu_{it}$$

- **Market-seeking models**

**Model 6**

$$logOFDI_{it} = \beta_0 + \beta_3 logpop_{it} + \beta_4 logGNI_{it} + \beta_1 tradeop_{it} + \beta_2 BIT_{it} + \beta_5 countrydummy_{it} + \beta_6 Inflation_{it} + \beta_7 logexchange_{it} + \mu_{it}$$

**Model 7**

$$logOFDI_{it} = \beta_0 + \beta_3 logpop_{it} + \beta_4 logGNI_{it} + \beta_1 tradeop_{it} + \beta_2 BIT_{it} + \beta_5 countrydummy_{it} + \beta_6 countrydummy_{p}op_{it} + \beta_7 Inflation_{it} + \beta_8 logexchange_{it} + \mu_{it}$$

**Model 8**

$$logOFDI_{it} = \beta_0 + \beta_3 logpop_{it} + \beta_4 logGNI_{it} + \beta_1 tradeop_{it} + \beta_2 BIT_{it} + \beta_5 countrydummy_{it} + \beta_6 countrydummy_{G}NI_{it} + \beta_7 Inflation_{it} + \beta_8 logexchange_{it} + \mu_{it}$$

**Model 9**

$$logOFDI_{it} = \beta_0 + \beta_3 logpop_{it} + \beta_4 logGNI_{it} + \beta_1 tradeop_{it} + \beta_2 BIT_{it} + \beta_5 countrydummy_{it} + \beta_6 countrydummy_{tradeo}p_{it} + \beta_7 Inflation_{it} + \beta_8 logexchange_{it} + \mu_{it}$$

**Model 10**

$$logOFDI_{it} = \beta_0 + \beta_3 logpop_{it} + \beta_4 logGNI_{it} + \beta_1 tradeop_{it} + \beta_2 BIT_{it} + \beta_5 countrydummy_{it} + \beta_6 countrydummy_{B}IT_{it} + \beta_7 Inflation_{it} + \beta_8 logexchange_{it} + \mu_{it}$$

**Model 11**

$$\begin{aligned} logOFDI_{it} = {} & \beta_0 + \beta_3 logpop_{it} + \beta_4 logGNI + \beta_1 tradeop_{it} + \beta_2 BIT_{it} \\ & + \beta_5 countrydummy_{it} + \beta_6 countrydummy_p op_{it} + \beta_7 countrydummy_G NI_{it} \\ & + \beta_8 countrydummy_{tradeo} p_{it} + \beta_9 countrydummy_B IT_{it} \\ & + \beta_{10} Inflation_{it} + \beta_{11} logexchange_{it} + \mu_{it} \end{aligned}$$

Model 1 to Model 5 are resource-seeking models while Model 6 to Model 11 are market-seeking models. In model 1, we have taken three resource-seeking factors- human resource development, natural resource endowment and infrastructure development along with country dummy taking a value 1 for African LDCs while a value 0 for Asian LDCs. Country dummy will help us understand and report the differences in FDI flows in African LDCs to that of Asian LDCs. Further, we take the interaction of country dummy with each of our resource-seeking variable- human resource development, natural resource endowment and infrastructure development in model 2, 3 and 4 respectively. Since there are structural differences between African and Asian region with respect to various factors such as resource endowments, infrastructure development, etc., it becomes worthwhile to interact the dummy with various independent variables to capture the specific impact on Asian and African LDCs. Finally, we take the interaction of country dummy with each of our resource-seeking factor simultaneously in model 5. In model 6, we take market-seeking factors- market size, income level of the country, trade openness and BIT along with our country dummy. The interaction between country dummy and each of our market seeking factor- market size, income level of the country, trade openness and BIT is taken in model 7,8,9 and 10 respectively. Finally, the interaction of country dummy with each of our market-seeking factors is taken simultaneously in model 11. In all the above models, we take inflation and exchange rate as our two control variables.

## 7. Results and discussion

Table 2 presents descriptive statistics for the underlying variables in our study, while also highlighting the differences between African LDCs and Asian LDCs. The statistics indicate that Asian LDCs on an average receive higher FDI from India in comparison to African LDCs. With respect to various market-seeking factors *viz.* market size (population), GNI per capita and trade openness, it is observed that the mean values of African LDCs and Asian LDCs do not differ significantly. However, they do report differences with respect to various resource-seeking factors *viz.* education of human resources, natural resource endowments and Infrastructure Development Index. As expected, African LDCs are endowed with greater natural resources in comparison to Asian LDCs. With respect to education of human resources as well infrastructural development, Asian LDCs have higher mean values in comparison to African LDCs.

The results for random effects Tobit model are reported in Table 3. The table presents the marginal effects for the impact of various resource-seeking and market-

Table 2. Descriptive statistics for variables.

| | Full sample | | | | African LDCs | | | | Asian LDCs | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variables | *Mean* | *Std Dev* | *Min* | *Max* | *Mean* | *Std Dev* | *Min* | *Max* | *Mean* | *Std Dev* | *Min* | *Max* |
| OFDI *(logOFDI)* | 0.3918 | 2.2957 | −4.3428 | 7.8818 | −0.3119 | 2.2867 | −4.3428 | 7.8818 | 1.5467 | 1.8095 | −2.057 | 5.2334 |
| Human Resource Education *(logeduc)* | 13.5918 | 1.3386 | 10.8608 | 16.5799 | 13.2913 | 0.7008 | 11.4846 | 14.5802 | 14.1929 | 1.9783 | 10.8608 | 16.5799 |
| Natural Resource Endowments *(lognatural)* | 1.7344 | 1.1891 | −0.9002 | 3.9959 | 2.3855 | 0.6458 | 1.1080 | 3.9959 | 0.2696 | 0.7390 | −0.9002 | 1.6400 |
| Infrastructure Development Index *(Infra)* | 3.50e − 08 | 1.0000 | −1.2972 | 1.7527 | −0.7142 | 0.5438 | −1.2972 | 0.6644 | 0.9862 | 0.5342 | −0.1345 | 1.7527 |
| Population *(logpop)* | 16.5332 | 1.2454 | 13.4174 | 18.8991 | 16.4679 | 0.7001 | 15.0082 | 17.8465 | 16.6801 | 1.9931 | 13.417 | 18.8991 |
| GNI per Capita *(logGNI)* | 6.6731 | 0.5237 | 5.8289 | 7.9963 | 6.5720 | 0.4564 | 5.8289 | 7.5229 | 6.9061 | 0.5957 | 6.0637 | 7.9963 |
| Trade Openness *(tradeop)* | 0.6274 | 0.2248 | 0.3223 | 1.3198 | 0.6352 | 0.2131 | 0.3223 | 1.3198 | 0.6097 | 0.2512 | 0.3530 | 1.1592 |

**Table 3.** Random effects tobit model regression results.

| Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Human Resource Education | 0.4884 | 0.4759 | 0.5785 | 0.4318 | 0.3825 | | | | | | |
| | (0.000)* | (0.000)* | (0.000)* | (0.001)* | (0.009)* | | | | | | |
| Natural Resources | 1.0082 | 1.0843 | 1.3360 | 0.9145 | 0.8823 | | | | | | |
| | (0.002)* | (0.000)* | (0.001)* | (0.010)* | (0.027)* | | | | | | |
| Infrastructure | −0.300 | −0.3388 | −0.2633 | −0.6473 | −0.7237 | | | | | | |
| | (0.322) | (0.248) | (0.375) | (0.082) | (0.050)* | | | | | | |
| Country Dummy | −3.003 | −13.654 | −1.5001 | −2.7641 | −15.382 | −0.7856 | −8.5918 | −2.4040 | 0.2649 | −0.6356 | −21.987 |
| (1= Africa, 0= Asia) | (0.000)* | (0.001)* | (0.223) | (0.001)* | (0.008)* | (0.050)* | (0.142) | (0.633) | (0.773) | (0.141) | (0.008)* |
| Country Dummy* Human Resource Education | | 0.7626 | | | 0.8722 | | | | | | |
| (Interaction Term) | | (0.011)* | | | (0.018)* | | | | | | |
| Country Dummy* Natural Resources | | | −0.7701 | | 0.3037 | | | | | | |
| (Interaction Term) | | | (0.130) | | (0.632) | | | | | | |
| Country Dummy* Infrastructure | | | | 0.8789 | 0.9617 | | | | | | |
| (Interaction Term) | | | | (0.109) | (0.082) | | | | | | |
| Market Size | | | | | | 0.5492 | 0.4951 | 0.5350 | 0.6630 | 0.5490 | 0.3822 |
| | | | | | | (0.000)* | (0.001)* | (0.001)* | (0.000)* | (0.000)* | (0.027)* |
| GNI Per Capita | | | | | | 0.4418 | 0.5023 | 0.3565 | 0.4320 | 0.3776 | 0.0357 |
| | | | | | | (0.146) | (0.087) | (0.354) | (0.137) | (0.228) | (0.915) |
| Trade Openness | | | | | | 2.2234 | 2.0080 | 2.2315 | 3.3746 | 2.4108 | 2.5402 |
| | | | | | | (0.001)* | (0.004)* | (0.001)* | (0.007)* | (0.001)* | (0.086) |
| BIT | | | | | | −0.1082 | −0.0819 | −0.0927 | −0.0678 | 0.0454 | 0.7300 |
| (Yes = 1, No = 0) | | | | | | (0.666) | (0.750) | (0.715) | (0.792) | (0.887) | (0.073) |
| Country Dummy* Market Size | | | | | | | 0.3284 | | | | 0.6403 |
| (Interaction Term) | | | | | | | (0.185) | | | | (0.016)* |
| Country Dummy* GNI Per Capita | | | | | | | | 0.1793 | | | 0.7904 |
| (Interaction Term) | | | | | | | | (0.719) | | | (0.070) |
| Country Dummy* Trade Openness | | | | | | | | | −1.5001 | | −0.6366 |
| (Interaction Term) | | | | | | | | | (0.266) | | (0.691) |
| Country Dummy* BIT | | | | | | | | | | −0.4402 | −1.0157 |
| (Interaction Term) | | | | | | | | | | (0.386) | (0.037)* |
| Inflation | −0.0289 | −0.0319 | −0.0292 | −0.0144 | −0.0167 | 0.0208 | 0.0229 | 0.0227 | 0.0198 | 0.0159 | 0.0276 |
| | (0.249) | (0.191) | (0.240) | (0.597) | (0.516) | (0.282) | (0.232) | (0.259) | (0.300) | (0.424) | (0.133) |
| Exchange Rate | 0.0316 | −0.1187 | −0.0882 | 0.0150 | −0.1155 | 0.0219 | −0.0319 | 0.0320 | −0.0160 | 0.0054 | −0.0996 |
| | ( | (0.336) | (0.519) | (0.899) | (0.387) | (0.851) | (0.795 | (0.788) | (0.893) | (0.964) | (0.408) |
| | (0.781) | | | | | | | | | | |
| Wald $Chi^2$ | 33.39 | 41.29 | 37.19 | 34.69 | 44.80 | 28.84 | 35.03 | 28.62 | 32.55 | 28.54 | 69.95 |
| Prob. > $Chi^2$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0002 | 0.0000 | 0.0004 | 0.0001 | 0.0004 | 0.0000 |
| Log Likelihood | −96.08 | −92.79 | −94.95 | −94.84 | −91.31 | −126.12 | −125.31 | −126.05 | −125.50 | −125.73 | −123.09 |

***Note:*** *The table reports marginal effects of random effects Tobit model. The dependent variable is OFDI from India in LDCs. OFDI, human resource education, natural resource endowment, GNI per capita and market size are measured in logs. The results are reported at 5% level of significance.*

seeking factors on OFDI from India in LDCs. It is observed that all the resource-seeking as well market seeking models are fit at 5% level of significance. Model 1 to model 5 are resource-seeking models while model 6 to model 11 are market-seeking models.

### 7.1. Results for resource-seeking models

In all the resource-seeking models, results indicate that endowment of natural resources in LDCs have a significant and positive relationship with OFDI from India. Hence, we fail to reject our *Hypothesis 1a*. The results were highly expected since it has been observed that over the years foreign investment in LDCs has majorly been natural resource-seeking. These results are also consistent with the statistics presented in World Investment Report (2019) which indicates that in 2018, the abundance of natural resources played a significant role in attracting foreign investments in LDCs, particularly in African LDCs. Existence of large pool of natural resources, particularly oil, make LDCs an attractive destination for investment by India, the country being the third largest importer of oil in the world.

Similarly, education of human resources in LDCs also has a significant and positive impact on OFDI from India in all our resource-seeking models. Thus, we fail to reject our *Hypothesis 1 b*. The results are consistent with past researchers Bhavan, Xu, and Zhong (2011) and Kahouli and Maktouf (2015), who have established positive relationship between the development of human capital in a country and its ability to attract more FDI. More than the cost or the quantity of human resources it is the quality of human resources that play a significant role in attracting FDI in the country. Skilled human resources help in raising the production efficiency since they are able to work well with advanced technology. Human resource development is particularly critical for LDCs since it is one of the criteria which defines the development in LDCs. Development of human resources has been the focal point of every ten-year *Programme of Action* for the growth and development of LDCs. Hence, the development of human resources in LDCs help in restoring investors' confidence, since it signals greater efforts and participation on part of LDCs towards the efficient development of their human resources.

Resource-seeking models indicate that infrastructure development in LDCs is insignificant in attracting FDI from India. Hence, we have sufficient evidence to reject our *Hypothesis 1c*. Until now, infrastructure development has not been the focal point in many LDCs, particularly African LDCs. Reports by United Nations indicate that road density and electricity coverage in LDCs has been low over the years (UNCTAD 2015; UNCTAD 2017). The communication and transportation networks are also not up to the required levels in LDCs. As a result, infrastructure in LDCs does not seem to play a significant role in attracting FDI.

The country dummy variable which takes the value 1 for African LDCs and 0 for Asian LDCs is significantly negative in all the resource-seeking models except for model 3. This indicates that OFDI from India is lower in African LDCs in comparison to Asian LDCs. It is in line with the UNCTAD statistics which indicate that FDI flows in African economies declined by 31% in 2017 while for Asian economies, the

flows reported an upward trend (UNCTAD 2018a). Also, FDI flows in 2018 touched a record high in Asian LDCs (UNCTAD 2019b). Although trade and investments by India in African LDCs has been rising over the years, it is majorly concentrated in oil and gas sector.

The results for interaction terms between country dummy and each of the resource variable are indicated in model 2,3,4 and 5. Of all the interaction terms, only the interaction between country dummy and education of human resources is positively significant in models 2 and 5. This indicates that as the human resources develop, the increase in OFDI by India will be more in African LDCs in comparison to Asian LDCs.

### 7.2. Results for market-seeking models

Model 6 to model 11 report the marginal effects of market-seeking factors on OFDI from India in LDCs. Consistently in all market-seeking models, trade openness has a significant and positive relationship with OFDI. Thus, we have sufficient evidence to not reject our *Hypothesis 2c*. This indicates that as LDCs open their trade with the partners abroad, the FDI in LDCs also increases. The results are consistent with past researchers (Asiedu 2006; Aziz and Mishra 2016; Campos and Kinoshita 2003; Tintin 2013), who have established a positive relationship between trade openness and FDI in a country. As the country opens its trade, the integration between the economies increases, thereby helping the foreign investors in improving their understanding of the host country local market conditions. However, BIT between LDCs and India is not significant in attracting FDI from India in all the market-seeking models. As a result, our *Hypothesis 2d* stands rejected. Although, over the years LDCs are taking greater interest in signing treaties with partners abroad, it is important to understand that mere signing the BIT is not enough in enhancing the relationship between economies; it is imperative that these treaties must be in force well in time. Statistics by United Nations indicate that there is usually a time lag between the signing of BIT and when it actually comes into force (United Nations 2020). This might be the reason that despite the active participation by LDCs in signing BITs, these treaties are not playing a significant role in attracting FDI from India.

Market size also has a significant and positive relationship with OFDI in all the market-seeking models. Thus, we fail to reject our *Hypothesis 2a*. The results are in line with past researchers Milner and Pentecost (1996), Nunnenkamp (2002) and Rjoub *et al.* (2017), who have established positive relationship between market size of an economy and its ability to attract more FDI. The rising population size of the host economy provides foreign investors with the growing market for their products. This entices the investors to set up their production facilities in such economies. This is particularly significant for LDCs since they represent greater potential for market growth. Markets in developed nations are already saturated and are characterized with ageing population. Thus, the untapped opportunities in underdeveloped economies like LDCs makes them an attractive destination for investment. With respect to GNI per capita, although the marginal effects for GNI per capita are positive in all

the models, however, they are not significant in attracting investment to LDCs in all the market-seeking models. Hence, our *Hypothesis 2 b* stands rejected.

Consistently in all market-seeking models, again a negative coefficient of country dummy indicate that OFDI by India in African LDCs is lower than in Asian LDCs. Also, model 7,8,9,10 and 11 include interaction of country dummy with each of the market-seeking variable. Of all the interaction terms, the interaction between country dummy and market size, and country dummy and BIT is significant. Interaction between country dummy and market size indicates that as the size of the market increases, OFDI by India is more in African LDCs in comparison to Asian LDCs. Also, the interaction between country dummy and BIT indicates that as the BITs between India and LDCs increases, the OFDI by India is higher in Asian LDCs in comparison to African LDCs. This may be owing to the fact that it is comparatively easy for India to strengthen its bilateral relations with its neighbouring countries due to similarity in their market dynamics to some extent.

In conclusion, the results indicate that OFDI from India into LDCs is both resource-seeking as well market-seeking in nature. It is majorly due to the endowment of rich resources and untapped potential in LDCs, which makes them an attractive destination for investment by India. Our overall results suggest that currently, OFDI from India is more in Asian LDCs in comparison to African LDCs. FDI inflows in Asian LDCs are mostly driven by strong investment in real estate, telecommunication and hydropower projects. Over the years, Asian LDCs have focused on developing their economies by undertaking several infrastructural development projects as well as by investing in diversified sectors. However, our results also indicate that as various market-seeking factors (market size and GNI per capita) and resource-seeking factors (quality of human resources and infrastructure development) develop/improve, OFDI from India will be more in African LDCs in comparison to Asian LDCs. Despite great potential, currently African LDCs are not able to attract much FDI from India mainly due to their large scale underdevelopment. It is only in terms of natural resources that African LDCs are ahead of Asian LDCs, while with respect to quality of human resources and infrastructural development, Asian LDCs have taken the lead. Hence, African LDCs must focus on developing their economies in order to provide conducive investment environment.

## 8. Conclusion and policy implications

The current study explores the impact of various host country (LDCs) resource-seeking as well market-seeking factors that affect the outward foreign direct investment from India. Over the years, India has been undertaking several initiatives to boost its trade and investment with LDCs. For instance, the Government of India implemented the Duty-Free Tariff Preference (DFTP) scheme in 2008 to grant duty free market access to all LDCs. The government has also undertaken several strategic initiatives in order to strengthen its relationship with African countries such as 'Focus Africa' programme, Asia-Africa Growth Corridor and India Africa Forum Summit. Notwithstanding these policy measures, India does not feature amongst the top ten investor economies in LDCs (UNCTAD 2019b; UNCTAD 2020). Hence, the current

study attempts to determine what causes OFDI flows from India into LDCs. For this purpose, the study employs a panel data of 13 LDCs from 2008 to 2018 and uses random effects Tobit model. The results indicate that OFDI from India into LDCs is both resource-seeking as well market-seeking in nature. With respect to geographical distribution, results indicate that OFDI by India is lower in African LDCs in comparison to Asian LDCs.

The study presents policy implications at three levels. Firstly, what can be done by India to boost its outward FDI into LDCs? It is evident that LDCs are emerging as an attractive destination of investment for Indian MNCs. Hence, we suggest a focused policy framework such as the 'Go Global or Go Out' policy implemented by Chinese Government in 1999, which can encourage and support Indian MNEs to fully exploit the untapped potential and take the benefit of 'L' advantages (OLI framework) in LDCs. There are several economies that provide financial assistance to its players for investing in developing and transition economies. Such as United States' OPIC (now DFC) encourages its private players to invest in developing and transition economies. Hence, our study suggests that such policy measures shall be undertaken by Government of India (GoI) in order to encourage its players to invest in underdeveloped economies like LDCs. Since the idea of investment in LDCs is perhaps in its nascent phase, the information and knowledge gap with respect to the underlying investment opportunities in LDCs exist. The investment opportunities in LDCs must be communicated well with the Indian MNEs. As a result, GoI shall provide information services to its MNCs such as providing data on business opportunities in LDCs, economic and political investment climate and information on the benefits of internationalization. Additionally, various financial measures shall be extended to Indian MNEs such as concessional loans, public-private risk sharing and grants to carry out feasibility studies and other pre-investment activities in LDCs in order to assess the cost-benefit analysis of setting up their offices in LDCs.

Secondly, our study presents implications for what can LDCs do in order to boost their inward FDI. The results indicate that OFDI from India into LDCs is both market-seeking as well as resource-seeking. The resource-rich nature of LDCs as well as their untapped potential, makes them an attractive destination of investment. Hence, LDCs shall leverage upon their 'L' advantages of the OLI framework to pull investment from Indian MNCs. Although the quality of human resources plays a significant role in attracting FDI from India, the skill development of human resources in LDCs is not up to the required levels. It is thus imperative for LDCs to moot the training and development of human resources under the ambit of their strategic goals. This is particularly important for African LDCs since the results indicate that as the human resources develop, the increase in OFDI by India will be more in African LDCs in comparison to Asian LDCs. Also, LDCs represent humungous growth potential in their infrastructural facilities. Hence, it is important for LDCs to focus more on building its infrastructure facilities to attract FDI from India. Further, our results indicate that African LDCs present greater FDI opportunities in coming years, since the interaction effect indicates that as the market size, GNI per capita, quality of human resources and infrastructure development improve, OFDI from India will be more in African LDCs in comparison to Asian LDCs. It is mainly due to their large

scale underdevelopment that currently African LDCs are not able to attract much FDI from India. Hence, African LDCs must focus on strengthening their communication networks, transportation networks, road density and electricity coverage in order to provide conducive investment environment. The focus shall be on developing sectors other than natural resources like power, telecommunication, real-estate, etc.

Further, there are certain measures that can be undertaken jointly by India and LDCs in order to boost their bilateral ties. Since, trade openness emerges as a significant variable in boosting OFDI from India into LDCs, it suggests that both the economies shall focus on enhancing their trade relationship. India has already implemented Duty-Free Tariff Preference (DFTP) scheme to grant duty free market access to all LDCs. In return, LDCs can also implement such policies to provide greater market access to Indian firms. Finally, it is imperative for both the economies to understand that mere signing of BIT is not sufficient, appropriate enforcement of BITs between India and LDCs is also important in order to strengthen the bilateral ties between the nations and make investment environment more conducive.

Finally, there are certain limitations in our study which provides the scope for future researchers. First, our sample is limited to 13 LDCs due to unavailability of the data. In future, researchers shall undertake a bigger sample and perform region-wise comparison among LDCs. Second, it would be worthwhile for future researchers to undertake industry analysis and examine which particular industries are attracting major investment in LDCs. Third, researchers can also examine if the distance between India and LDCs plays a major role in explaining OFDI from India into LDCs by employing the gravity model of international trade. Fourth, in the light of various policy initiatives implemented by the GoI towards the development of LDCs, it becomes interesting to explore the motives of investment by Indian MNCs in LDCs. Are these investments community driven or profit driven? Further, are these investments driven by state-owned enterprises or private enterprises?

## Acknowledgements

The manuscript was presented at the 7th Copenhagen Conference on Outward Investment from Emerging Economies organised by Copenhagen Business School on 14–15 October, 2021 and has been revised on the basis of the reviews received.

A version of the paper was adjudged the best paper at International Conference on 'Resilience for Sustainability: Revisiting Management Practices and Strategizing for the Future' organized by School of Management and Liberal Studies at the NorthCap University (NCU), Gurugram, India, held on March 26, 2021

We are thankful to the reviewers for their valuable and constructive comments on the manuscript.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Note

1. There are 27 Low-income countries and 46 Least Developed Countries. As per the World Bank classification, all low-income countries are also classified as least developed countries except Democratic People's Republic of Korea and Syrian Arab Republic.

## Notes on contributors

*Apoorva Jain*, Research Scholar, is pursuing her Doctorate of Philosophy (Ph.D.) from Delhi School of Management, Delhi Technological University (DTU), having completed her Masters in Commerce (M. Com) from Department of Commerce, University of Delhi and Bachelors in Commerce [B. Com (Hons.)] from Shri Ram College of Commerce, University of Delhi. Her area of interest lies in International Business and Family Business Management.

*Dr. Sonal Thukral* is an Assistant Professor at Delhi School of Management, Delhi Technological University (DTU). She was earlier employed with Shri Ram College of Commerce, University of Delhi as Assistant Professor for over seven years. Prior to academic experience, she worked with S&P Capital IQ and Tecnova India. An alumna of Shri Ram College of Commerce and Hindu College, University of Delhi, she is the recipient of various best research paper awards at conferences (including IIM Indore). She has published various papers in national and international journals. Dr. Thukral has an avid interest in pursuing multidisciplinary studies. She has completed her doctoral studies in the area of finance and international business.

## ORCID

Sonal Thukral http://orcid.org/0000-0003-0926-5438

## References

Adhikari, R. (2019), "Targeting Aid for Trade for Impactful Capacity-Building in the Least Developed Countries." *Global Policy* 10 (3): 408–412. doi:10.1111/1758-5899.12705.

Aleksynska, M., and O. Havrylchyk. 2013. "FDI from the South: The Role of Institutional Distance and Natural Resources." *European Journal of Political Economy* 29: 38–53. doi:10.1016/j.ejpoleco.2012.09.001.

Anh, N. T., and L. H. Ngoc. 2016. "Introduction to the Theoretical Framework of Dunning's Investment Development Path." *VNU Journal of Science: Education Research* 32 (5): 1–11.

Asiedu, E. 2006. "Foreign Direct Investment in Africa: The Role of Natural Resources, Market Size, Government Policy, Institutions and Political Instability." *The World Economy* 29 (1): 63–77. doi:10.1111/j.1467-9701.2006.00758.x.

Aziz, O. G., and A. V. Mishra. 2016. "Determinants of FDI Inflows to Arab Economies." *The Journal of International Trade & Economic Development* 25 (3): 325–356. doi:10.1080/09638199.2015.1057610.

Barnard, H., A. Cuervo-Cazurra, and S. Manning. 2017. "Africa Business Research as a Laboratory for Theory-Building: Extreme Conditions, New Phenomenon, and Alternative Paradigms of Social Relationships." *Management and Organization Review* 13 (3): 467–495. doi:10.1017/mor.2017.34.

Basten, S., and J. C. Cuaresma. 2014. "Modelling the Macroeconomic Impact of Future Trajectories of Educational Development in Least Developed Countries." *International Journal of Educational Development* 36: 44–50. doi:10.1016/j.ijedudev.2013.12.003.

Belkhodja, O., M. Mohiuddin, and E. Karuranga. 2017. "The Determinants of FDI Location Choice in China: A Discrete-Choice Analysis." *Applied Economics* 49 (13): 1241–1254. doi:10.1080/00036846.2016.1153786.

Beri, R. 2015. "Third India Africa Forum Summit: Key Outcomes," December. Manohar Parikar Institute for Defense Studies and Analyses. https://idsa.in/africatrends/third-india-africa-forum-summit_rberi_1215

Bhavan, T., C. Xu, and C. Zhong. 2011. "The Relationship between Foreign Aid and FDI in South Asian Economies." *International Journal of Economics and Finance* 3: 143–149.

Blomström, Magnus, Robert E. Lipsey, and Magnus Blomstrom. 1991. "Firm Size and Foreign Operations of Multinationals." *The Scandinavian Journal of Economics* 93 (1): 101–107. doi: 10.2307/3440424.

Cameron, A. C., and, and P. K. Trivedi. 2005. *Microeconometrics, Methods and Applications*. 1st ed. New York: Cambridge University Press.

Campos, N. F., and Y. Kinoshita. 2003. "Why does FDI Go Where it Goes? New Evidence from the Transition Economies," IMF Working Papers 03/228, International Monetary Fund. www.imf.org/external/pubs/ft/wp/2003/wp03228.pdf.

Cardamone, P., and M. Scoppola. 2012. "The Impact of EU Preferential Trade Agreements on Foreign Direct Investment." *The World Economy* 35 (11): 1473–1501. doi:10.1111/j.1467-9701.2012.01434.x.

Choi, C. 2003. "Does the Internet Stimulate Inward Foreign Direct Investment?" *Journal of Policy Modeling* 25 (4): 319–326. doi:10.1016/S0161-8938(02)00202-8.

Cuervo-Cazurra, A., and M. Genc. 2008. "Transforming Disadvantages into Advantages: developing-Country MNEs in the Least Developed Countries." *Journal of International Business Studies* 39 (6): 957–979. doi:10.1057/palgrave.jibs.8400390.

Deichmann, J. I. 2001. "Distribution of Foreign Direct Investment among Transition Economies in Central and Eastern Europe." *Post-Soviet Geography and Economics* 42 (2): 142–152. doi:10.1080/10889388.2001.10641167.

Driffield, N., and K. Taylor. 2000. "FDI and the Labour Market: A Review of the Evidence and Policy Implications." *Oxford Review of Economic Policy* 16 (3): 90–103. doi:10.1093/oxrep/16.3.90.

Dunning, J. H. 1982. "Explaining the International Direct Investment Position of Countries: Towards a Dynamic or Developmental Approach." In *International Capital Movements. International Economics Study Group*, edited by J. Black, and J. H. Dunning, 30–46. London: Palgrave Macmillan.

Dunning, J. H., and R. Narula. 1994. *Transpacific Foreign Direct Investment and the Investment Development Path: The Record Assessed*. Center for International Business Education and Research, Division of Research, College of Business Administration, University of South Carolina.

Elfakhani, S. M., and L. M. Matar. 2007. "Foreign Direct Investment in the Middle East and North Africa Region." *J. for Global Business Advancement* 1 (1): 49–70. doi:10.1504/JGBA.2007.012549.

Ezeoha, A. E., and N. Cattaneo. 2012. "FDI Flows to Sub-Saharan Africa: The Impact of Finance, Institutions, and Natural Resource Endowment." *Comparative Economic Studies* 54 (3): 597–632. doi:10.1057/ces.2012.18.

Hakro, A. N., and A. M. Omezzine. 2011. "Governance Infrastructure and FDI Flows in MENA Region Countries." *Journal of Middle Eastern Finance and Economics* 12: 144–158.

International Trade Centre. 2015. *India's Duty Free Tariff Preference Scheme for LDCs: A Business Guide*. Geneva: International Trade Centre.

Jain, S. C., and S. Vachani. 2006. "The Role of MNCs in Alleviating Global Poverty." In *Multinational Corporations and Global Poverty Reduction*. Cheltenham: Edward Elgar Publishing.

Janicki, H. P., and P. V. Wunnava. 2004. "Determinants of Foreign Direct Investment: Empirical Evidence from EU Accession Candidates." *Applied Economics* 36 (5): 505–509. doi:10.1080/00036840410001682214.

Jarvis, P. 2006. *Towards a Comprehensive Theory of Human Learning*. New York: Routledge.

Jorgenson, A. K., K. Austin, and C. Dick. 2009. "Ecologically Unequal Exchange and the Resource Consumption/Environmental Degradation Paradox: A Panel Study of Less-

Developed Countries, 1970–2000." *International Journal of Comparative Sociology* 50 (3-4): 263–284. doi:10.1177/0020715209105142.

Kahouli, B., and S. Maktouf. 2015. "The Determinants of FDI and the Impact of the Economic Crisis on the Implementation of RTAs: A Static and Dynamic Gravity Model." *International Business Review* 24 (3): 518–529. doi:10.1016/j.ibusrev.2014.10.009.

Kumari, R., and A. K. Sharma. 2017. "Determinants of Foreign Direct Investment in Developing Countries: A Panel Data Study." *International Journal of Emerging Markets* 12 (4): 658–682. doi:10.1108/IJoEM-10-2014-0169.

Laabas, B., and W. Abdmoulah. 2009. "Determinants of Arab Intraregional Foreign Direct Investments." *Journal of Business & Policy Research* 4: 138–169.

Lewin, A. Y., S. Massini, and C. Peeters. 2009. "Why Are Companies Offshoring Innovation? The Emerging Global Race for Talent." *Journal of International Business Studies* 40 (6): 901–925. doi:10.1057/jibs.2008.92.

Lim, D. 1983. "Fiscal Incentives and Direct Foreign Investment in Less Developed Countries." *The Journal of Development Studies* 19 (2): 207–212. doi:10.1080/00220388308421859.

Loree, D. W., and S. E. Guisinger. 1995. "Policy and Non-Policy Determinants of U.S. Equity Foreign Direct Investment." *Journal of International Business Studies* 26 (2): 281–299. doi:10.1057/palgrave.jibs.8490174.

Maddala, G. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Malgwi, C. A., V. Owhoso, K. C. Gleason, and I. Mathur. 2006. "Who Invests in the Least Developed Countries?" *Journal of African Business* 7 (1–2): 201–227. No doi:10.1300/J156v07n01_10.

Medvedev, D. 2012. "Beyond Trade: The Impact of Preferential Trade Agreements on Foreign Direct Investment Inflows." *World Development* 40 (1): 49–61. doi:10.1016/j.worlddev.2011.04.036.

Milner, C., and E. Pentecost. 1996. "Locational Advantage and US Foreign Direct Investment in UK Manufacturing." *Applied Economics* 28 (5): 605–615. doi:10.1080/00036849600000040.

Ministry of External Affairs. 2011. *Note on India and Least Development Countries (LDCs)*. February 6. Ministry of External Affairs, Government of India: https://mea.gov.in/bilateral-documents.htm?dtl/4389/Note+on+India+and+Least+Development+Countries+LDCs#:~:text=Dependence%20of%20LDCs%20on%20global,natural%20resources%20for%20economic%20growth.&text=India%20has%20excellent%20bilateral%20relations,longsta

Mohiuddin, M., and Z. Su. 2013. "Manufacturing Small and Medium Size Enterprise's Offshore Outsourcing and Competitive Advantage: An Exploratory Study on Canadian Offshoring Manufacturing SMEs." *Journal of Applied Business Research (JABR)* 29 (4): 1111–1130. doi:10.19030/jabr.v29i4.7920.

Montanari, M. G., J. D. M. Giraldi, and S. V. R. Galina. 2019. "Relationship between Country Brand and Internationalization: A Literature Review." *Benchmarking: An International Journal* 27 (7): 2148–2165. doi:10.1108/BIJ-09-2018-0277.

Musteen, M., L. Rhyne, and C. Zheng. 2013. "Asset or Constraint: Corporate Reputation and MNCs' Involvement in the Least Developed Countries." *Journal of World Business* 48 (3): 321–328. doi:http://dx.doi.org/10.1016/j.jwb.2012.07.016.

Na, L., and W. S. Lightfoot. 2006. "Determinants of Foreign Direct Investment at the Regional Level in China." *Journal of Technology Management in China* 1 (3): 262–278. doi:10.1108/17468770610704930.

Nunnenkamp, P. 2002. "Determinants of FDI in Developing Countries: Has Globalization Changed the Rules of Game?." *Kiel Working Paper No. 1122*, Keil Institute of World Economics, http://www.ifw-members.ifw-kiel.de/publications/de … the-game/kap1122.pdf.

Poelhekke, S., and F. Van der Ploeg. 2010. "Do Natural Resources Attract FDI? Evidence from Non-Stationary Sector-Level Data." *CEPR Discussion Paper 8079*. Centre for Economic Policy Research.

Pradhan, J. P. 2017. "Indian Outward FDI: A Review of Recent Developments." *Transnational Corporations* 24 (2): 43–70. doi:10.18356/32864c71-en.
Prahalad, C. K. 2004. *The Fortune at the Bottom of the Pyramid: Eradicating Poverty through Profits*. Philadelphia: Wharton School Publishing.
Prahalad, C. K., and A. Hammond. 2002. "Serving the World's Poor, Profitably." *Harvard Business Review* 80 (9): 4–11.
Prahalad, C. K., and K. Lieberthal. 1998. "The End of Corporate Imperialism." *Harvard Business Review* 76 (4): 68–78.
Research and Information System for Developing Countries. 2017. *Asia Africa Growth Corridor: Partnership for Sustainable and Innovative Development*. Ahemdabad: Research and Information System for Developing Countries.
Rjoub, H., M. Aga, A. A. Alrub, and M. Bein. 2017. "Financial Reforms and Determinants of FDI: Evidence from Landlocked Countries in Sub-Saharan Africa." *Economies* 5 (1): 1–12. doi:10.3390/economies5010001.
Root, F. R., and A. A. Ahmed. 1978. "The Influence of Policy Instruments on Manufacturing Direct Foreign Investment in Developing Countries." *Journal of International Business Studies* 9 (3): 81–93. doi:10.1057/palgrave.jibs.8490670.
Sauvant, K. P. 2015. "Attracting Foreign Direct Investment and Benefiting from It: Challenges for the Least Developed Countries." *Transnational Corporations Review* 7 (2): 125–127. doi:10.5148/tncr.2015.7201.
Tessema, M. T., B. P. Winrow, and M. M. Teclezion. 2012. "The Transfer of Training at Macro Level in Least Developed Countries: A Case Study of the 'Brain-Drain' in Eritrea." *International Journal of Training and Development* 16 (4): 247–262. doi:10.1111/j.1468-2419.2012.00412.x.
The Economic Times. 2002. "Maran Launch's 'Focus: Africa' Programme." *The Economic Times*, March 31. https://economictimes.indiatimes.com/news/economy/policy/maran-launches-focus-africa-programme/articleshow/5455788.cms?from=mdr.
The Economic Times. 2020. WTO: India pushes for building broadband infra, digital skills in LDCs, developing countries. *The Economic Times*, May 19. https://economictimes.indiatimes.com/news/economy/policy/wto-india-pushes-for-building-broadband-infra-digital-skills-in-ldcs-developing-countries/articleshow/75828611.cms
Tintin, C. 2013. "Foreign Direct Investment Inflows and Economic Freedoms: evidence from Central and Eastern European Countries." *Advances in Business-Related Scientific Research* 4 (1): 1–12.
Tripathi, V., and S. Thukral. 2016. "Financing the Internationalisation of Indian MNEs." *Transnational Corporations Review* 8 (3): 215–229. doi:10.1080/19186444.2016.1233732.
Tripathi, V., and S. Thukral. 2018a. "Determinants of Financing of Outward Foreign Direct Investment by Indian MNEs: A Three-Level Analysis." *International Journal of Emerging Markets* 13 (5): 1154–1181. doi:10.1108/IJoEM-12-2016-0333.
Tripathi, V., and S. Thukral. 2018b. "Impact of Characteristics of Emerging Market's Firm on Financing Their Outward Foreign Direct Investment." *International Journal of Public Sector Performance Management* 4 (2): 190–213. doi:10.1504/IJPSPM.2018.090737.
Turnbull, C., S. Sun, and S. Anwar. 2016. "Trade Liberalisation, Inward FDI and Productivity within Australia's Manufacturing Sector." *Economic Analysis and Policy* 50: 41–51. doi:10.1016/j.eap.2016.02.004.
Ubah, C. N. 1980. "Western Education in Africa: The Igbo Experience, 1900–1960." *Comparative Education Review* 24 (3): 371–388. doi:10.1086/446154.
UNCTAD. 2001. "29 Bilateral Investment Treaties Signed by Least Developed Countries in Brussels," United Nations Conference on Trade and Development, May 18. https://unctad.org/en/pages/PressReleaseArchive.aspx?ReferenceDocId=2914
UNCTAD. 2011. *Foreign Direct Investments in LDCs: Lessons Learned from the Decade 2001-2010 and the Way Forward*. New York and Geneva: United Nations Publication.
UNCTAD. 2015. *The Least Developed Countries Report 2015: Transforming Rural Economies*. Geneva: United Nations Publication.

UNCTAD. 2017. *The Least Developed Countries Report 2017: Transformational Energy Access*. Geneva: United Nations Publication.
UNCTAD. 2018a. *World Investment Report 2018: Investment and New Industrial Policies*. New York and Geneva: United Nations Publication.
UNCTAD. 2018b. *The Least Developed Countries Report 2018: Entreprenuership for Structural Transformation: Beyond Business as Usual*. New York and Geneva: United Nations Publication.
UNCTAD. 2019a. *The Least Developed Countries Report 2019: The Present and Future of External Development Finance- Old Dependence, New Challenges*. Geneva: United Nations Publication.
UNCTAD. 2019b. *World Investment Report 2019: Special Economic Zones*. Geneva: United Nations Publication.
UNCTAD. 2020. *World Investment Report 2020: International Production beyond Pandemic*. Geneva: United Nations Publication.
UNCTAD. 2022. *Foreign Direct Investment in LDCs: Investment Trends and Policies since LDC VI and the Way Forward*. Geneva: United Nations Publication.
United Nations. 2012. "LDC Portal: International Support Measures for Least Developed Countries," United Nations, October 2. https://www.un.org/ldcportal/india-imports-from-ldcs/
United Nations. 2020. "International Investment Agreements Navigator," Investment Policy Hub-United Nations https://investmentpolicy.unctad.org/international-investment-agreements/countries/96/india
United Nations Department of Economic and Social Affairs. 2018. *Handbook on the Least Developed Country Category: Inclusion, Graduation and Special Support Measures*. 3rd ed. United Nations Publications. https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/2018CDPhandbook.pdf
United Nations Secretariat. 2015. *Strengthening Investment Promotion Regimes for Foreign Direct Investment in the Least Developed Countries*. New York: UNOHRLLS, United Nations Secretariat.
UNOHRLLS. 2006. *Brussels Programme of Action: Addressing the Needs of the Least Developed Countries*. New York: United Nations Publication.
UNOHRLLS. 2020a. "Istanbul Declaration and Programme of Action – 2011," UN Offuce of the High Repressentative for the Least Developed Countries, Landlocked Developing Countries and Small Island Developing States (UN-OHRLLS) http://unohrlls.org/about-ldcs/istanbul-programme-of-action/
UNOHRLLS. 2020b. "About LDCs," UN Office of the High Representative for the Least Developed Countries, Landlocked Developing Countries and Small Islands Developing States (UN-OHRLLS). https://unohrlls.org/about-ldcs/
Wadhwa, K., and S. S. Reddy. 2001. "Foreign Direct Investment into Developing Asian Countries: The Role of Market Seeking, Resource Seeking and Efficiency Seeking Factors." *International Journal of Business and Management* 6 (11): 219–226. doi:10.5539/ijbm.v6n11p219.
Wheeler, David, and Ashoka Mody. 1992. "International Investment Location Decisions." *Journal of International Economics* 33 (1–2): 57–76. doi:10.1016/0022-1996(92)90050-T.
Winters, L. A., N. McCulloch, and A. McKay. 2004. "Trade Liberalization and Poverty: The Evidence so Far." *Journal of Economic Literature* 42 (1): 72–115. doi:10.1257/002205104773558056.

Hindawi
Complexity
Volume 2022, Article ID 4815623, 20 pages
https://doi.org/10.1155/2022/4815623

*Research Article*

# Wind Effects on Rectangular and Triaxial Symmetrical Tall Building Having Equal Area and Height

**Astha Verma,[1] Rahul Kumar Meena,[2] Hrishikesh Dubey,[2] Ritu Raj,[2] and S. Anbukumar[2]**

[1]*Department of Civil Engineering, Govind Ballabh Pant University of Agriculture and Technology, Pantnagar, India*
[2]*Department of Civil Engineering, Delhi Technological University, New Delhi, India*

Correspondence should be addressed to Ritu Raj; rituraj@dtu.ac.in

Received 25 February 2022; Revised 11 May 2022; Accepted 23 May 2022; Published 20 July 2022

Academic Editor: Mostafa Al-Emran

Copyright © 2022 Astha Verma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The study aims to investigate wind effects on the equal area and the same height building model having the plan cross-sectional shape in the form of regular and irregular shape bulidings. The ratio of modification in the cross-sectional shape is kept same for regular and irregular shaped models because limited studies are available for such type of comparison. The responses of the structure to resist the wind load is increased by applying these modifications in the corner configuration such as corner cut, chamfer, and fillet in the plan shape of the structure. The numerical study is performed using ANSYS CFX and wind incidence angle varies in the range of 0° to 180° at the interval of 15°. The results of velocity and pressure distribution are presented in the form of pressure contours and wind force coefficients for all four-building models which are presented in the form of a graph. The comparison of numerical simulation results on two models are compared with the different international standards and with the experimental results. The regular shape with a corner cut and a "Y"-shaped structure with a fillet corner model performed the best among all models for resisting overall wind force. The results of pressure distribution of all four models are presented in the form of pressure contours for 0° and 90° wind angles.

## 1. Introduction

The tall structure is the most appropriate solution for the current situation since land availability in the conventional shape is approaching at saturation. The population is also booming and there is a scarcity of space available for individual residences consequently, accommodating such a large population within the available land is feasible with tall buildings. Various international standards [1–10] are already discussed different parameters to design tall buildings. There is significant constraint on the design of tall buildings as the majority of international standards are silent on the modification of the shape of plan cross-sectional area buildings. Tavakoli et al. [11] discussed the important aspects of high rise buildings. Dadkhah et al. [12] investigated the performance of the steel structure during an earthquake. Kamgar and Rahgozar [13] analyzed different lateral resisting systems to resist lateral loads which are produced during an earthquake. Kamgar et al. [14] investigated the soil-structure interaction for a different storey outrigger braced building resting on two different types of soil. Dadkhah et al. [15] performed nonliner seismic analyses using incremental dynamic analysis, which is most accurate and time-consuming, and also need a high computational cost. Dadkhah et al. [16] presented the optimization of a tuned mass damper to improve the seismic performance of a six storey steel structure based on the ductility damage index. Kamgar and Samani [17] conducted the numerical study for evaluating the effect of length to the height ratio on the behavior of a concrete frame retrofitted with steel infill plates and observed that the value of absorbed energy by two diagonal viscous dampers is higher than that of the chevron viscous damper. Numerous studies have been conducted in the subject of wind engineering, the majority of which focus on modifying the shape or height of the structure but the current study focuses on equal area structures of regular and

irregular shapes. When the regular shape of the rectangular building model is contrasted to the irregular shape of the "Y"-shaped building model, the designer can choose the shape of the building when constructing such shapes.

Investigation of wind effects is possible through two techniques like wind tunnel and CFD, the various amount of research is already conducted through the wind tunnel, such as Kwok [18] studied the effect of wind on the shape of a building by applying different measures to control the wind effect and found that the building with the chamfer corner is best among all to resist the along and across wind response. Jozwiak et al. [19] presented the aerodynamic interference effect on pressure distribution on the building model and observed that there is a significant difference in pressure distribution with and without modeling of the boundary layer. Pal et al. [20] presented the experimental investigation performed in an open circuit boundary layer wind tunnel on the "fish plan shape" and observed that pressure on the windward face in the case of a fish shape is less than the square shape when the wind incidence angle is 0°. Nagar et al. [21] performed an experiment in BLWT on the "H" shaped model and found that pressure increases until the wind incidence angle reach up to 60° wind, and it was found that the interference factor for the square and the "H" shaped building is less than one. Pal et al. [22] carried out a comparative study of wind-induced mutual interference effects on the twin square and the fish plan shaped model, which is having equal volume, and the experiment is performed in the boundary layer wind tunnel at the length scale of 1 : 300. This is observed from the experimental study that the drag and lift force for the fish shaped model is lesser than that of the square shaped model in the case of 0° and 180° wind. Nagar et al. [23] investigated the wind-induced interference effect between two plus shaped high rise buildings on mean and RMS pressure coefficients and concluded that these interference effects on local wind pressure are significantly higher on the windward side face near the recessed corner. Pal et al. [24] performed an experimental investigation on a square and a remodeled triangle shaped building at the length scale of 1 : 300. This is found that the front-to-front interference case of the remodeled triangle shaped model performed best in resisting along and across the wind. The maximum suction of the testing model at the FBI condition is decreased by nearly 40% to that of the isolated case of the SPS model at the same condition. Amin and Ahuja [25] presented a review on an aerodynamic modification such as corner cut, chamfering, rounding, horizontal slots, vertical slots, dropping of corners, and many more to the shape of the tall building. The major outcome from this review is that the opening along the cross-wind direction at the top significantly reduces the wind excitation of the building, and the tapering of tall buildings has significant effects on wind force in across wind direction than that of the along wind response. Blackmore [26] estimated the wind effects on and around the building model, wind forces are dynamic and fluctuate continuously in both magnitude and position. Most of the international standards considered static loads and having limitations. Bhattacharyya and Dalui [27] presented the study on the "*E*" plan shaped building, which has symmetry about both the axis and wind response is measured from 0° to 330° at the interval of 30°. The result obtained from the experimental study and numerical study were compared. Various amounts of research work is performed using the wind tunnel test, however, this study is performed by using the $k$-$\varepsilon$ turbulence model and most of the studies in the present times are performed using the $k$-$\varepsilon$ turbulence model to obtained the solution using the numerical simulation.

Numerical simulation is the better option for the estimation of wind loads because the wind tunnel test is time taking and costly, and it also needs heavy machinery and resources to investigate the wind effects. Various studies are also available and presented on the various aspect of CFD simulation, such as Tominaga and Stathopoulos [28] investigated the turbulent scalar flux by assuming the gradient diffusion hypothesis using Reynold averaged Navier stock equation in CFD. Blocken [29] discussed the important aspect of CFD for defining urban physics and gave ten important tips and tricks towards the accurate and reliable simulations. Behera et al. [30] performed an experiment in the boundary layer wind tunnel to investigate the wind interference effect on tall buildings with varying plan ratios, and observed that the interference zone extends over a larger area as the building plan ratio increases. Tamura et al. [31] performed a series of wind tunnel tests to determine the aerodynamic performance and pedestrian level wind characteristics of different super tall buildings having square, rectangular, and elliptic plan with corner cuts. Mittal et al. [32] performed the numerical simulation on tall buildings and discussed the effect of the shorter building placed in front of a taller building. Blocken et al. [33] conducted the CFD simulation of the atmospheric boundary layer and discussed the effect of using wall function. The accuracy of CFD simulation can be seriously compromised when wall function roughness modification based on the experimental data for sand grain roughened pipes and channels are applied at the bottom of the computational domain. Numerical studies are already conducted on the different types of the plan shape structure, such as Tian et al. [34] on a rectangular shape, Gaur and Raj [35] on corner modification on a square and a plus shaped building, Raj et al. [36], Keerthana and Harikrishna [37]on the "*H*" plan, Kumar and Raj [38] on an octagonal plan shape, Raj et al. [39] obtained response of a square and a plus shaped building by varying wind loads, Gaur et al. [40] studied the interference effect using CFD, Mukherjee et al. [41], Sanyal and Dalui [42–45], Goyal et al. [46] on the "Y" shape, Meena et al. [47], Kumar and Raj [48] on the "L" shape, Raj and Ahuja [49], on the "+" shape, Sanyal and Dalui [50] on the rectangular shape having some modification, Meena et al. [51] investigated the wind effects on the regular shaped structure having a rectangular plan with various corner modifications, Mallick et al. [52] on the "*C*" shape, Amin and Ahuja [53] on the side ratio of a rectangular building, Amin and Ahuja [54] obtained response of a tall building against the wind load. The considerable outcomes after modifying the shapes of the structure are as the Strouhal number is not sensitive to the aspect ratio, flow reattachment of the separated shear layer in the corner cut model present suction into corners, which is further

caused in the reduction of drag, interference position of the building is highly effecting principal building pressure distribution, the length/width (side ratio) makes changes into the upwash, downwash, and the stagnation zone on upstream of flow, shielding effects were the main factors which is controlling the spacing between the buildings, highest wind velocity is obtained at the edges of the windward side and minimum at the leeward side, maximum suction is observed at the corner of the building, it is important to give due consideration for suction on the corner region, by increasing the $L/W$ ratio horizontal force coefficient and overturning moment coefficient increases by the significant amount, the model with the rounded corner is best to reduce wind loads and overturning moments, in comparison of the chamfer corner, rounded corner are efficient in reducing the wind load, deflection is observed due to the presence of vortices in the wake region of buildings, the accuracy of the $k$-$\varepsilon$ turbulence model is more than the SST model.

The current study will examine the wind pressure on tall structures of both regular and irregular shapes. The majority of earlier research has concentrated on the wind effects of either regular or irregular shaped buildings, however, the current study examines the results of these two types of buildings by altering the corners. This study is particularly noteworthy because the cross-sectional area and height of the building model are kept constant to minimize wind loads. Numerous corner modifications, such as chamfer, fillet, and corner cut, are also employed and the ratio of cuts is maintained, allowing the designer to select the most efficient structure for the situation. Additionally, two models are validated and the numerical simulation results are compared to experimental results and a number of relevant standards. Furthermore, the result of velocity streamlines is illustrated graphically in the form of contours, pressure distribution, and wind force coefficient for wind incidence angles ranging from 0° to 180° at 15°-degree intervals.

## 2. Methodology

The present study concentrates on the investigation of wind effects and compares the results of the regular and irregular shaped models. The numerical simulation results are validated before the starting of the work and the results of numerical simulation are found identical with experimental results.

*2.1. Method of Study.* The present study is performed using the numerical simulation tool ANSYS CFX and investigated the wind effects on tall buildings having an equal plan area and same height. Most of the available studies are either for regular shape or for irregular shape, but these studies present the comparison among such shapes having both regular and irregular shape. The geometry is selected and applied in the CFX-Pre setup. After the geometry, the meshing is performed and flow physics are defined in the solution and simulation is performed and the results are presented in different graphical forms.

*2.2. Numerical Simulation.* Numerical simulation is performed using ANSYS CFX and boundary conditions used into this simulation are kept as the boundary condition used by Raj [55] in the experiment performed in the boundary layer wind tunnel at IIT Roorkee, India. The numerical simulation performed into this study is performed using the k-$\varepsilon$ turbulence model [56].

(i) $k$-$\varepsilon$ turbulence model

This method is generally used to solve a complex fluid problem and it is a two-equation model. The solution is generated after solving two different equations during the entire numerical simulation. This model performs nearly equivalent to the experimental problem and the k-$\varepsilon$ model also uses the scalable wall-function to increase the efficiency of the solver and this model performs more robustness in the case of the fine mesh.

It is very less expensive and mostly used to simulate turbulent flow characteristics. The solution is solved by two transport equations, i.e., turbulent kinetic energy ($k$) and the turbulent dissipation rate ($\varepsilon$). It has the positive advantage of not including any geometry-related parameters in the modeling. Turbulent kinetic energy and the turbulence dissipation rate are two variables introduced into the system of equations for the model. The values for turbulence are calculated according to turbulence eddy dissipation using

(ii) Basic equations

The basic equation used to study fluid flow problems using Navier-Stokes and continuity equation.

(iii) Navier stokes equation is written as

$$\frac{\partial(\rho u_i)}{\partial t} = -\frac{\partial(\rho u_i u_j)}{\partial x_j} - \frac{\partial P}{\partial x_j} + \frac{\partial}{\partial_{x_j}}\left[\mu\left(\frac{\partial_{u_i}}{\partial_{x_j}} + \frac{\partial_{u_j}}{\partial_{x_i}}\right)\right] + F. \tag{1}$$

(iv) Continuity equation is written as

$$\frac{\partial_\rho}{\partial_t} + \frac{\partial_{\rho_i}}{\partial_{x_i}} = 0. \tag{2}$$

The standard $k$-$\varepsilon$ model uses the following transport equations for the turbulence kinetic energy and turbulence dissipation.

$$\rho\frac{\partial k}{\partial t} + \rho\overline{u_i}\frac{\partial k}{\partial x_j} = \tau_{ij}\frac{\partial \overline{u_i}}{\partial x_j} - \rho\varepsilon + \frac{\partial}{\partial_{x_j}}\left[\left(\mu + \frac{\mu_t}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial x_j}\right]. \tag{3}$$

Momentum equation is written as

$$\frac{\partial(\rho U_i)}{\partial t} = -\frac{\partial(\rho U_i U_j)}{\partial x_j} - \frac{\partial P}{\partial x_i} + \frac{\partial}{\partial_{x_j}}\left[\mu_{eff}\left(\frac{\partial_{U_i}}{\partial_{x_j}} + \frac{\partial_{U_j}}{\partial_{x_i}}\right)\right] + S_M. \tag{4}$$

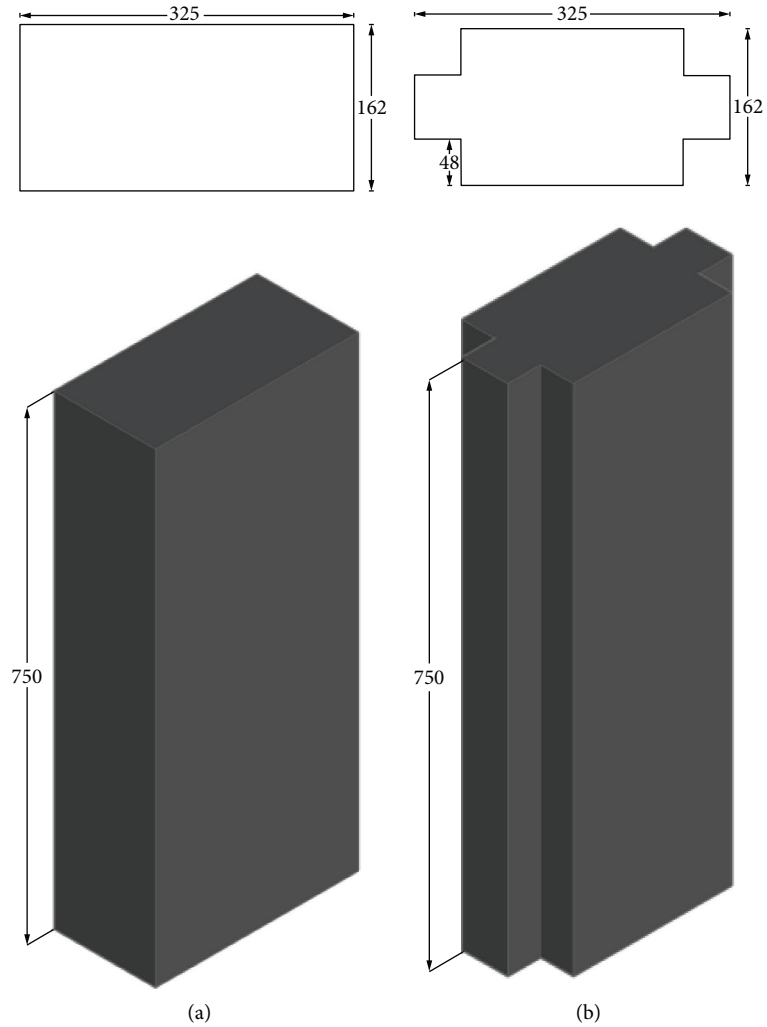



Figure 1: Regular shaped building model (a) rectangle and (b) rectangle with corner cuts.

*2.2.1. Advantages*

(1) It performs best for the boundary layer separation fluid problem.

(2) It is the most use turbulence model for validation purposes.

(3) Generally, it is used by the industrialist to solve the complex fluid problem.

(4) It provided the solution based on boundary conditions.

*2.2.2. Disadvantage*

(1) It does not perform well in the case where flow is changing instantaneously.

(2) It is unable to generate the accurate solution for the rotating fluid problem.

(3) It is notable to generate the accurate solution for a low Reynold number.

*2.3. Geometrical Configuration of Model.* The regular shape of a rectangular building Figure 1 represents the building having no corner modification, corner cut model and irregular shaped model of Y-shape is represented in Figure 2 are considered in this study. The plan cross-sectional area and height are kept the same for the study so that the results can be compared on an equal volume building model. Both sorts of building models have the same modification ratio.

*2.4. Meshing and Domain.* Meshing contains important flow features, which are dependent upon the flow parameters, such as grid refinement inside the wall boundary layer. The meshing applied after the geometry and the name selection for each part of geometry are required before meshing to understand the CFX-Pre about the geometry. The name selection is required to define flow physics. It is better to provide name selection before the meshing so that the surface mesh exactly matches with nodes on the two sides of the boundary, which allows a more accurate fluid solution

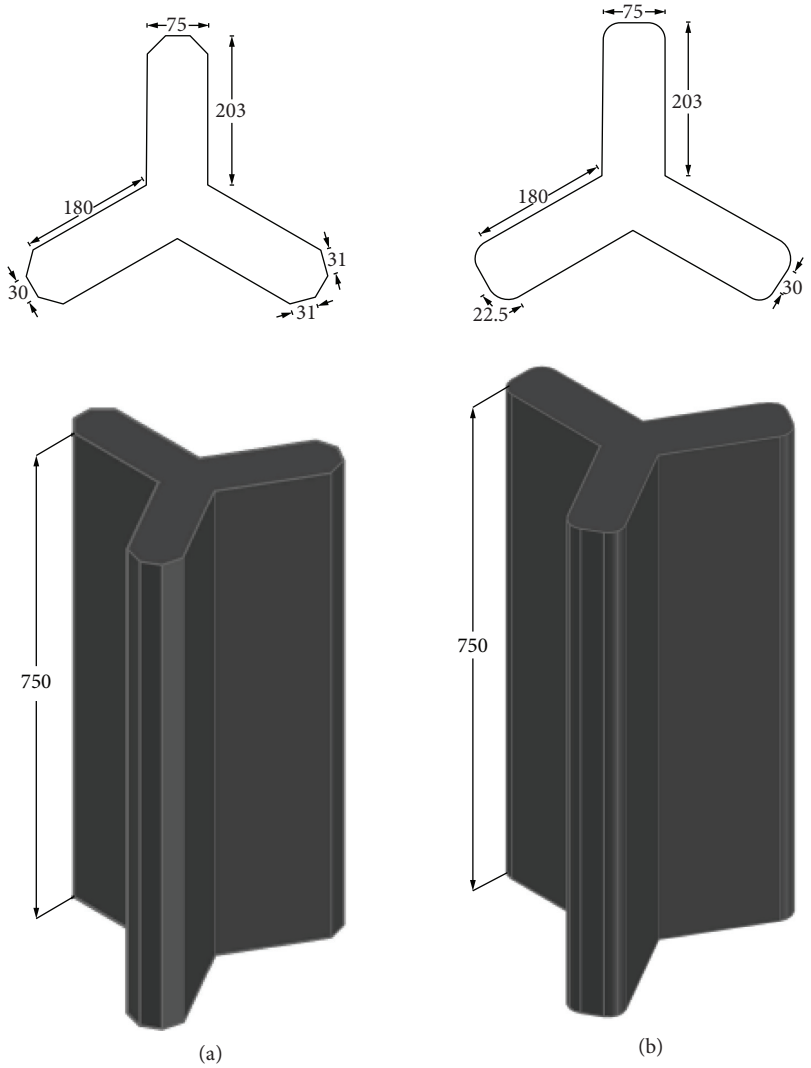



Figure 2: Y-shape building model (a) Y-shape with chamfer and (b) Y-shape with fillet.

[57]. Name selection also helps the program to control the inflation, i.e., automatically select for the wall and inflation automatically provided during the auto mesh generation. The meshing is represented in Figure 3 for domain, building, and inflation.

The mesh generation steps are automatically worked into the program; however, this can be controlled by varying the element size, type of mesh to generate, and where and how the mesh should be refined. Tetra dominant meshing, which is patch independent, is suitable for the CAD model having many surface patches and if the geometry has small edges then this is suitable [58]. Hex meshing is used both in general sweep and thin sweep; it is also recommended for the model which has the clean cad geometry. Mapped and free meshing is adopted where the fluid problem needs a different type of meshing like structured (mapped) and unstructured (free); however, this is suitable for the problem where the sweeping method does not work without extensive geometry decomposition.

Inflation is provided to capture the flow properly at the interface and the same can be provided by various methods

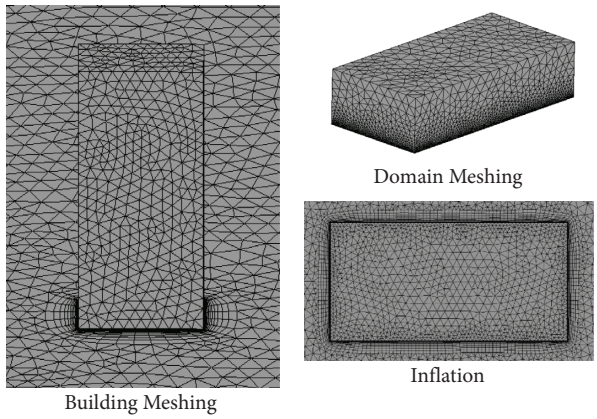



Figure 3: CFD model representing meshing.

available in the CFD model. Smooth transition is the default option and it uses the local tetrahedron elements size to compute each local height and total height so that the rate of volume change is smooth. Each triangle that is being inflated

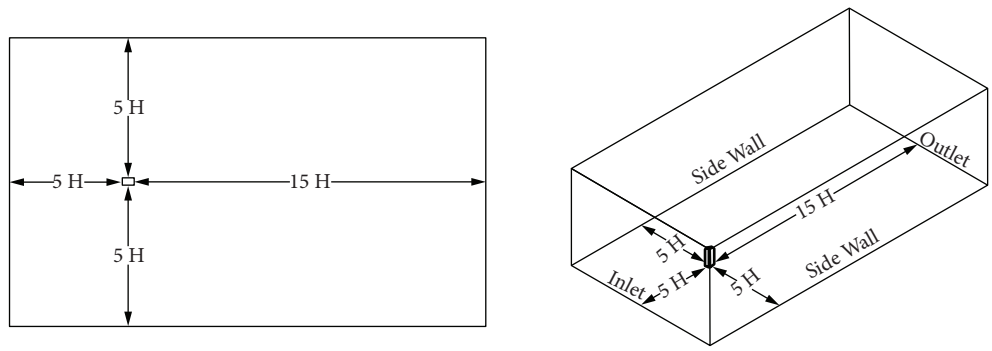



Figure 4: CFD domain.

will have an initial height that is computed with respect to its area averaged at the nodes. This means that for the uniform mesh, the initial height will be roughly same, while for a varying mesh, the initial height will vary. An increment in the value of the growth rate that controls the reduction in the total height of inflation layers.

The numerical simulation predicts the result based on boundary conditions, as the finite element method work well with boundary conditions. Inlet wind speed is provided as the power law where the reference height in the simulation is considered as 1 m, while the reference velocity is defined as 10 m/s. The wall of the domain is considered as a free slip wall, while the ground of the domain is a rough wall [43].

The faces of the model are considered as no slip. Wall function defines that if a coarse mesh near the wall than it assumes that the logarithmic low applies as wall function. At the same time, the fine mesh near the wall and various turbulence models account for a low Reynold number. Wall functions are useable to low Reynold number problems. In general, if the fine mesh is defined near the wall then the gradient changes are considered automatically by the solver. The domain is represented in Figure 4 in both the top and isometric view. A domain feature is considered after the various recondition provided by the past studied [59–62] numerical model. No-slip walls is the wall where fluid immediately next to the wall assumes the velocity of the wall as zero; free slip wall is the wall where the shear stress at the wall is zero and the velocity of the fluid near the wall is not retarded by the wall friction effects.

*2.5. Grid Sensitivity.* In this study a grid sensitivity study is performed on model-A and tabulated in Table 1, which is having no corner modification into the rectangular shape model. A grid sensitivity study is essential to CFD programming as it helps to achieve the proper meshing size, which predicts the most accurate result for the numerical simulation. For the present study GS-3 is adopted for all the wind incidence angle.

*2.6. Mean Wind Speed and Turbulence Intensity Profile*

*2.6.1. Mean Wind Speed.* Wind flow over the tall building involves complex flow patterns, although wind flow is generally separated from the surface, where the flow is mostly recirculating. Flow patterns are unsteady around the buildings model, therefore creating turbulence. The development of turbulences near the ground level is observed due to the presence of other obstacles. Such effects are also considered while calculating the turbulence intensity and wind flow up to the gradient height and after that it becomes constant.

Mean wind speed defined in the CFX-Pre is represented in Figure 5, the variation in the form of a power law between height and mean wind speed is as follows:

$$\frac{U(Z)}{U\left(Z_{ref}\right)} = \left(\frac{Z}{Z_{ref}}\right)^{n}, \tag{5}$$

where $U(Z)$ = mean velocity at height $Z$; $U(Z_{\text{ref}})$ = mean velocity at reference height $Z_{\text{ref}}$; and $n$ = power law exponent, a measure of ground roughness, varies between 0.13 and 0.15 in open terrain.

The variation in the mean wind speed with height is generally expressed in an alternate form in a logarithmic form.

$$\frac{U(Z)}{U_{*}} = \left(\frac{1}{k}\right) ln\left(\frac{Z}{Z_{o}}\right). \tag{6}$$

Also, a simplification of this logarithmic equation is as follows:

$$\frac{U(Z)}{U_{(Z_{\text{ref}})}} = \frac{\ln\left(Z/Z_{o}\right)}{\ln\left(Z_{\text{ref}}/Z_{o}\right)}, \tag{7}$$

where $U_{*}$ = shear velocity; $k$ = Von Karman constant, 0.4; ln = natural log function; and $Z_{o}$ = effective roughness length, another measure of ground roughness, varies from 0.01 to 0.05 meters in open terrain.

*2.6.2. Turbulent Intensity.* Turbulence intensity is a non-dimensional quantity derived from the variance and for the mean wind speed.

$$I_{z} = \frac{\sigma_{z}}{U(z)}, \tag{8}$$

where $I_{z}$ = turbulence intensity at height $z$; $\sigma_{z}$ = standard deviation of the wind speed at height $z$; and $U(z)$ = mean wind speed at a reference height.

Table 1: Mesh independent test result for model-A.

| Name | Type of meshing | No of elements | Mean external pressure Face | | | | % error | | | | Reynold No |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | *A* | *B* | *C* | *D* | *A* (%) | *B* (%) | *C* (%) | *D* (%) | |
| GS-1 | Coarse | 957324 | 0.59 | −0.49 | −0.24 | −0.49 | 36 | 32 | 26 | 32 | |
| GS-2 | Medium-1 | 1284687 | 0.73 | −0.56 | −0.27 | −0.56 | 10 | 16 | 12 | 16 | |
| GS-3 | Medium-2 | **1439589** | **0.78** | **−0.64** | **−0.30** | **−0.64** | **3** | **2** | **0** | **2** | $3.56\times10^6$ to $3.90\times10^6$ |
| GS-4 | Medium-3 | 1561140 | 0.81 | −0.69 | −0.32 | −0.69 | 1 | 5 | 6 | 5 | |
| GS-5 | Fine | 2497236 | 1.02 | −0.98 | −0.43 | −0.98 | 33 | 33 | 30 | 33 | |

Bold values are indicating that if the GS-3 case is adopted for the numerical simulation, then the percentage error will be less in comparison to other meshing cases investigated in this study.

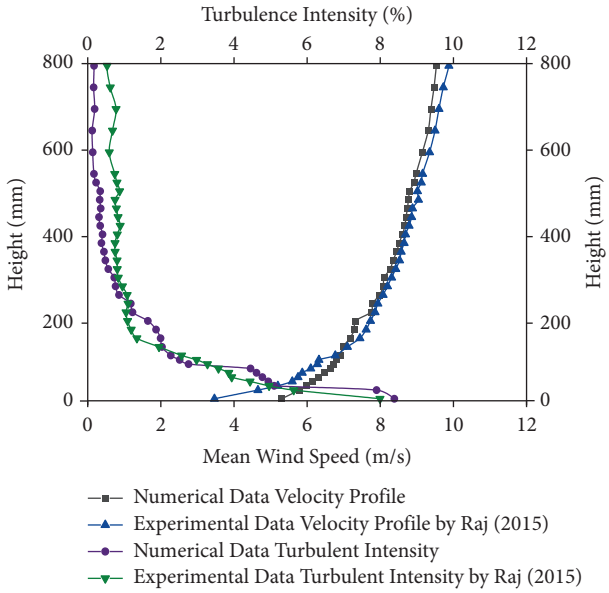



Figure 5: Variation of mean wind speed and turbulent intensity with height.

*2.7. Validation.* Validation is a prerequisite part of numerical simulation and two different models are considered in this study to validate with available international standards and an experimental study. Such buildings model are selected and external $C_p$ compared with the available experimental results. This is also clearly depicted in Figure 6 that the result obtained in this numerical simulation show nearly an identical result for the pressure coefficient.

The results plotted for the normal and corner cut rectangular building model are representing a very close match with the experimental results. Therefore, the result obtained in this study are very much accurate for designing the model, which are considered in this entire study. The rectangular shape model is also simulated for the wind load at 0° and 90° wind and a pressure coefficient is compared with available international standards mentioned at Table 2.

The external pressure coefficient "$C_p$" is calculated using (9).

$$C_p = \frac{p - p_o}{(1/2)\rho U_H^2}, \tag{9}$$

where $p$ = pressure derived from the external lines plotted on the model; $p_o$ = static pressure at reference height; $\rho$ = air density (1.225 kg/m$^3$); $U_H$ = mean wind velocity at the building reference height.

## 3. Result and Discussion

CFD postprocessed results enables the visualization and analysis of wind effects in variety of graphical representations. The pressure distribution along the face is a function of the shape and size and the same is illustrated using contours. The wind response is also represented graphically in the form of the wind force coefficient.

*3.1. Velocity Contours.* Velocity contours are represented for model-A in Figures 7 and 8. The main cause of the pressure difference is the velocity field around and along with the building model. The velocity at downstream is mentioned at Figure 7, and the size of the wake is increasing with respect to the frontal exposed area, which changes according to wind incidence angle.

Velocity along the height of model-A is depicted in Figure 8 and the wind velocity distribution pattern is illustrating the variation of the wind at the downstream. The suction observed in the case of 15° and 75° are of the same type. The suction on the roof of the building is also observed and demonstrated in Figure 8.

*3.2. Wind Forces.* The wind-generated effect in the form of wind force is evaluated after the numerical simulation performed on the building model. The value of $F_x$ and $F_y$ are obtained using CFD postprocessed result while the frontal exposed area is calculated at each wind incidence angle. Drag force is calculated as per equations (16) and (17). Along wind force, i.e., the drag for model-A, is obtained a maximum of −4.72 is between 60° and 120°, while the minimum 0.0 is noted when the wind is striking to the model at 0° and 90°. The corner cut model of rectangular shape, i.e., model-B, has the highest magnitude of drag is −4.39 in the case of 30° and 150° wind and the smallest drag is 0.04 in the case of 0° and 180° wind. The maximum amount of drag is reduced for model-B with respect to model-A.

$$F_x = \left(0.5\rho_{U_h^2}.A_p\right)C_{f_x},$$
$$F_y = \left(0.5\rho_{U_h^2}.A_p\right)C_{f_y}. \tag{10}$$

An irregular "Y" shaped model having corner modifications such as model-C and D are having chamfer and fillet, respectively, the drag is measured using numerical

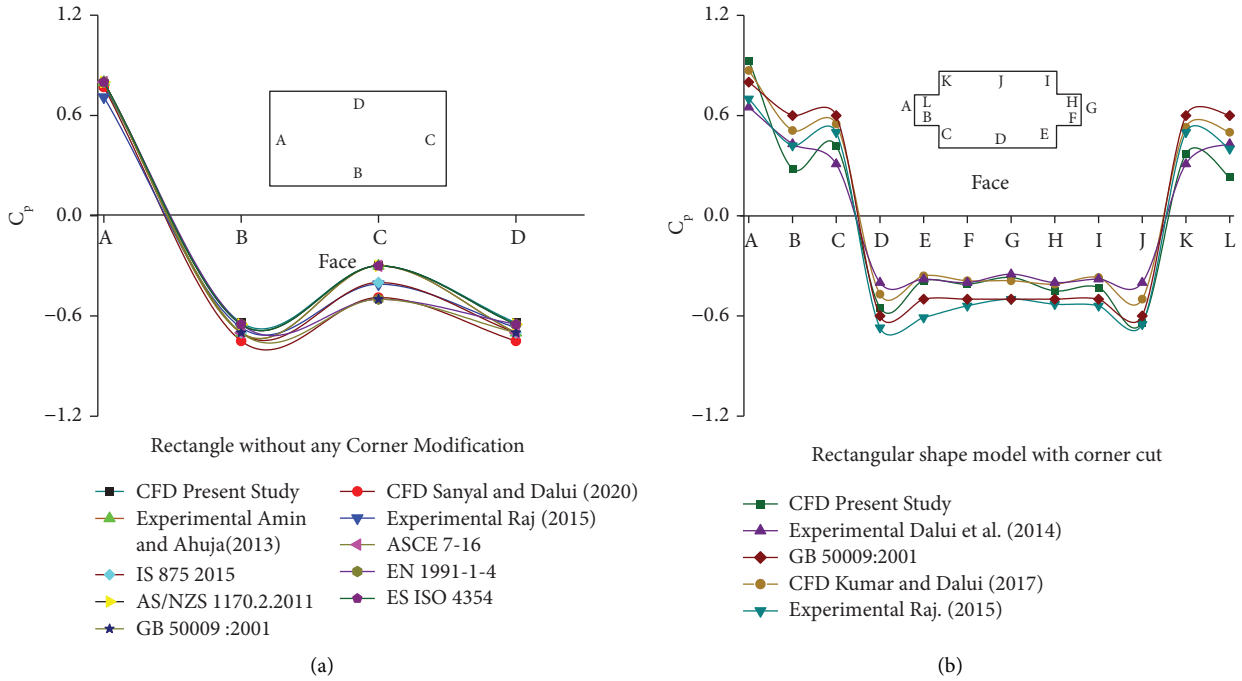



Figure 6: Validation result: (a) rectangular shape and (b) corner cut model.

Table 2: Comparison of average face pressure coefficient (*Cp*) on the rectangular shape tall building.

| International code | Wind angle (°) | Windward side | Side wall | Leeward side |
|---|---|---|---|---|
| CFD results | 0 | 0.78 | −0.64 | −0.30 |
| | 90 | 0.70 | −0.61 | −0.45 |
| IS 875 (part 3) | 0 | 0.7 | −0.7 | −0.4 |
| | 90 | 0.8 | −0.5 | −0.1 |
| ASCE/SEI 7-16 | 0 | 0.8 | −0.7 | −0.5 |
| | 90 | 0.8 | −0.7 | −0.5 |
| AS/NZS 1170.2.2011 | 0 | 0.8 | −0.65 | −0.5 |
| | 90 | 0.8 | −0.65 | −0.5 |
| EN 1991-1-4 | 0 | 0.8 | −0.5 | −0.7 |
| | 90 | 0.8 | −0.5 | −0.7 |
| BS 6399-2 | 0 | 0.8 | −0.5 | −0.7 |
| | 90 | 0.8 | −0.5 | −0.7 |
| GB 50009-2001 | 0 | 0.8 | −0.5 | −0.7 |
| | 90 | 0.8 | −0.5 | −0.7 |
| NSCP 2015 | 0 | 0.8 | −0.5 | −0.7 |
| | 90 | 0.8 | −0.5 | −0.7 |
| ES/ISO 4354: 2012 | 0 | 0.8 | −0.65 | −0.7 |
| | 90 | 0.8 | −0.65 | −0.7 |

simulation and graphically plotted in Figure 9. The drag is also measured for model-C (chamfer-Y shape) and the maximum magnitude is −9.44 in the case of 150° wind. Model-D (fillet–Y shape) is investigated for along wind forces and the maximum drag of −7.92 is noticed in the case of 150° wind, while the lowest drag is −0.04 in the case of 180° wind is noticed for model-D.

The across wind, i.e., lift force in the *y*-direction, is calculated and graphically plotted in Figure 10 for model-A, B, C, and D. The regular shaped model of the rectangular shape (model-A & B) is having some different type of corner modification so that lift force can be reduced on the tall building by making some modification in the corners. The maximum lift of −13.73 is noted in the case of 90° wind for model-A, while the smallest of −5.23 is spotted in the case of 0° and 180° wind. Model-B (corner cut rectangle) is having the highest lift of −11.28 in the case of 90° wind, while the lowest lift of −4.73 is detected in the case of 15° and 165° wind.

The irregular "Y" shaped model-C and *D* are numerically investigated for the across wind response and lift force for model-D (fillet-Y) is having the maximum lift force of

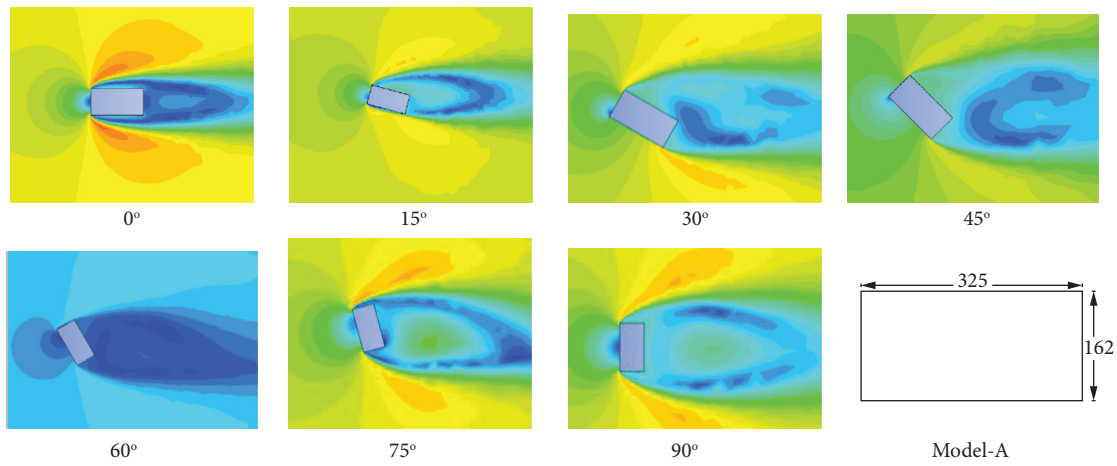



Figure 7: Velocity contours building model-A in plan.

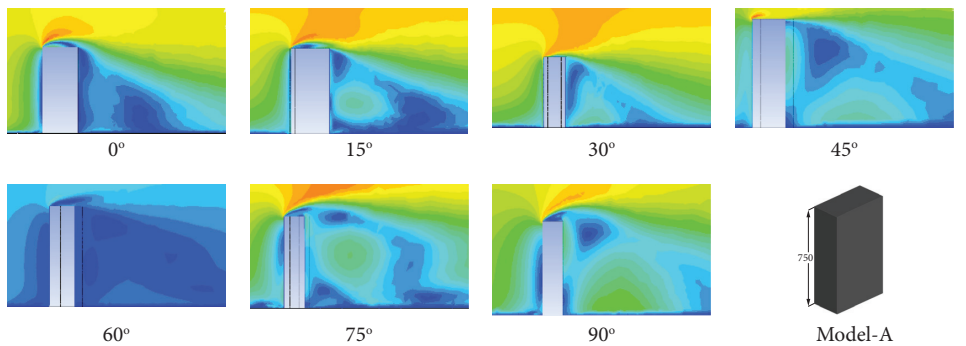



Figure 8: Velocity contours for model-A in elevation.

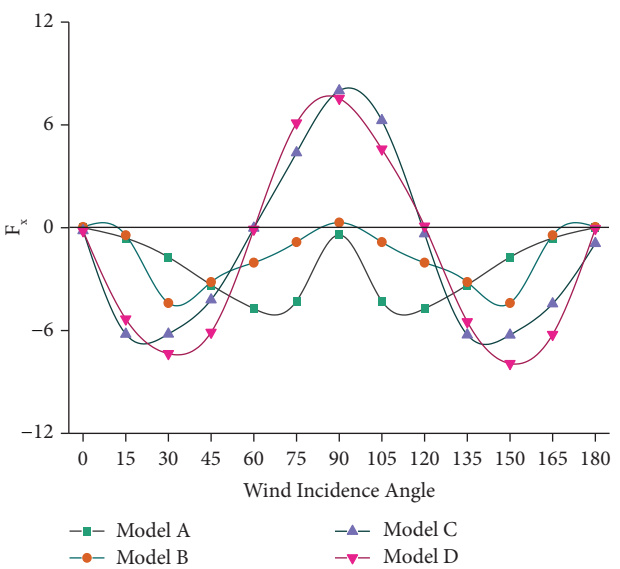



Figure 9: Wind force coefficient in $x$-direction $F_x$.

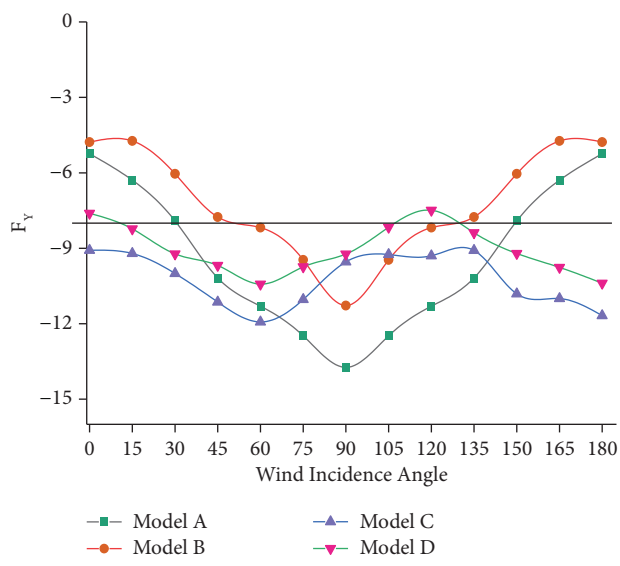



Figure 10: Wind force coefficient in $y$-direction $F_y$.

−11.92 in 60° wind, while the minimum of −9.08 for 0° wind 135° wind. The along wind, lift force is found greatest of −10.42 in the case of model-D (fillet-Y) for 60° wind, while the smallest of −7.6 is noticed in the case of 0° wind. Among the Y-shaped model, the lowest lift force is observed for the model having a fillet corner into each limb of the "Y" shape [63].

*3.3. Pressure Contours.* Pressure distribution is represented in the form of contours for each surface of the building model. Pressure distribution for rectangular building model-A is represented in Figure 11 for 0° wind, while for 90° wind, the pressure distribution pattern along the height of the

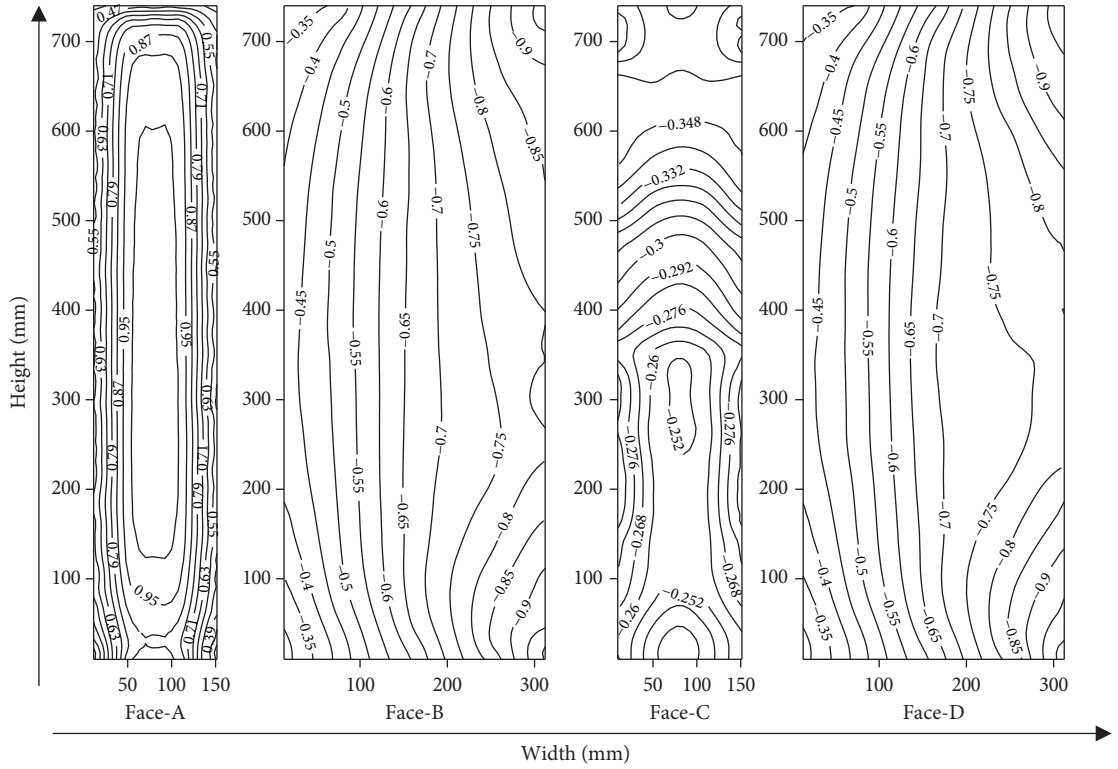



Figure 11: Pressure contours at 0° wind for model-A.

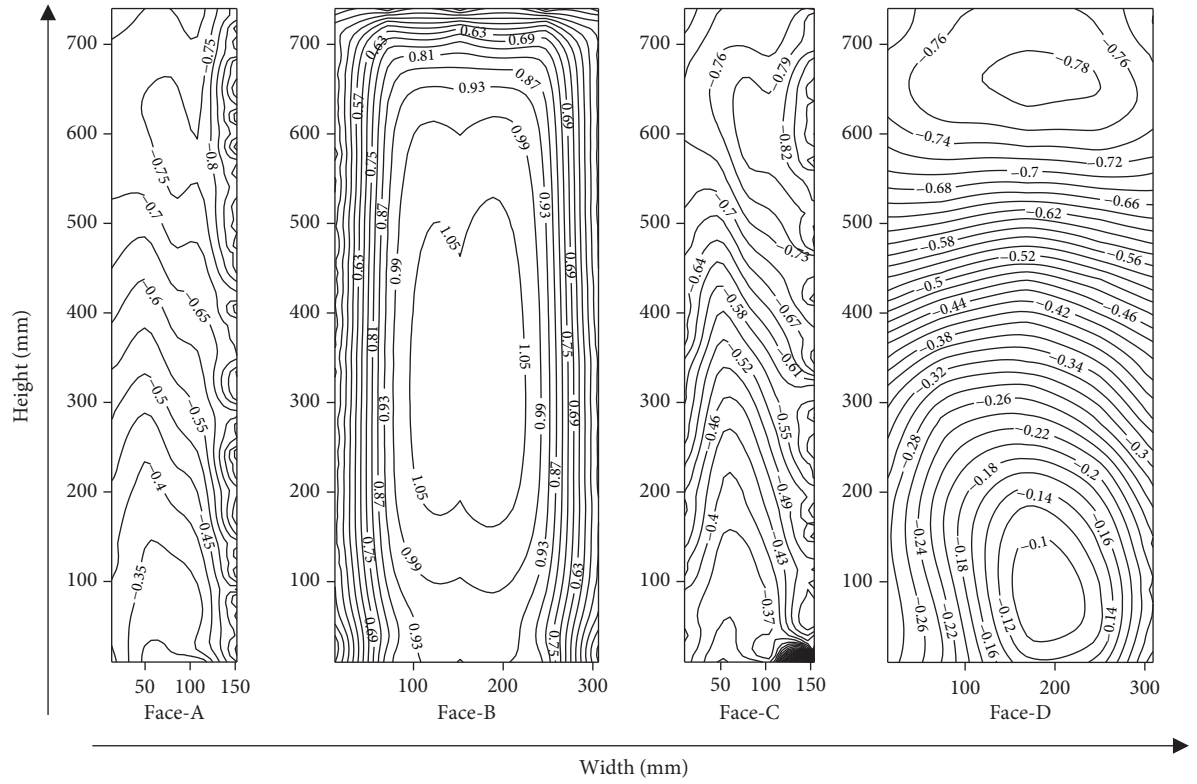



Figure 12: Pressure contours at 90° wind for model-A.

model is represented in Figure 12. The pressure on the wind ward face in both the case is having nearly the same pressure distribution pattern. The maximum pressure is observed in the case of 0° wind of 0.95 in the central part of face-A, while face-B in 90° wind is having the maximum of around 1.05, which is more than face-A for 0° wind because of the face

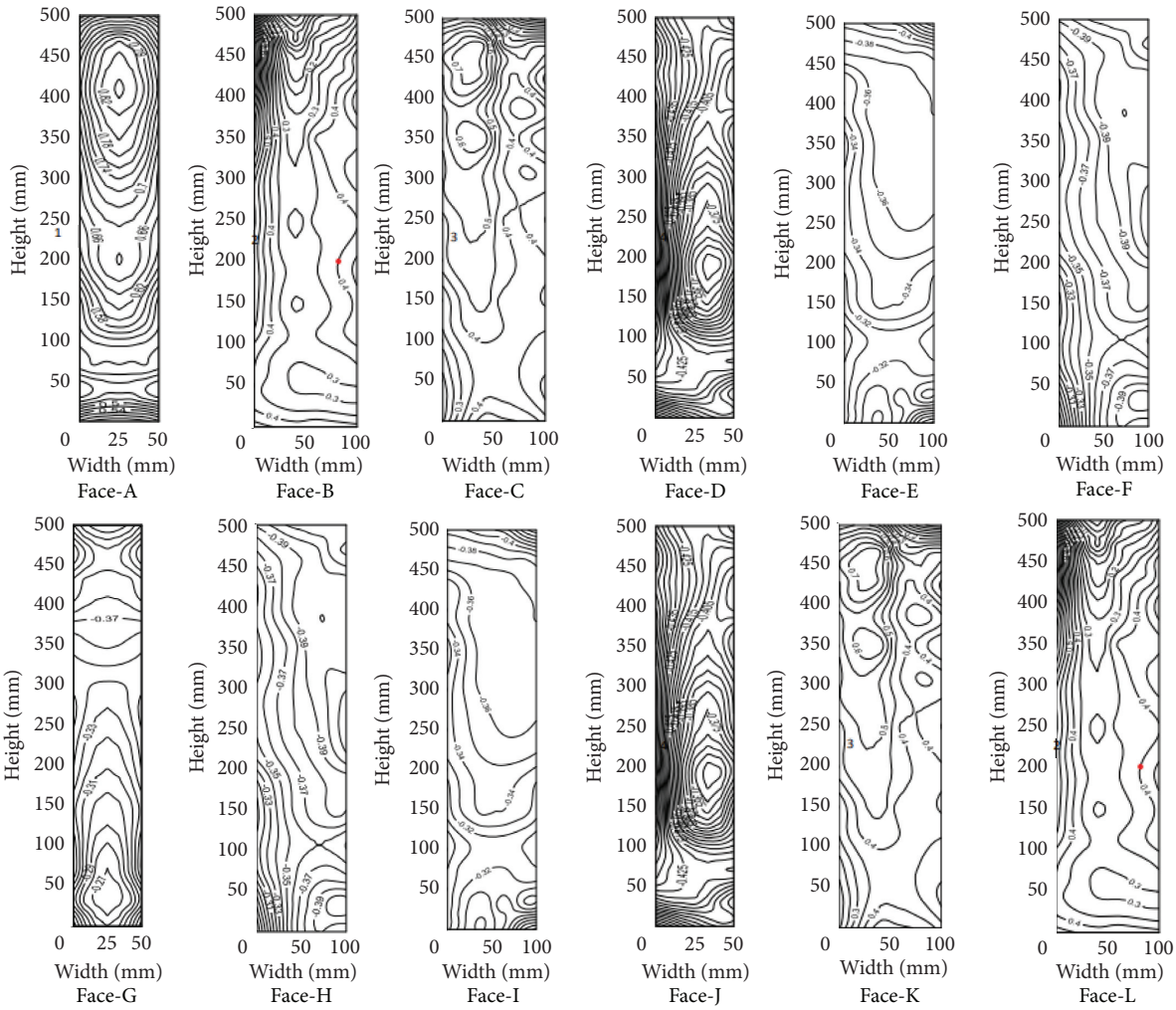



Figure 13: Pressure contours for 0° wind "+" shape [63].

width is increased and that is why the pressure exposure area is increased. For 0° wind, pressure along face-B starts to increase from upstream to downstream of wind, this is because of reattachment of flow. Face-B and face-D is having almost the same pressure distribution pattern, this is because of the symmetric faces of a building model. Leeward face, i.e., face-C, is having suction (negative pressure) and the maximum is on the bottom part of the building, while suction is increased at face-C from base to top of the model.

Figure 12 represents the pressure distribution variation of model-A. Pressure is presented in the form of contours, which is varying as per the shape and size of the particular face. In the case of 90° wind angle, face-A and face-C are having almost the identical pressure distribution pattern. This is increased from the ground of the model to the top of the building model. Also, pressure in the negative is increased as the wind moves from upstream to downstream. Pressure along face-D (Leeward) is having highest negative pressure in the top one-third part of the building model.

Figure 13 is pictorially representing [63] the pressure distribution in the form of contours for the "+" shaped building model, while present studies numerically investigate the wind effects on model-B. Model-B (rectangular model having corner cuts) pressure contours for 0° wind is represented in Figure 14, while for 90° wind the pressure contours for model-B are represented in Figure 15. The nature of pressure distribution on windward face-A is identical and the maximum pressure of 1.04 is spotted on face-A, while the face-A of the "+" shaped building is having maximum of 1.06. This is clearly depicted that each face of model-B is having a pressure distribution pattern and is the same as that with the experimental study performed on the "+" shaped model [63]. The pressure distribution variation is not much dependent on the size of the face, while the height more or less controls the pressure patterns.

Model-C (Y-Chamfer) is investigated numerically for wind effects and the pressure distribution pattern for 0° wind is in Figure 16, and for 90° wind, pressure contours are represented in Figure 17. The maximum positive pressure of 0.55 is on face-D, while the maximum negative of −1.3 is on face-E. The pressure variation along the shape and size is

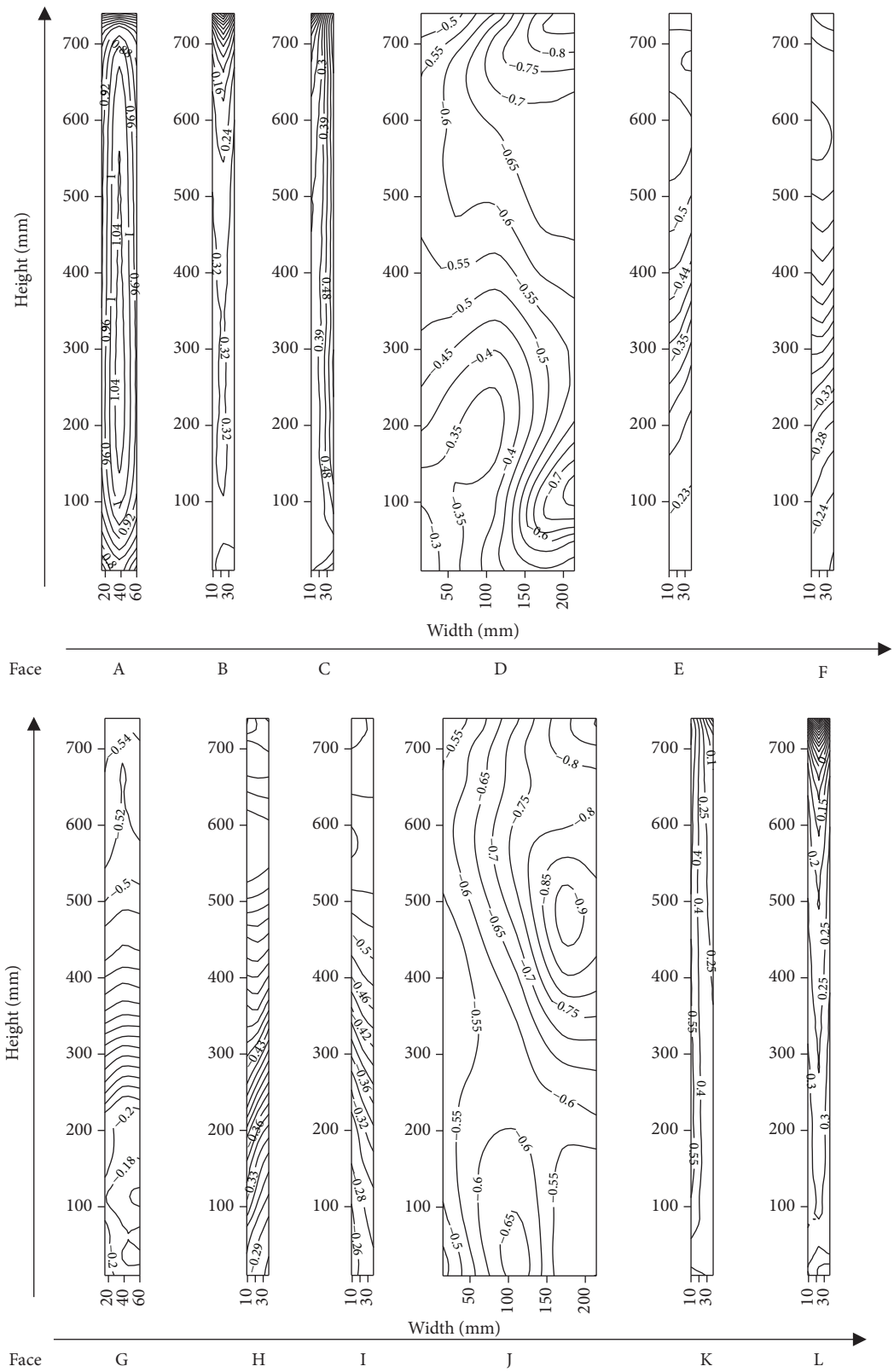



Figure 14: Pressure contour at 0° wind for model-B.

clearly demonstrating the magnitude of pressure as contours are showing with the label, and such labels are showing the increment or decrement in pressure. The actual phenomenon along model-C is valuable for the designer to investigate the wind effects on such types of corner configurations.

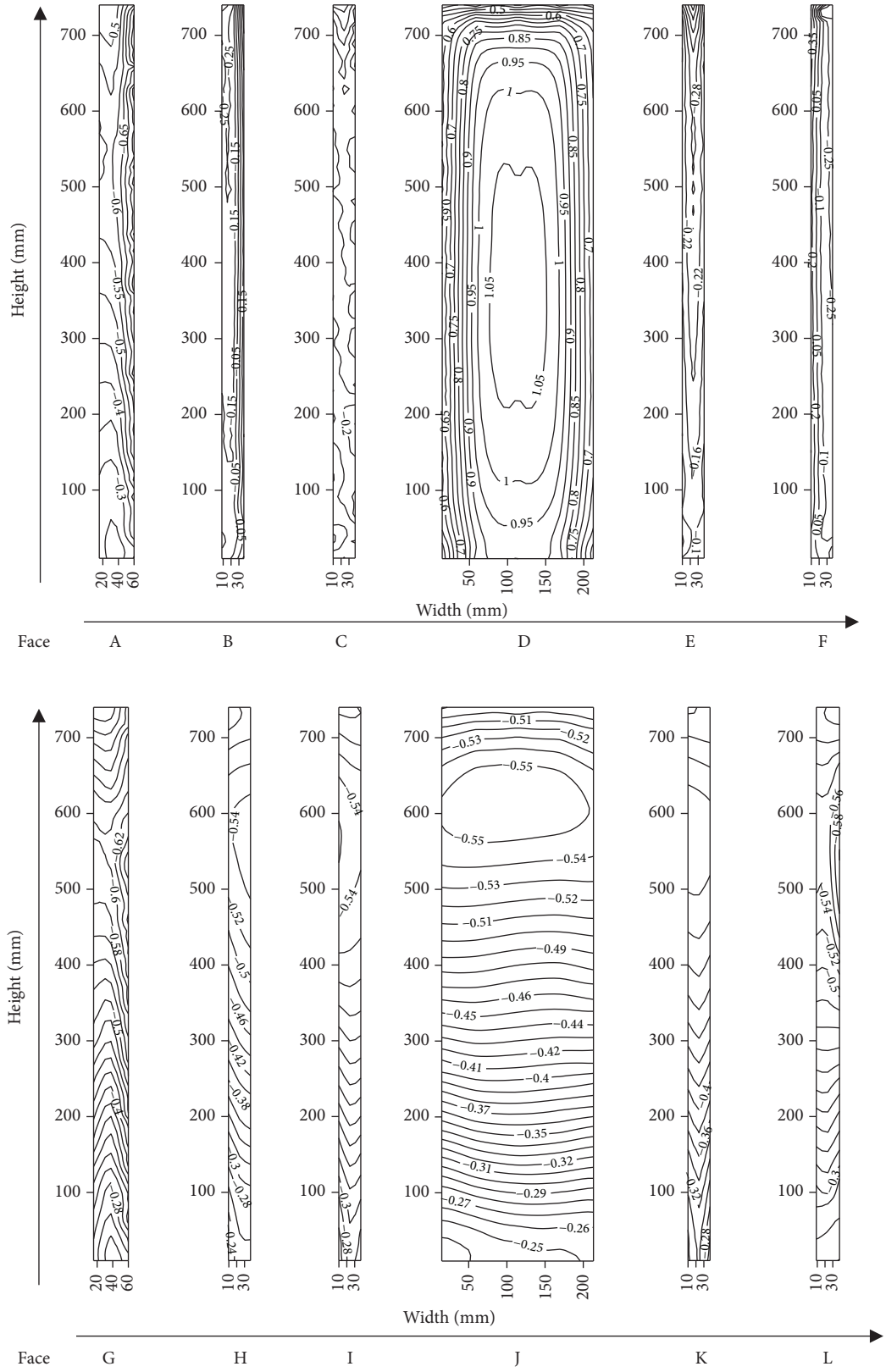



Figure 15: Pressure contour at 90° wind for model-B.

Pressure distribution on model-D is represented graphically in Figure 18 for 0° wind, while pressure contours in Figure 19 is for 90° wind. The positive pressure is observed on face-A, B, C, D, M, and N, while the negative pressure is noticed on face-E, F, G, H, I, J, K, L, and O in the case of 0° wind. The maximum positive pressure of 0.64 is observed on

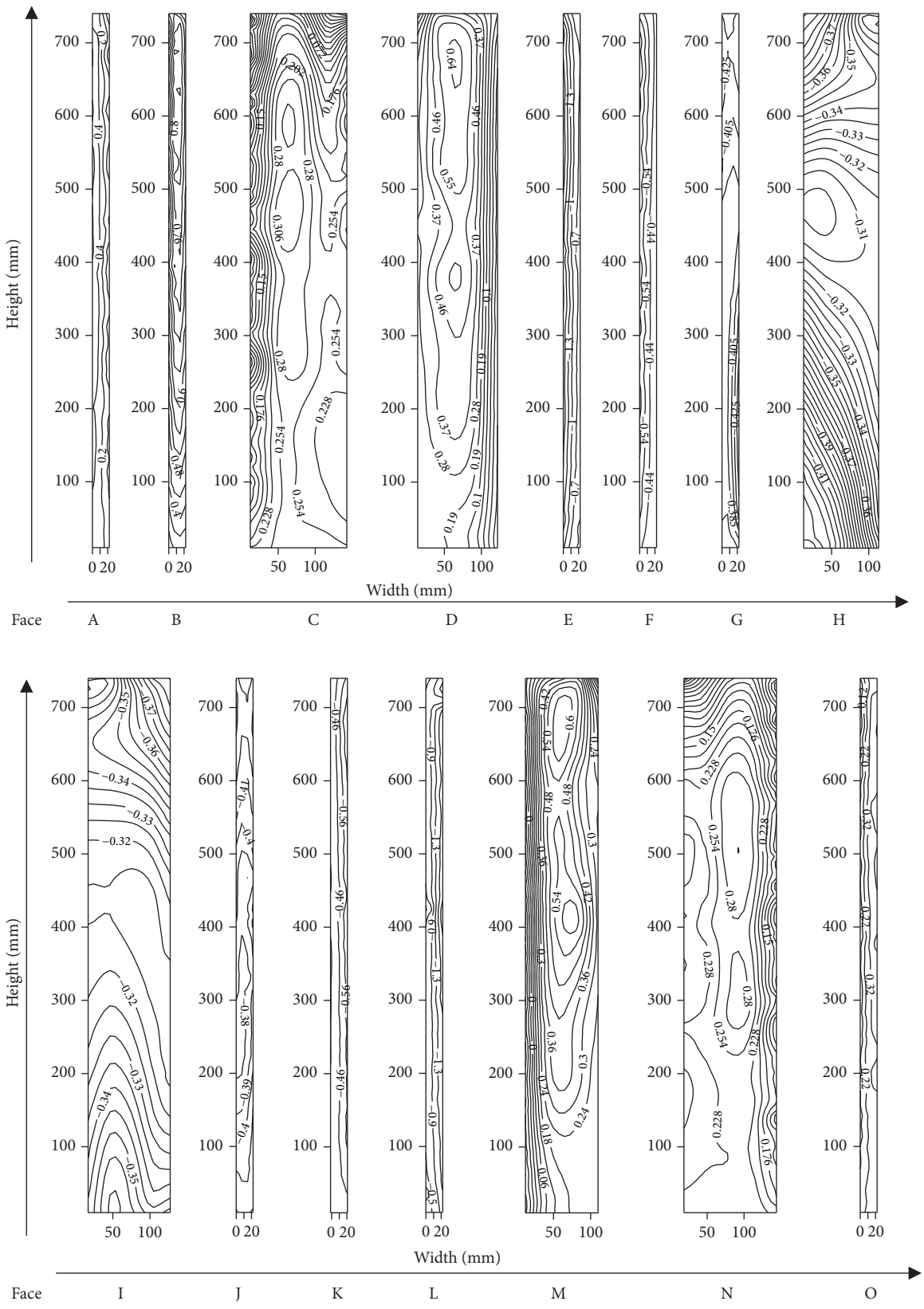



Figure 16: Pressure contours at 0° wind for model-C.

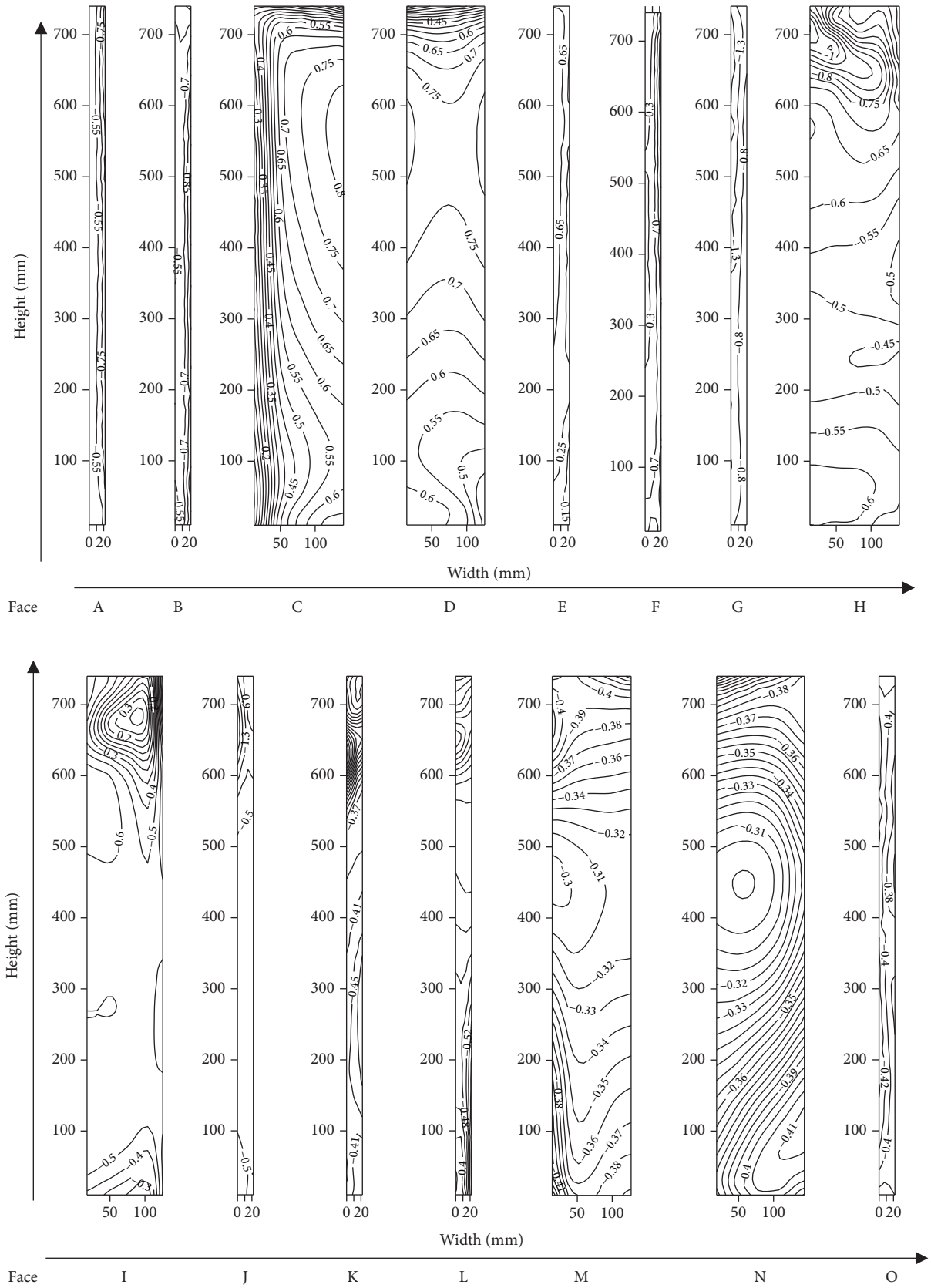



Figure 17: Pressure contours at 90° wind for model-C.

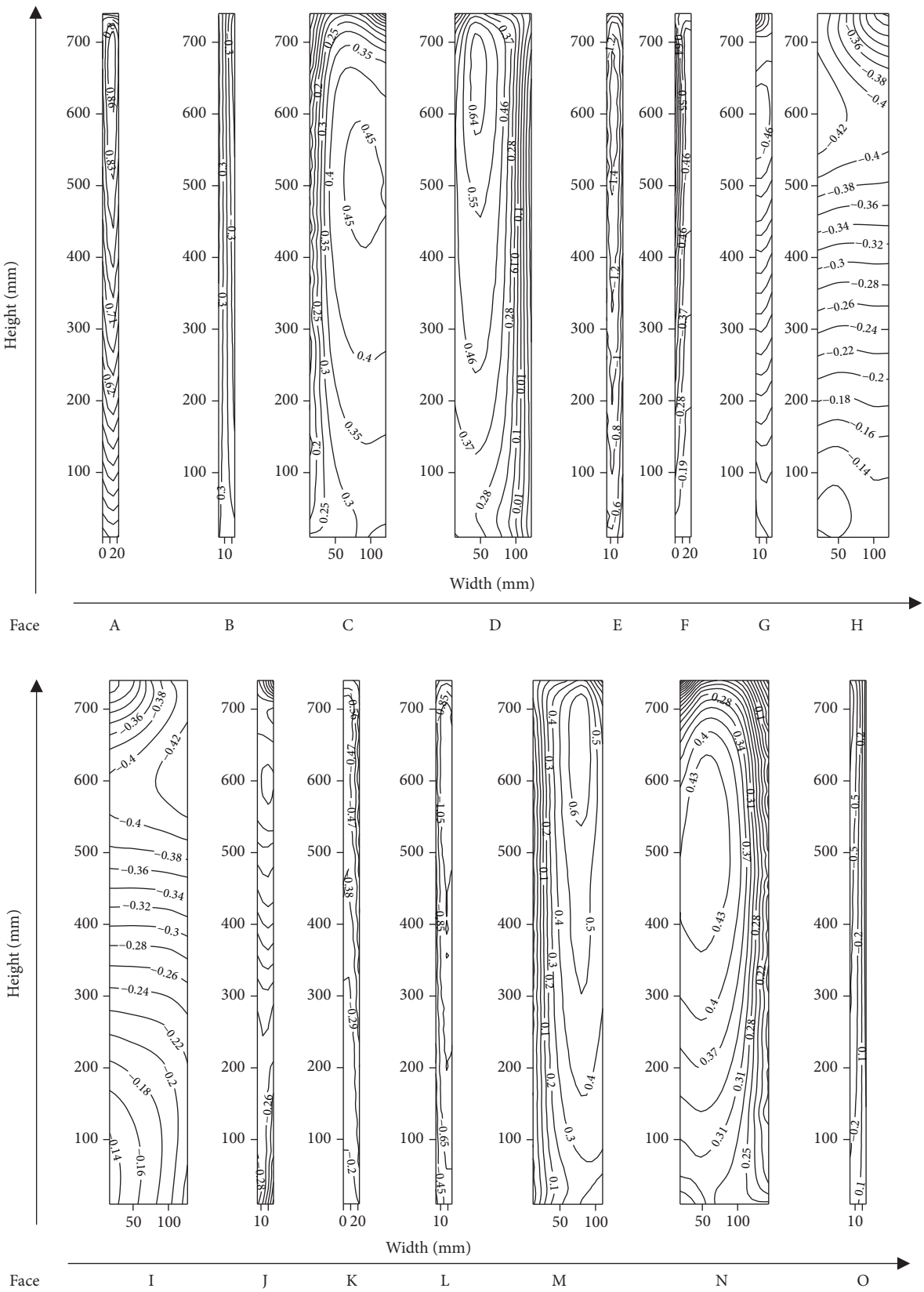



Figure 18: Pressure contours at 0° wind for model-D.

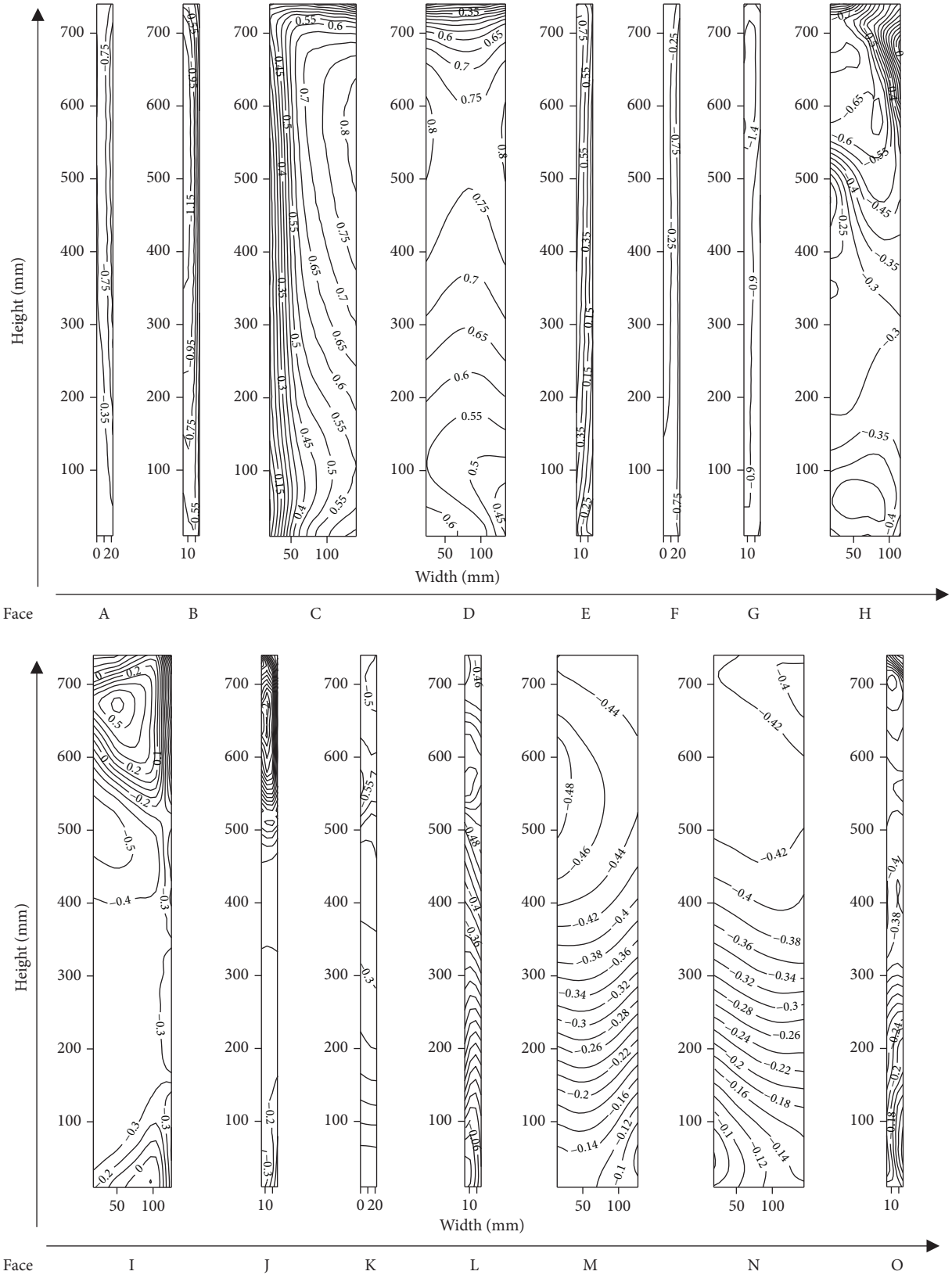



Figure 19: Pressure contours at 90° wind for model-D.

face-*D*, while the maximum suction of −1.4 is spotted on face-E in the case of 0° wind. Pressure contours for 90° wind for model-D (Y-fillet) is represented graphically Figure 19. The positive pressure is spotted on face-C, D, and E only in the case of 90° wind, while other faces are under the effect of negative pressure. The maximum positive pressure of 0.75 is observed on face-C and D, while the negative pressure of −1.4 is noticed on face-G for model-D (Y-fillet). This is also noticed from pressure contours that symmetrical faces are having same pressure distribution patterns.

## 4. Conclusion

The present numerical study is performed to investigate the wind effects on the corner effects of the tall building having various corners (corner cut, chamfer, and fillet) in the corner of the cross-sectional plan area of the building. The present study is performed using ANSYS CFX by utilizing the $k$-$\varepsilon$ turbulence model. The wind incidence angle varies in the range of 0° to 180° at the interval of 15° each. All four building models have the same corner modification ratio, as well as the same area and height. Validation studies show the closer agreement of the numerical simulation result with the experimental studies and different international standards. Various results of the pressure distribution wind force coefficient are calculated and presented. The notable outcomes from the present are as follows:

(i) Validation investigations indicate a strong degree of congruence between numerical results and experimental results and a number of international standards. Validation is carried out on two distinct models of rectangles, one with a regular shape and one with a corner cut.

(ii) The shape of the building is having the significant effect in determining the wind load on the tall building, wind forces are varying with the wind incidence angle. A regular plan shape model with no corner modification and a model with corner modification exhibit identical along the wind response, but the overall lift is less for the model having corner modification.

(iii) The maximum wind forces at 30°, 90°, and 150° wind angles were observed for an irregular "Y" shaped building with modified corners. These types of structures are well-suited to withstand such high-intensity forces, if appropriate openings are provided in the cladding unit.

(iv) Among the regular shape models, a model with corner cuts is optimal; however, an irregular "Y" shaped model with fillet corners into each limb is optimal for resisting overall wind force. To avoid excessive wind loads, the base shear of all four buildings is also investigated.

(v) The majority of available international standards are silent on providing data for irregular shape models; however, this study investigates wind load information for such buildings.

(vi) The designer can choose the best model from one of these four models, since this study compares the wind effects on equal area building models with the same ratio of modifications and the results are presented in the form of pressure contours for the regular and irregular models.

*4.1. Future Scope.* The present study analyses and compares the result of the simple type of corner and corner cut on a regular shaped building model, while for irregular shaped modifications of chamfer and fillet shape is provided. Different types of modification can be applied and checked for stability against the wind load.

## Nomenclature

| | |
|---|---|
| $\rho$: | Density of air |
| $F$: | Body force per unit volume |
| $u_i$: | The filtered scale velocity field |
| $\mu$: | Dynamic viscosity of air |
| $t$: | Time step |
| $\mu_t$: | Eddy or turbulent viscosity |
| $x$, $y$, $z$: | System of rectangular cartesian coordinates |
| $\in$: | Dissipation rate of $k$ |
| $P$: | Pressure |
| $k$: | von Karman's constant |
| $u, v, w$: | Fluctuating wind |
| $\tau_{ij}$: | Turbulent stress tensor |
| $z_o$: | Reference height |
| $z$: | Height above the ground |
| $n$: | Power law index |
| $v_z$: | Mean wind speed at any height |
| $v_o$: | Mean wind speed at reference height |
| $P$: | Pressure at the point |
| $P_0$: | Static pressure at reference height |
| $U_H$: | Mean wind speed at reference height |
| $C_F$: | Force coefficient |
| $C_M$: | Base moment coefficient |
| $A_p$: | Area projected |
| $H$: | Height of the building model |
| $\beta$, $\beta^*$, $\sigma^*$, $\sigma$: | Closure coefficients |
| $k$: | Turbulent kinetic energy |
| $\overline{u_i u_j}$: | Reynold stress |
| $i$: | General values of $u, v, w$ component at point |
| BLWT: | Boundary Layer Wind Tunnel |
| TTU: | Texas Tech University |
| SST: | Sher Stress Transport |
| GC: | Grid Convergence |
| ABL: | Atmospheric boundary layer flow |
| CFD: | Computational fluid dynamics |
| WT: | Wind Tunnel |
| atm: | Atmospheric. |

## Data Availability

All data, models, and code generated or used during the study appear in the submitted article. Data are available from

the corresponding author upon request in the form of contour plots and graphs.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The authors would like to express their sincere gratitude to the Delhi Technological University, Delhi, India, for providing funding to conduct the research work and for the institutional fellowship for the second author under Institutational fellowship for the Ph.D.

## References

[1] I S, "Indian Standard Design Loads (Other than Earthquake) for Buildings and Structures-Code of Practice,part 3(wind loads)," Indian Standard, Mirzapur, I. S 875, 2015.

[2] As/Nzs, "Structural Design Actions - Part 2," As/Nzs:1170, p. 2, Wind actions. Standards Australia/Standards New Zealand, Sydney, Wind actions. Standards Australia/Standards New Zealand, 2011.

[3] Asce, "Minimum Design Loads and Associated Criteria for Buildings and Other Structures," *Structural Engineering Institute of the American Society of Civil Engineering, Reston*, 2017.

[4] Gb 50009-2001, *National Standard Of The People's Republic Of China*, General Administration of Quality Supervision, Inspection and Quarantine of The People's Republic of China, China, 2002.

[5] BE, *Code of Practice on Wind Effects in Hong Kong 2004*, pp. 68–70, Buildings Department, Hong Kong, 2004.

[6] NSCP C101-15, *National Structural Code of the Philippines 2015*, Association of Structural Engineers of the Philippines, Quezon, Philippines, 2015.

[7] Mnbc, *Myanmar National Building Code 2020*, International relation and legal Section Department of Building Minis, 2020.

[8] Ministry of Housing and Urban Development, "National building code," pp. 1–476, Federal. Repub. Nigera, Nigera West Africa, 2006.

[9] Mvcs, *National Building Regulations*, Bureau of indian standards, New Delhi, India, 2018.

[10] Bnbc, *Bangladesh National Building Code (BNBC) 2020*, House Build. Res. Inst, 2020.

[11] R. Tavakoli, R. Kamgar, and R. Rahgozar, "Optimal location of energy dissipation outrigger in high-rise building considering nonlinear soil-structure interaction effects," *Periodica Polytechnica: Civil Engineering*, vol. 64, pp. 887–903, 2020.

[12] M. Dadkhah, R. Kamgar, H. Heidarzadeh, A. Jakubczyk-Galczyńska, and R. Jankowski, "Improvement of performance level of steel moment-resisting frames using tuned mass damper system," *Applied Sciences*, vol. 10, pp. 34–38, 2020.

[13] R. Kamgar and P. Rahgozar, "Optimum location for the belt truss system for minimum roof displacement of steel buildings subjected to critical excitation," *Steel and Composite Structures*, vol. 37, pp. 463–479, 2020.

[14] R. Kamgar, R. Tavakoli, P. Rahgozar, and R. Jankowski, "Application of discrete wavelet transform in seismic nonlinear analysis of soil–structure interaction problems," *Earthquake Spectra*, vol. 37, 2021.

[15] M. Dadkhah, R. Kamgar, and H. Heidarzadeh, "Reducing the cost of calculations for incremental dynamic analysis of building structures using the discrete wavelet transform," *Journal of Earthquake Engineering*, vol. 26, 2020.

[16] M. Dadkhah, R. Kamgar, and H. Heidarzadeh, "Improving the nonlinear seismic performance of steel moment-resisting frames with minimizing the ductility damage index," *SN Applied Sciences*, vol. 3, 2021.

[17] R. Kamgar and M. R. B. Samani, "Numerical study for evaluating the effect of length-to-height ratio on the behavior of concrete frame retrofitted with steel infill plates," *Practice Periodical on Structural Design and Construction*, vol. 27, pp. 1–10, 2022.

[18] K. C. S. Kwok, "Effect of building shape on wind-induced response of tall building," *Advances in Wind Engineering*, vol. 28, pp. 381–390, 1988.

[19] R. Jozwiak, J. Kacprzyk, and J. A. Zuranski, "Wind Tunnel Investigation of Interference Effects on Pressure Distribution on a Building," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 57, 1995.

[20] S. Pal, R. K. Meena, R. Raj, and S. Anbukumar, "Wind tunnel study of a fish - plan shape model under different isolated wind incidences," vol. 5, pp. 353–366, 2021.

[21] S. K. Nagar, R. Raj, and N. Dev, "Experimental study of wind - induced pressures on tall buildings of different shapes," *Wind Struct. An Int. J*, vol. 5, pp. 441–453, 2020.

[22] S. Pal, R. Raj, and S. Anbukumar, "Comparative study of wind induced mutual interference effects on square and fish-plan shape tall buildings," *Sādhanā*, vol. 46, no. 2, p. 86, 2021.

[23] S. K. Nagar, R. Raj, and N. Dev, "Proximity effects between two plus-plan shaped high-rise buildings on mean and RMS pressure coefficients," *Scientia Iranica*, vol. 29, 2021.

[24] S. Pal, R. Raj, and S. Anbukumar, "Bilateral interference of wind loads induced on duplicate building models of various shapes," *Latin American Journal of Solids and Structures*, vol. 18, 2021.

[25] J. A. Amin and A. K. Ahuja, "Aerodynamic modifications to the shape of the buildings: a review of the state-of-the-art," *Asian J. Civ. Eng*, vol. 11, pp. 433–450, 2010.

[26] P. A. Blackmore, "The role of wind tunnel testing in the design of building structures," *Proceedings of the Institution of Civil Engineers - Structures and Buildings*, vol. 122, no. 3, pp. 253–265, 1997.

[27] B. Bhattacharyya and S. K. Dalui, "Experimental and numerical study of wind-pressure distribution on irregular-plan-shaped building," *Journal of Structural Engineering*, vol. 146, Article ID 04020137, 2020.

[28] Y. Tominaga and T. Stathopoulos, "Turbulent Schmidt numbers for CFD analysis with various types of flowfield," *Atmospheric Environment*, vol. 41, no. 37, pp. 8091–8099, 2007.

[29] B. Blocken, "Computational Fluid Dynamics for urban physics: importance, scales, possibilities, limitations and ten tips and tricks towards accurate and reliable simulations," *Building and Environment*, vol. 91, pp. 219–245, 2015.

[30] S. Behera, D. Ghosh, A. K. Mittal, Y. Tamura, and W. Kim, "The effect of plan ratios on wind interference of two tall buildings," *The Structural Design of Tall and Special Buildings*, vol. 29, 2020.

[31] Y. Tamura, X. Xu, H. Tanaka, Y. C. Kim, A. Yoshida, and Q. Yang, "Aerodynamic and pedestrian-level wind characteristics of super-tall buildings with various configurations," *Procedia Engineering*, vol. 199, pp. 28–37, 2017.

[32] A. K. Mittal, D. Ghosh, S. Behera, I. A. Siddiqui, and D. S. Dharmshaktu, "Wind flow simulation in the vicinity of tall buildings through CFD," in *Proceedings of the 8th Asia-Pacific Conf. Wind Eng. APCWE*, pp. 682–690, Chennai, India, December 2013.

[33] B. Blocken, T. Stathopoulos, and J. Carmeliet, "CFD simulation of the atmospheric boundary layer: wall function problems," *Atmospheric Environment*, vol. 41, no. 2, pp. 238–252, 2007.

[34] X. Tian, M. C. Ong, J. Yang, and D. Myrhaug, "Unsteady RANS simulations of flow around rectangular cylinders with different aspect ratios," *Ocean Engineering*, vol. 58, pp. 208–216, 2013.

[35] N. Gaur and R. Raj, "Aerodynamic mitigation by corner modification on square model under wind loads employing CFD and wind tunnel," *Ain Shams Engineering Journal*, vol. 13, 2021.

[36] R. Raj, T. Rana, T. Anchalia, and U. Khola, "Numerical study of wind excited action on H Plan-shaped tall building," *International Journal on Emerging Technologies*, vol. 11, pp. 591–605, 2020.

[37] M. Keerthana and P. Harikrishna, "Application of CFD for assessment of galloping stability of rectangular and H-sections," *Journal of Scientific & Industrial Research*, vol. 72, pp. 419–427, 2013.

[38] A. Kumar and R. Raj, "Study of pressure distribution on an irregular octagonal plan oval-shape building using CFD," *Civil Engineering Journal*, vol. 7, no. 10, pp. 1787–1805, 2021.

[39] R. Raj, A. Sharma, and S. Chauhan, *Response of Square and Plus Shaped Buildings on Varying Wind Loads*, in *Proceedings of the ASCE India Conference 2017*, pp. 206–215, New Delhi , India, December 2018.

[40] N. Gaur, R. Raj, and P. K. Goyal, "Interference effect on corner - configured structures with variable geometry and blockage configurations under wind loads using CFD," *Asian J. Civ. Eng*, vol. 22, 2021.

[41] S. Mukherjee, S. Chakraborty, S. K. Dalui, and A. K. Ahuja, "Wind induced pressure on 'Y' plan shape tall building," *Wind and Structures*, vol. 19, no. 5, pp. 523–540, 2014.

[42] P. Sanyal and S. K. Dalui, "Effect of corner modifications on Y' plan shaped tall building under wind load," *Wind Struct. An Int. J*, vol. 30, pp. 245–260, 2020.

[43] P. Sanyal and S. K. Dalui, "Effects of Side Ratio for 'Y' Plan Shaped Tall Building under Wind Load," *Building Simulation*, vol. 14, 2020.

[44] P. Sanyal and S. K. Dalui, "Effects of internal angle between limbs of "Y" plan shaped tall building under wind load," *Journal of Building Engineering*, vol. 33, Article ID 101843, 2021.

[45] P. Sanyal and S. K. Dalui, "Comparison of aerodynamic coefficients of various types of Y-plan-shaped tall buildings," *Asian Journal of Civil Engineering*, vol. 21, no. 7, pp. 1109–1127, 2020.

[46] P. K. Goyal, S. Kumari, S. Singh, R. K. Saroj, R. K. Meena, and R. Raj, "Numerical study of wind loads on Y plan-shaped tall building using CFD," *Civil Engineering Journal*, vol. 8, no. 2, pp. 263–277, 2022.

[47] R. K. Meena, R. Raj, and S. Anbukumar, *Numerical Investigation of Wind Load on Side Ratio of High-Rise Buildings Numerical Investigation of Wind Load on Side Ratio of High-Rise Buildings*, Springer, Singapore, 2021.

[48] A. Kumar and R. Raj, "CFD study of flow characteristics and pressure distribution on Re-entrant wing faces of L-shape buildings," *Civil Engineering and Architecture*, vol. 10, no. 1, pp. 289–304, 2022.

[49] R. Raj and A. K. Ahuja, "Wind loads on cross shape tall buildings," *J. Acad. Ind. Res*, vol. 2, pp. 111–113, 2013.

[50] P. Sanyal and S. K. Dalui, "Effects of courtyard and opening on a rectangular plan shaped tall building under wind load," *International Journal of Advanced Structural Engineering*, vol. 10, no. 2, pp. 169–188, 2018.

[51] R. K. Meena, R. Raj, and S. Anbukumar, "Wind excited action around tall building having different corner configurations," *Advances in Civil Engineering*, vol. 2022, Article ID 1529416, 2022.

[52] M. Mallick, A. Mohanta, A. Kumar, and V. Raj, "Modelling of wind pressure coefficients on C-shaped building models," *Modelling and Simulation in Engineering*, vol. 2018, Article ID 6524945, 2018.

[53] J. A. Amin and A. K. Ahuja, "Effects of side ratio on wind-induced pressure distribution on rectangular buildings," *Journal of Structures*, vol. 2013, pp. 1–12, 2013.

[54] J. A. Amin and A. K. Ahuja, "Characteristics of wind forces and responses of rectangular tall buildings," *International Journal of Advanced Structural Engineering*, vol. 6, no. 3, pp. 1–14, 2014.

[55] R. Raj, "Effects Of Cross-Sectional Shapes On Response Of Tall Buildings Under Wind Loads," iit Roorkee, 2015, http://hdl.handle.net/123456789/14664.

[56] P. J. Richards and R. P. Hoxey, "Appropriate boundary conditions for computational wind engineering models using the k-$\epsilon$ turbulence model," in *Proceedings of the 1st International Symposium on Computational Wind Engineering (CWE 92)*, pp. 145–153, Tokyo, Japan, August 1993.

[57] Ansys, *ANSYS Meshing User's Guide*, Taylor and francis group an informa business, UK, 2010.

[58] E. M. Alawadhi, "Meshing guide," *Finite Elem. Simulations Using ANSYS*, vol. 15317, pp. 407–424, 2020.

[59] Ansys, "ANSYS CFX-Solver Modeling Guide," ANSYS INC, 2020, http://www.ansys.com.

[60] J. Revuz, D. M. Hargreaves, and J. S. Owen, "On the domain size for the steady-state CFD modelling of a tall building," *Wind and Structures*, vol. 15, no. 4, pp. 313–329, 2012.

[61] J. Franke, A. Hellsten, H. Schlünzen, and B. Carissimo, *Best Practice Guideline for the Cfd Simulation of Flows in the Urban Environment. Cost 732: Quality Assurance and Improvement of Microscale Meteorological Models*, New Delhi, India, 2007.

[62] J. H. Ferziger and M. Peric, "Computational Methods for Fluid Dynamics," 2002, http://marefateadyan.nashriyat.ir/node/150.

[63] S. Chakraborty, S. K. Dalui, and A. K. Ahuja, "Experimental investigation of surface pressure on "+" plan shape tall building," *Jordan J. Civ. Eng*, vol. 8, pp. 251–262, 2014.

**How to Cite**:
Singh, H., Goyal, H., Singh, M., & Raj, R. (2022). Wind interference effect on a hexagonal-shaped high building with orifices. *International Journal of Health Sciences*, *6*(S1), 8963–8982.
https://doi.org/10.53730/ijhs.v6nS1.7040

# Wind interference effect on a hexagonal-shaped high building with orifices

**Harmanbir Singh**
Undergraduate Student, Department of Civil ngineering, Delhi Technological University (DTU), Delhi, India

**Harsh Goyal**
Undergraduate Student, Department of Civil ngineering, Delhi Technological University (DTU), Delhi, India

**Mayank Singh**
Undergraduate Student, Department of Civil ngineering, Delhi Technological University (DTU), Delhi, India

**Ritu Raj**
Assistant Professor, Department of Civil Engineering, Delhi Technological University (DTU), Delhi, India
Corresponding author email: rituraj@dtu.ac.in

***Abstract*---**This research examined at how the wind load and pressure on a tall hexagonal structure varied as the apertures changed. Wind angles ranging from 0 to 90 degrees with a 30-degree gap were employed in various situations. The distance between the interference building and the model building was also gradually increased, at 50mm intervals, from 50mm to 200mm. The major building model was examined at a scale of 1:500. The validity of the ANSYS CFX package was also evaluated using a rigid model with dimensions of 100mm x 100mm and a height of 700mm. The k-ε -turbulence model was used in the study, which was carried out using Computational Fluid Dynamics (CFD) and the ANSYS CFX software package. To demonstrate the link between distinct scenarios, interference factors (IF) and contours were employed. Interfering buildings' impacts on the hexagonal-shaped building are explored. The pressure coefficient of the faces rises as the distance between the interference building and the model building grows. The impacts of apertures on the model building were also examined, with 5-20% openings given.

***Keywords*---**Hexagonal-Shape Building, ANSYS (CFX), Wind Pressure Coefficient, Contours, Interference Factor

## Introduction

The introduction of new technologies and the pace of scientific advancement have resulted in a growth in the height, width, and scale of structures, making them even more vulnerable to the impacts of wind. As a result, the influence of wind loads on these sorts of structures becomes increasingly critical [1]. In this work, wind loads on the model structure were first analysed, and then the change in wind loads was investigated when openings were added to our model.

Currently, various research have been conducted utilising computational fluid dynamics (CFD) to investigate the impacts of wind on tall buildings. ANSYS (CFX) is a popular tool for analysing wind load on many types of constructions. Buildings with complex forms may also be simulated for numerical analysis, and the shape can be subdivided into various smaller pieces using proper meshing [2]. The research [3] investigated the mean pressure coefficient of a 'E' plan shaped tall building experimentally and computationally. Wind tunnel testing was used for experimental purposes, while CFD simulation was used for numerical analysis. These two tests were found to be in good agreement with one another.

Another study [4] conducted focused on the numerical analysis of tall structures utilising the Commonwealth Advisory Aeronautical Council (CAARC) building model. In the investigation, a numerical data comparison with measurements in a boundary wind tunnel revealed that an accurate time-dependent analysis is critical. The study [5] looked on differences in the stress and pressure coefficients of a tall building with an unsymmetrical or uneven 'E' plan form.. Again, experimental and numerical analyses were conducted utilising wind tunnel testing and CFD, and they were found to be in good agreement with one another. They determined that along wind hits had the largest positive force and pressure coefficients because the surface area is greater and the most wind energy is wasted on the frontal surface.

The current Indian Standard Code for wind loads over structures includes information on characteristics such as pressure fluctuation and force coefficients. However, when openings are supplied to irregular buildings, it fails to handle the intricacies of wind loads over them [6]. The major goal of this research is to examine wind loads over isolated structures as well as the influence of interference when a second building is provided during wind testing with openings of 10-15%.

## Methodology

### Model Shape Selection

The initial stage was to decide on the structure's form. Because several research on three and four-sided buildings have been conducted, the hexagonal shape was chosen for a realistic and distinct analysis [7,8,9].

### Model Structure

Using AutoCAD software, a 3D model of the building was constructed. The suitable height: width ratio was chosen based on cases of the IS 875 (Part 3) 2015 [6].

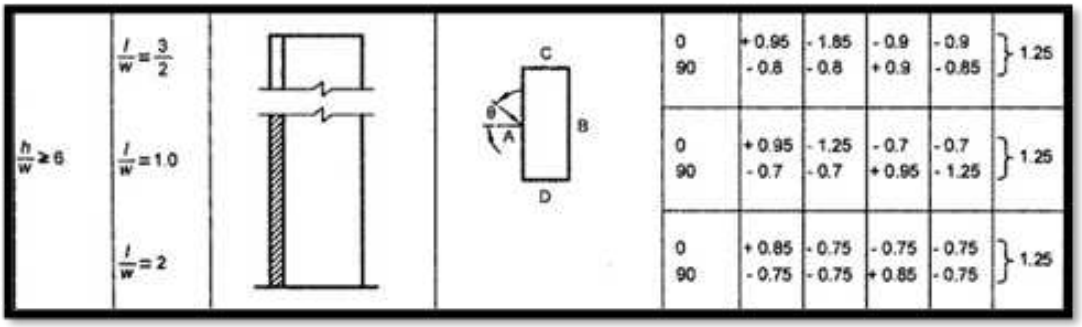



Figure 1: Describes the criteria for selecting the appropriate height: width ratio using IS 875 (Part 3): 2015 [6]

The structure that is being simulated in this study is explained in the next chapter. It displays all of the calculations, including the proportion of openings in the structure, as well as its form and size.

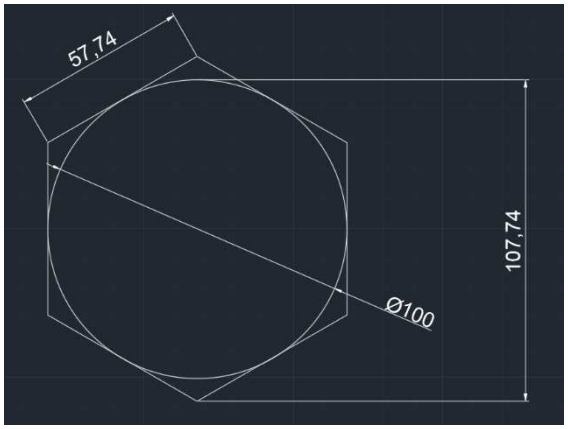



Figure 2: Top view of Model Building

The original height of the building is taken as 350 metres and the diameter of the circle is taken as 50 metres. The scale we chose for this study is 1:500 and thus the new dimensions for height and diameter become 700mm and 100mm respectively. Wall thickness of 1.2mm is taken.

**Openings**

| | | |
|---|---|---|
| Total area of the building | = | Area of roof + Area of walls |
| | = | 8661.74603 + 242508 |
| | = | 251170 $mm^2$ approximately. |
| Area of the openings | = | 42000 $mm^2$ |
| Percentage of openings | = | 16.72% |
| Size of a openings | = | 40 X 25 |
| Openings on a side | = | 7 |

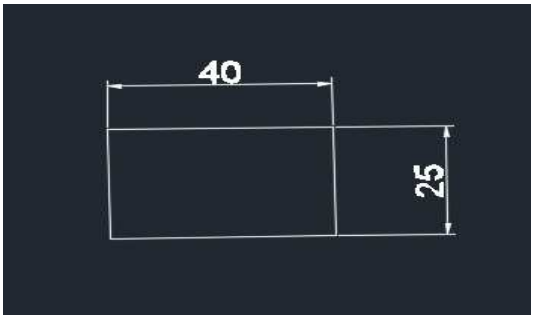



Figure 3: Dimensions of Openings

**Modelling in AutoCAD**

We utilised AutoCAD software to create the models. We employed a variety of workstation stations, including drafting and annotations, as well as 3d Basics. The two photos show the realistic and 2D wireframes of the model creation. The same design was used for both the model and the interference buildings.

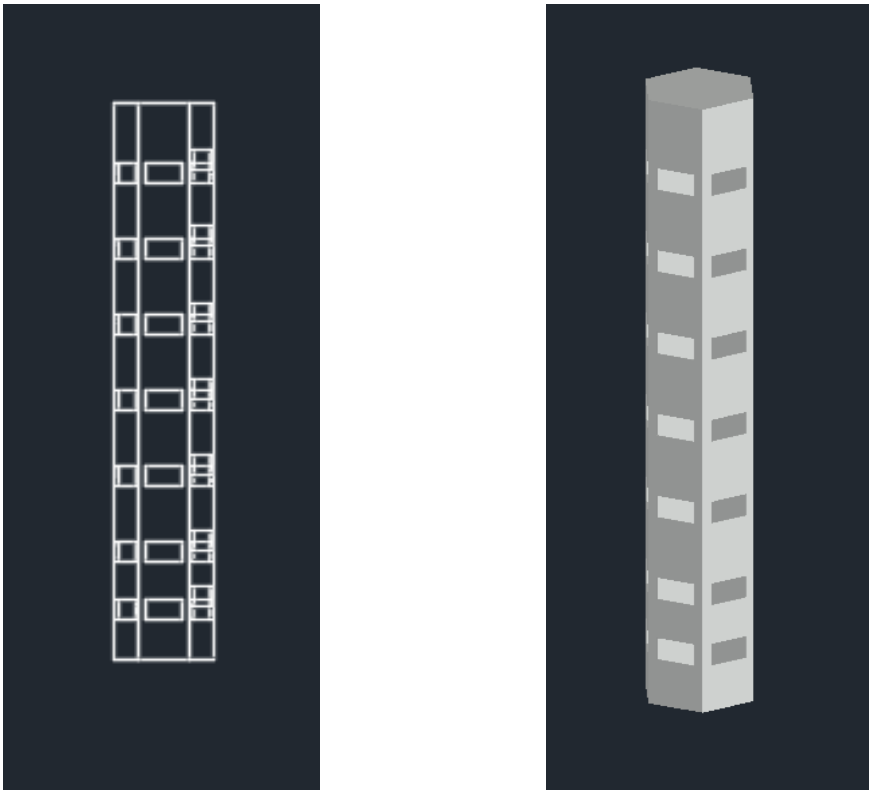


Figure 4: Includes the 2-D wire frame and realistic view of the model building which includes where openings have also been depicted.

**Validation using ANSYS**

The ANSYS CFX Program required to be confirmed before starting with the in-depth numerical examination. This requires the establishment of a domain.

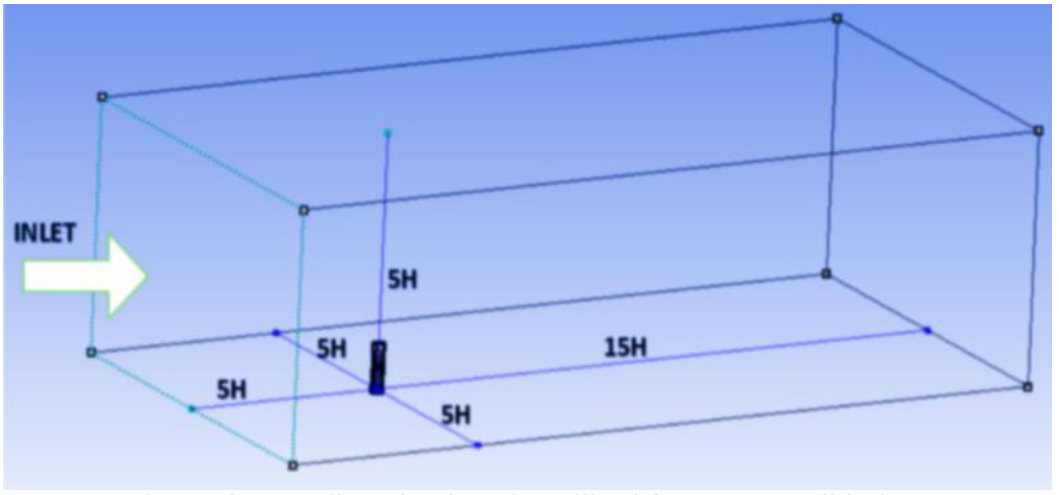



Figure 5: Describes the domain utilized for ANSYS validation.

The wind tunnel setting employed by [7] in their experimental investigation was like this region. The graph (Figure 6) displays the variability of velocity and turbulence intensity over the area and is equivalent to [7]

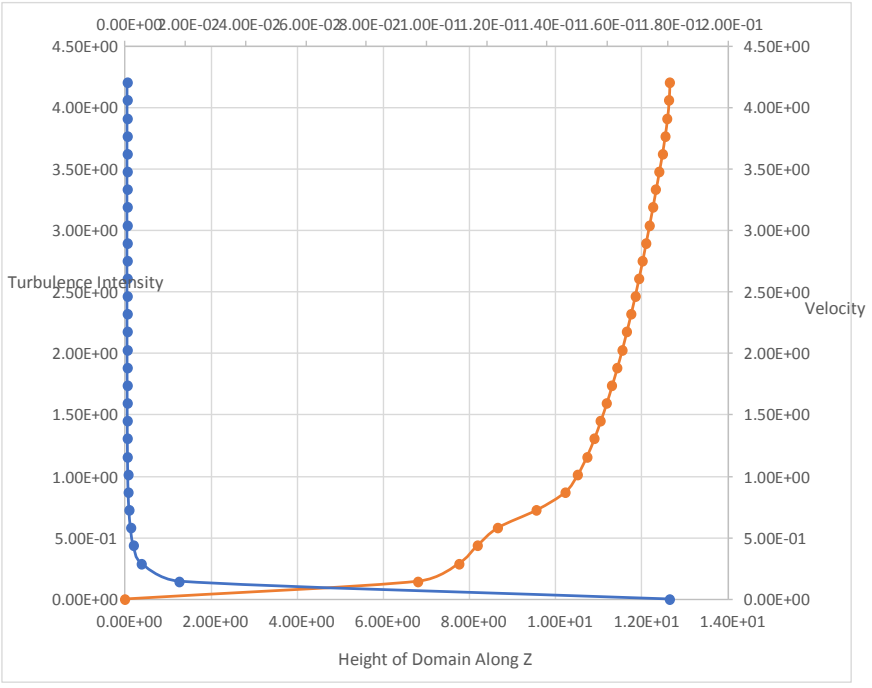



Figure 6: The fluctuation of velocity and turbulence intensity throughout the region is shown.

The model with L/W ratio one was validated in the domain using the k-epsilon [10] and SST [11] turbulence models discussed above. Wind impacts at 0° were explored for the models, and the results were compared to IS 875 (Part 3): 2015 [6].

**Domain**

The domain (Figure 5) was created in compliance with the parameters of [12]. The intake and outflow distances were calculated to be 5H and 15H, respectively, where H represents the height of the building model. Both the top clearance and the side aspect were rated 5H. As originally indicated, the length scale was changed to 1:500. This domain size is large enough to allow for the formation of vortices while avoiding the wind. The main causes of numerical simulation inaccuracy are the interference effects of the side and top faces. Reverse flow may occur due to a shortage of space in the outer direction, hindering the convergence of the CFD analysis. These can be prevented by following the advice of [8,12].

**Meshing**

The meshing of the domain was done with tetrahedral components (Fig. 7). This is where 0.2-meter-diameter elements were made. To properly measure the wind characteristics, more delicate meshing was used around the building model. Face scaling was the name of the game, and the element size was 0.05 metres. To ensure a steady flow, mesh inflating was erected around the outside of the structure. Mesh inflation has a total of 20 layers to choose from.

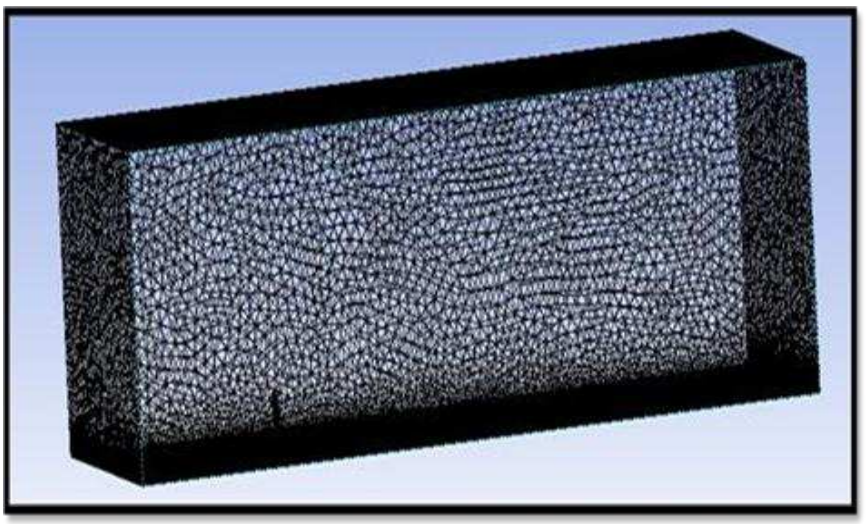


Figure 7: Shows the domain meshing, which was done with tetrahedral components

**Contours**

Contours are defined as shadowing on the simulating body, with various hues representing different values of the contour's parameter. The value increases as the contour becomes darker, and lowers as the shade becomes lighter.

We plotted the contour for Cp values in this investigation. Cp is the pressure coefficient, which represents how much pressure is applied and in which direction it is exerted. The negative sign reflects suction force in the opposite direction of the wind, whereas the positive sign represents force in the same direction as the wind.

A legend is always supplied on the side to provide us with the values representing different colours so that we may know which colour represents which value. The photographs below show the isolated building's faces A and B, which have sixteen percent openings.

The numbers on the legend show that face A on the front side has positive values, indicating that force is applied in the direction of the wind, but face B on the rear side has a negative Cp value, indicating the development of a tiny vortex on the back side of the structure.

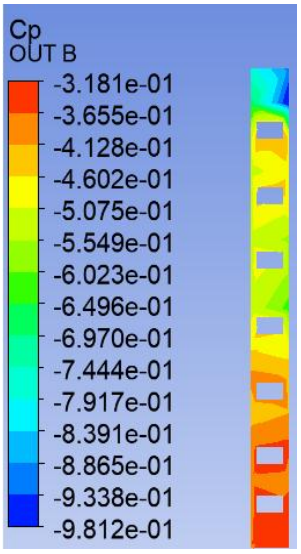



Fig 8: $C_P$ of Face A

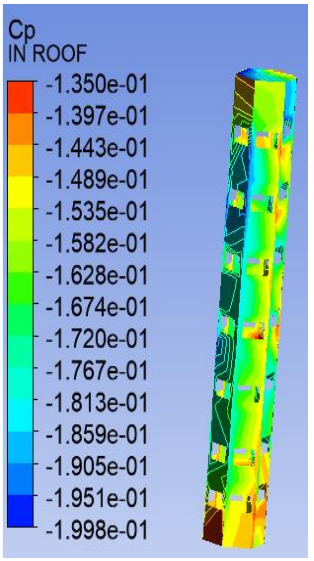



Fig 9: $C_P$ of Face B

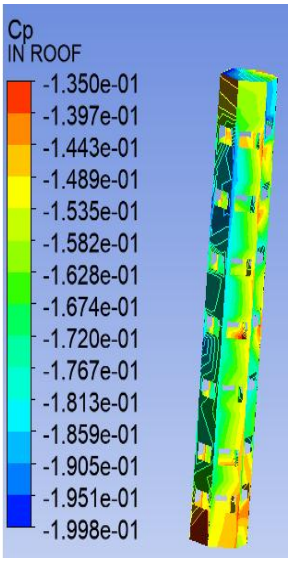



Fig 10: Contours of Inside Building

As the structure rotates, the face that confronts the wind direction experiences significant force, causing the force on face B to decrease and become less negative. This demonstrates that the face that facing the wind always feels the most pressure. When the building is rotated 30 degrees, the pressure is split across the two sides, A and D, indicating that this orientation is preferable than the one at 0 degrees.

Because all of the legend numbers are negative, the outlines of the inside building depict the suction force. The rotation of the building has essentially little influence on the interior structure since the total internal pressure of all the faces is almost constant.

The photographs below demonstrate that the value of Cp grows as the distance between the buildings increases, however we can detect a fall in Cp value between

100mm and 150mm spacing, which might be due to the creation of a vortex between the buildings, which lowers the Cp values. As the distance goes above 150mm, the effect of interreference building begins to disappear.

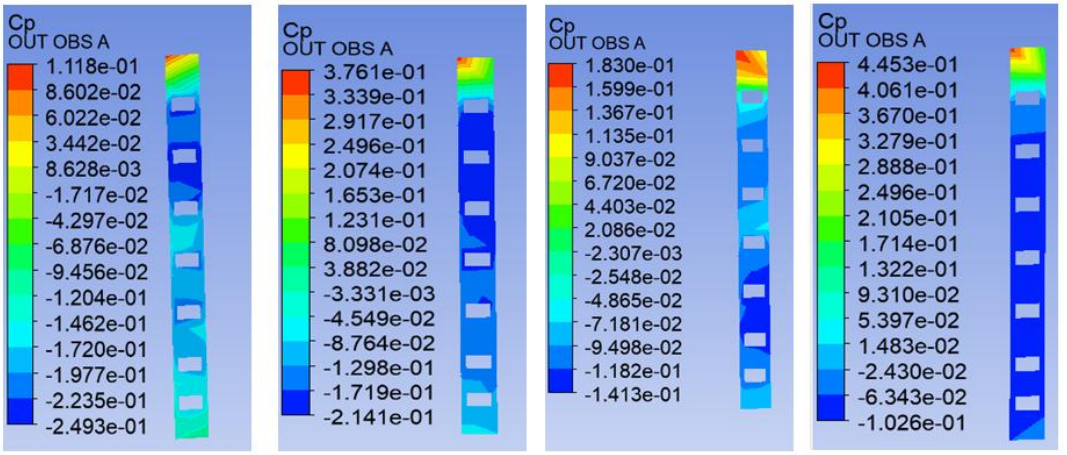



Figure 11: Contours of Face A for different spacing in ascending order.

Face B, on the other hand, has a normal rise in Cp values. The Cp value of this face is about half that of the isolated structure. This demonstrates that the existence of interference structures lessens suction force applied on that face.

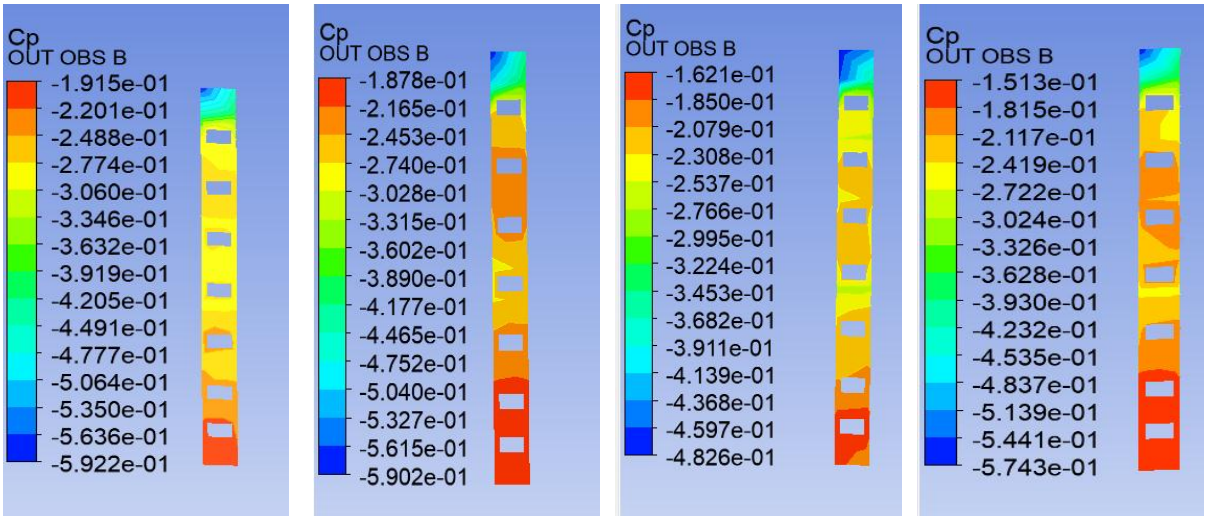



Figure 12: Contours of Face B for different spacing in ascending order

The photos below indicate that the presence of the interference building has greatly lowered the internal pressure of the building. Though increasing the distance increases the internal pressure, the increase is very minimal in all four circumstances.

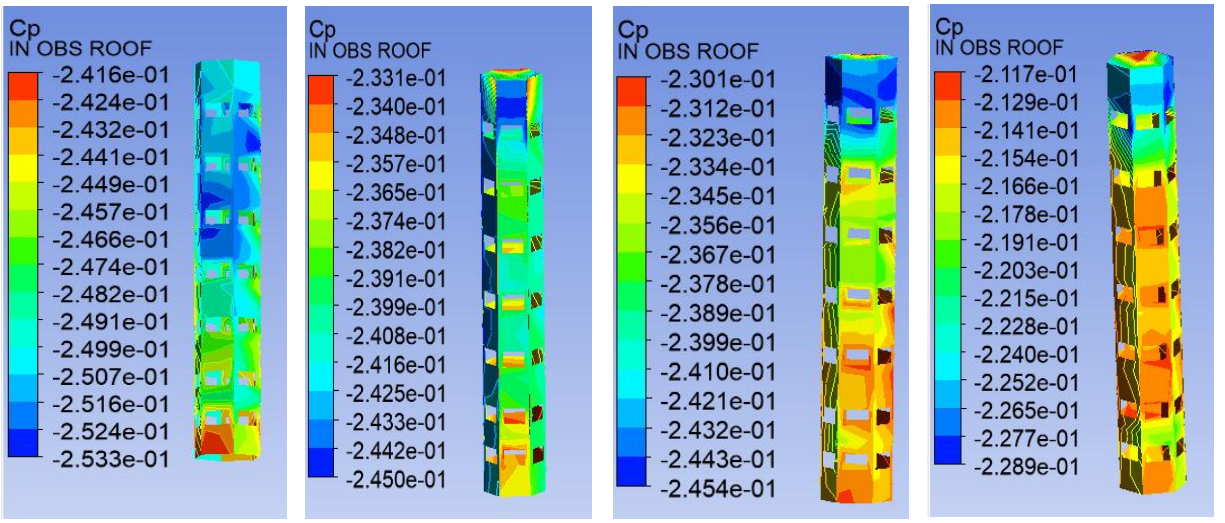



Figure 13: Contours of Internal Faces for different spacing in ascending order

**Streamlines**

Streamlines are a visual representation of wind particles and how they will flow after they are halted by a structure. This animation vividly depicts how the wind would travel around the structure. This provides a clear picture of the vortex creation and aids in understanding the variety in Cp values.
The streamline perspective of the isolated building is displayed, which demonstrates the creation of a vortex on the back side of the building, resulting in very little pressure on the backside. This explains why suction force forms on the Face B.

A vortex forms between the buildings, which multiplies as the distance between them climbs to 150 mm, causing the Cp value of face A to fall as the distance increases. Because the vortex generated at 200 mm is modest, the Cp value on Face A rises. An increase in the spacing and intensity of the vortex formation inside the observation building is seen by an increase in the internal pressure of the building.

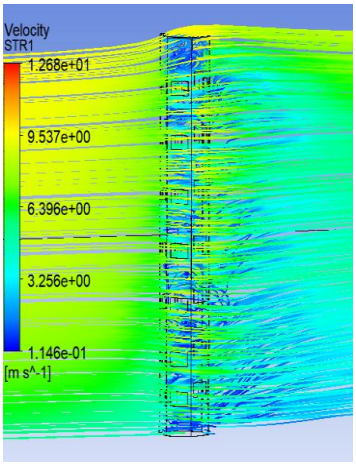



Figure 14 Streamline of Isolate Building

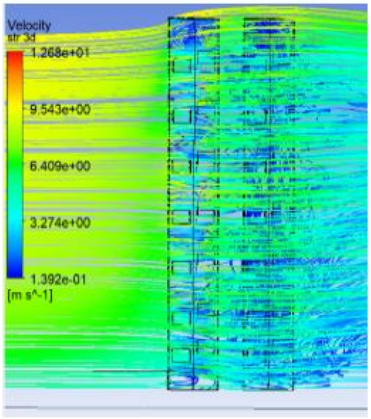



Figure 15: Streamline of 50mm spacing

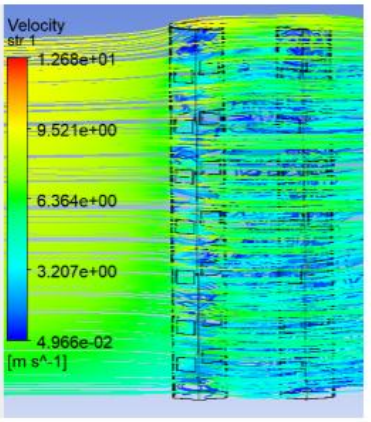



Figure 16: Streamline of 100mm spacing

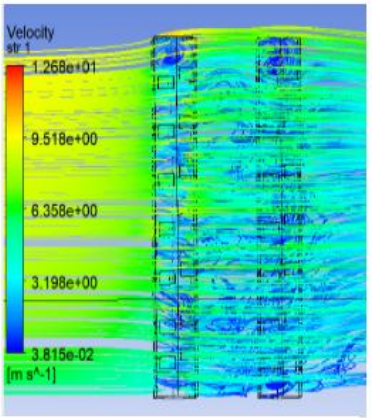



Figure 17: Streamline of 150mm spacing

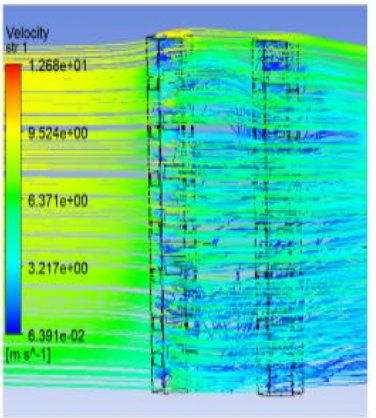



Figure 18: Streamline of 200mm spacing

**Results and Discussions**

The results that we found by studying various graphs and calculating the Cp values are mentioned below.

**Cp Values**

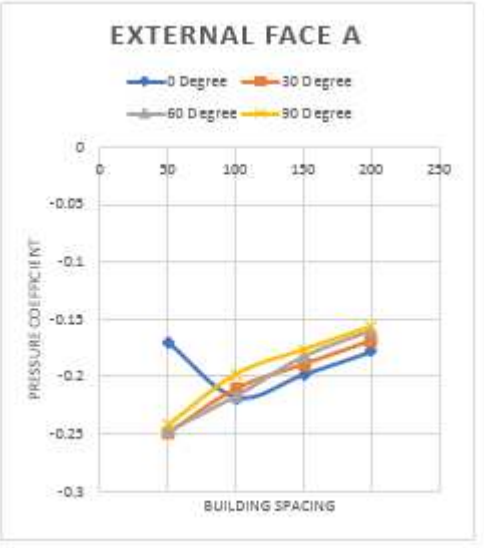
The Pressure coefficient values for each face of the instrumental building is calculated using the expression given below.

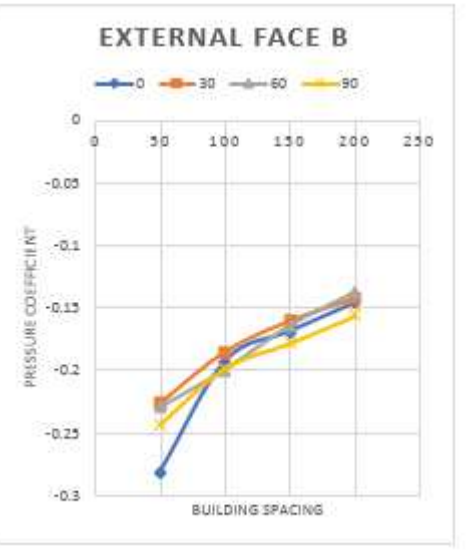
$Cp = Pressure/(0.5*rho*uRef^2)$

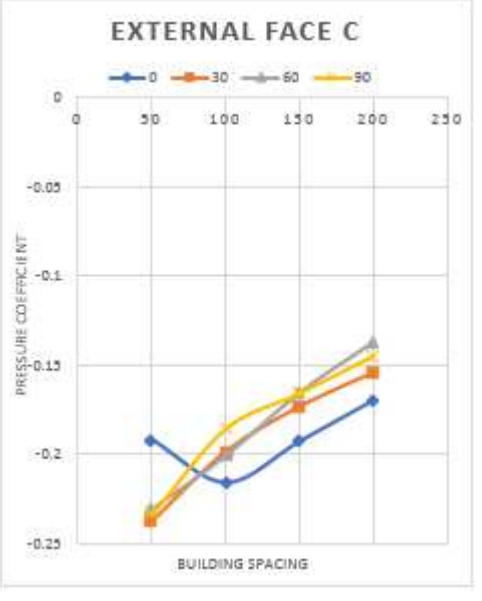
Where Pressure denotes the actual wind pressure, rho denotes air density, and uRef is the reference velocity at the building height. The external pressure coefficients, Cp (facial average value), for the model's many faces [8].
The graphs for each and every face is plotted for all its angle to see the variation on how the Cp values changes as the instrumental building rotates. The conclusions drawn from the graphs are mentioned below.

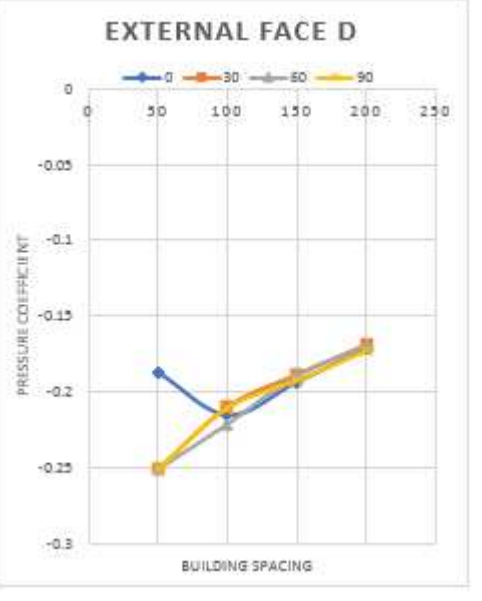
It can be observed that as the distance between the interference building and the observational building grows, so does the value of Cp, with the exception of Face A at 0 degree. This might be due to vortex development when the distance between the buildings increases from 50mm to 100mm. As the distance grows, the Cp becomes less negative as the vortex's impact diminishes. Because of the creation of a vortex behind the observational building, Face B has the largest negative Cp value in the situation of 0 degree at 50mm distance.

Faces C and D are next to face A and hence have graphs that are almost identical. This also implies that faces E and F will have the same trendline as face B. Because there is considerable vortex formation, suction force rises, and therefore the Cp values on Faces C and D become increasingly negative as the distance grows in the case of 0 degrees, but the converse happens on Faces E and F.


EXTERNAL FACE A
0 Degree
30 Degree
60 Degree
90 Degree
PRESSURE COEFFICIENT
BUILDING SPACING


EXTERNAL FACE B
0
30
60
90
PRESSURE COEFFICIENT
BUILDING SPACING


EXTERNAL FACE C
0
30
60
90
PRESSURE COEFFICIENT
BUILDING SPACING


EXTERNAL FACE D
0
30
60
90
PRESSURE COEFFICIENT
BUILDING SPACING

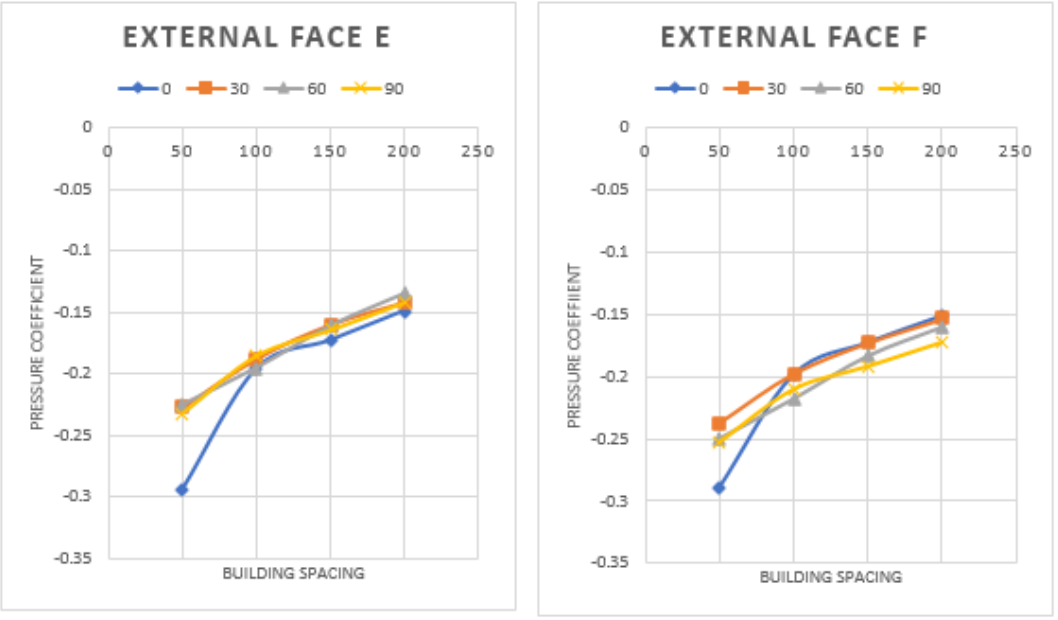
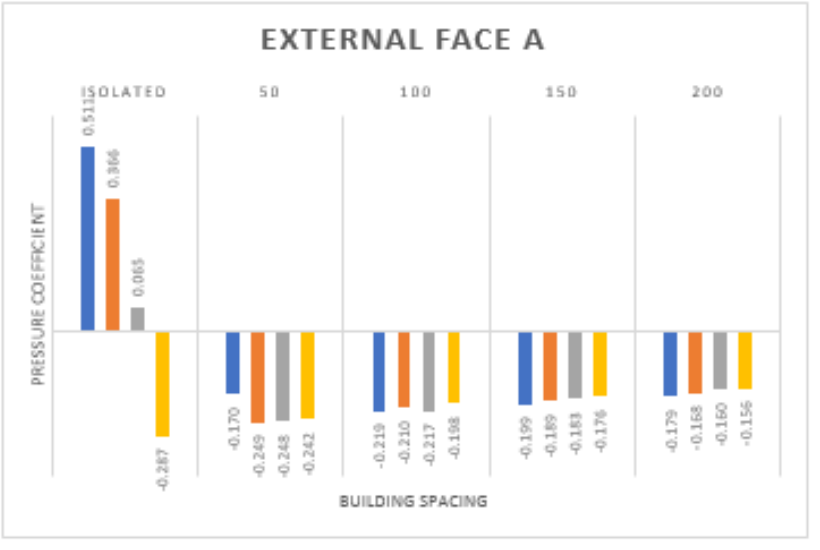
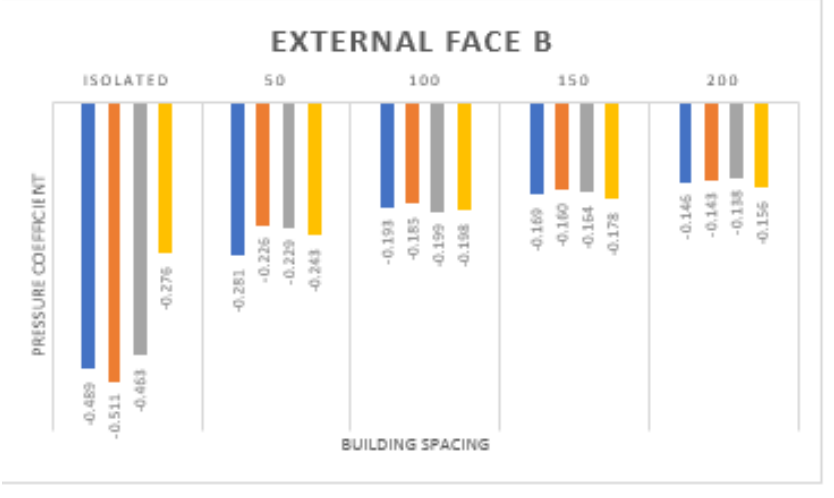
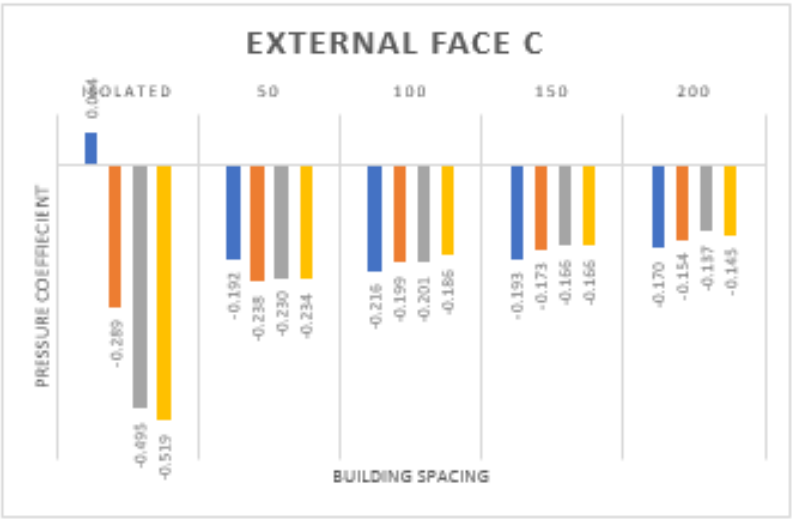


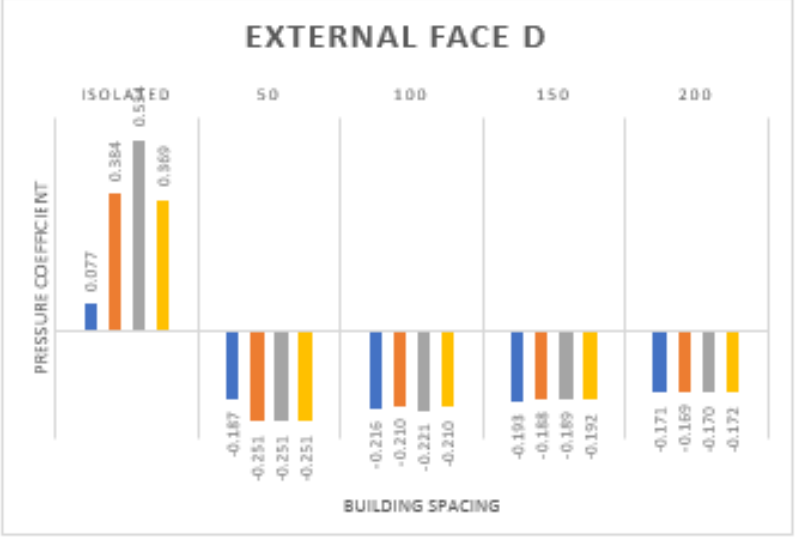

Figure 19 : Cp for Faces Comparison for Various Models Arranged at Different Angles

The graphs show the Cp values for each face in comparison to the isolated building. The graph of face A indicates that when the interference building is placed in front, the Cp values shift from positive to negative, indicating that the direction of force changes. The reduction in Cp values on Face B suggests that the interference building greatly reduces the pressure on both the windward and leeward sides.


EXTERNAL FACE A
ISOLATED
50
100
150
200
PRESSURE COEFFICIENT
0.512
0.366
0.065
-0.287
-0.170
-0.249
-0.248
-0.242
-0.219
-0.210
-0.217
-0.198
-0.199
-0.189
-0.183
-0.176
-0.179
-0.168
-0.160
-0.156
BUILDING SPACING
0
30
60
90


EXTERNAL FACE B
ISOLATED
50
100
150
200
PRESSURE COEFFICIENT
-0.489
-0.511
-0.463
-0.276
-0.281
-0.226
-0.229
-0.243
-0.193
-0.185
-0.199
-0.198
-0.169
-0.160
-0.164
-0.178
-0.146
-0.143
-0.138
-0.156
BUILDING SPACING


EXTERNAL FACE C
ISOLATED
50
100
150
200
PRESSURE COEFFICIENT
-0.289
-0.495
-0.519
-0.192
-0.238
-0.230
-0.234
-0.216
-0.199
-0.201
-0.186
-0.193
-0.173
-0.166
-0.166
-0.170
-0.154
-0.187
-0.145
BUILDING SPACING


EXTERNAL FACE D
ISOLATED
50
100
150
200
PRESSURE COEFFICIENT
0.077
0.384
0.969
-0.187
-0.251
-0.251
-0.251
-0.216
-0.210
-0.221
-0.210
-0.193
-0.188
-0.189
-0.192
-0.171
-0.169
-0.170
-0.172
BUILDING SPACING

0
30
60
90

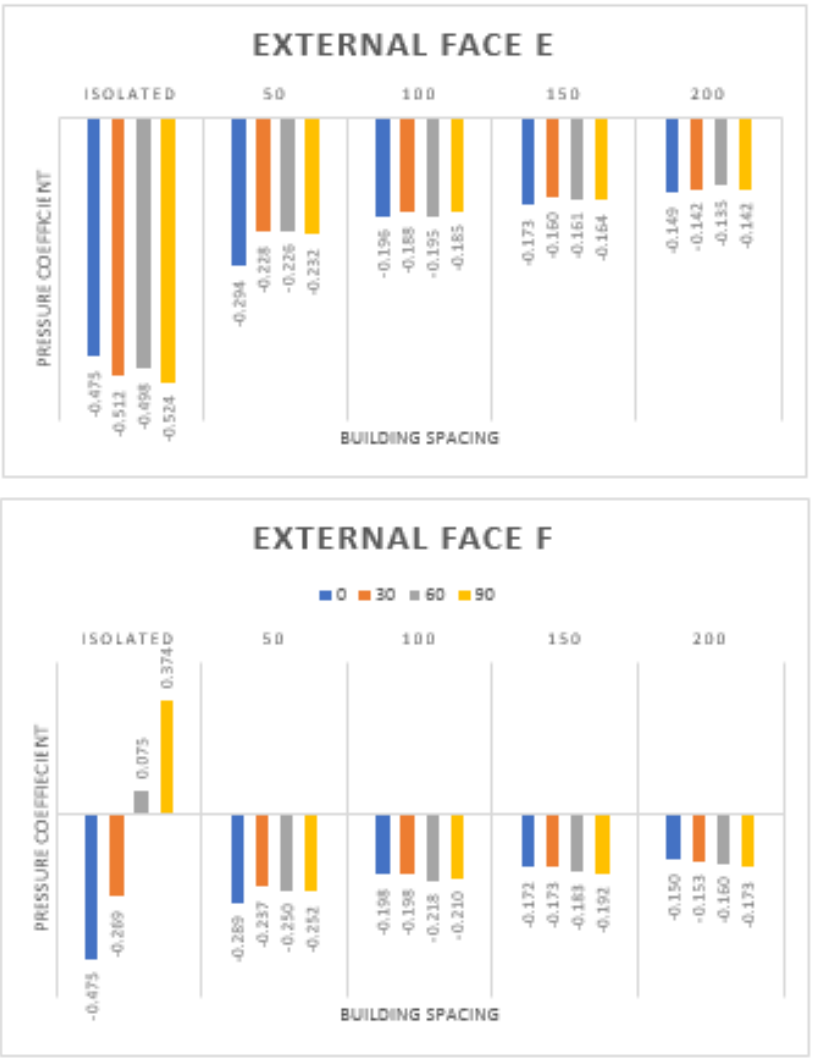



Figure 20: Cp values of the Isolated Building compared to other Interfering Models

**Interference Factor**

The interference effect is a metric that evaluates the strength of an obstruction in a fluid's path that prevents it from reaching the primary or test item. The more severe the interference, the higher the value of the interference factor. To investigate wind interference effects along wind direction, the force along the normal x-direction has been computed for all major models with varying configurations of interference construction sites and orientation. [9].

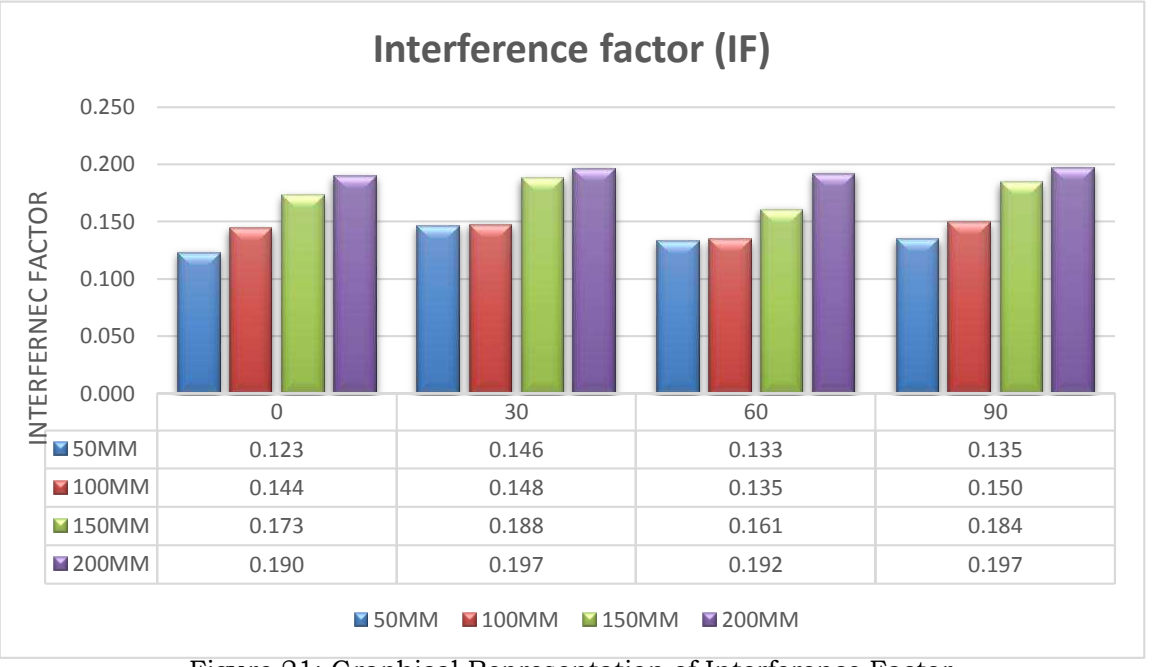



Figure 21: Graphical Representation of Interference Factor

- $\text{IF} = \frac{\text{Force on Instrumental Building in the presence of Interference Building}}{\text{Force on isolated Instrumental Building}}$

Graphing IF values for various distances and orientations is used to assess the interference influence on the instrumentation building. Values less than one imply that the observing building is shielded, whilst values greater than one indicate higher loading. [13].

It can be seen that all of the examples exhibit a shielding effect. The greater the value, the greater the shielding effect. Buildings with orientations of 0 and 60 degrees have a greater shielding effect than those with other angles. The first example, with a distance of 50mm between the buildings and a 0 degree orientation, has the most shielding impact, whereas the cases with a distance of 50mm between the buildings have the least shielding effect.

**Lift force and Drag force**

For studying the variation of wind effects due to different wind angles we also plotted the various graph for $C_{FX}$ and $C_{FY}$ whose equations are given below:

$$C_{FX} = \frac{F_X}{0.5\rho L_Y H U_H^2} \qquad C_{FY} = \frac{F_Y}{0.5\rho L_X H U_H^2}$$

- $C_{FX}$ and $C_{FY}$ are the force coefficients of the whole building along X and Y directions, respectively

- $F_X$ and $F_Y$ are the total forces over the building along the X- and Y-direction respectively
- ρ is the density of wind; $U_H$ is the reference velocity at the building height
- $L_X$ and $L_Y$ are the projected length of building along to the X and Y axes, respectively
- H is the height of the building [8].

Air resistance, often known as drag, is a force created by air that operates in the opposing direction of an item moving through it. It is the point at which air particles collide with the front of the item, slowing it down. The higher the surface area, the greater the number of air particles that impact it and the greater the resistance. The drag coefficient is proportional to the drag force exerted on the structures.

The drag coefficient is inversely related to the density of the fluid, i.e., water has a larger drag coefficient than air. The drag coefficient is inversely related to the fluid velocity. $C_{FX}$ stands for drag force. Any upward pressure given to a structure that has the ability to raise it relative to its surroundings is referred to as an uplift force. Pressure from the ground below, wind, surface water, and other variables can all create uplift forces. Lift force is indicated by the symbol $C_{FY}$ [14].
As can be observed, the greater the surface area, the greater the drag force, and hence the examples with angles 30 and 90 have higher drag force than the other two angles. Also, as the distance rises, the influence of the vortex decreases, and so the drag force increases.

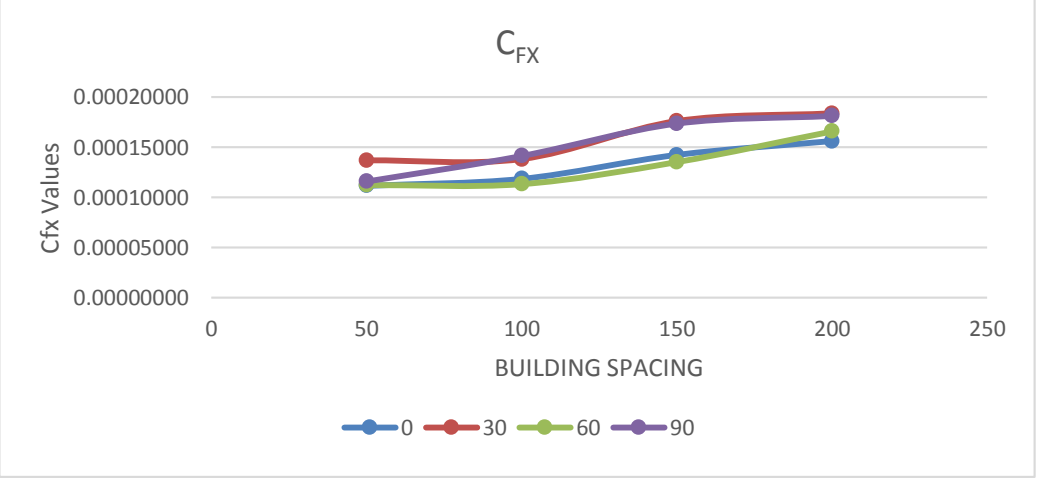



Figure 22.1: Graph of $C_{FX}$

Unlike $C_{FX}$, $C_{FY}$ shows the unusual trend. The maximum $C_{FY}$ is applied to a building with a 30 degree orientation, while the lowest is specified by a 60 degree orientation. Except for the 30 degree inclination, all of the trend lines follow a similar pattern. $C_{FY}$ rises in all orientations when we proceed from 50mm to 100mm spacing, except in 30 degrees direction due to the creation of a vortex. The maximum $C_{FY}$ force is applied to the case with 200mm spacing and a 30

degree orientation, while the lowest is applied to the case with 50mm spacing and a 60 degree orientation.

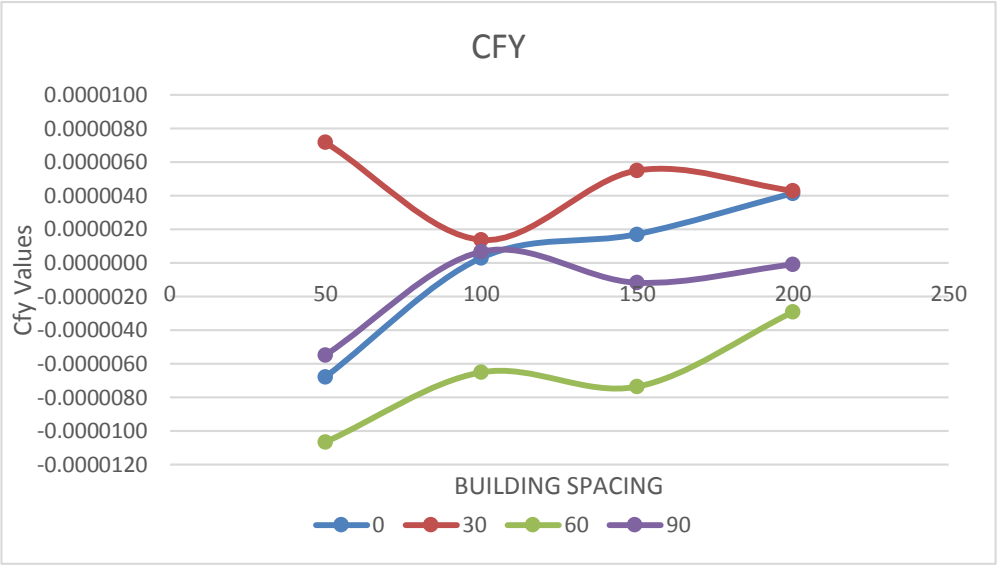



Figure 22.2: Graph of $C_{FY}$

**Conclusions**

The primary conclusions of this current study on the "hexagonal" plan shaped high structure with openings are as follows.:

- The internal pressure coefficient was discovered to vary as the model building was rotated based on the contours. The streamline diagrams also revealed that the majority of the air was trapped between the buildings owing to the holes.
- The streamline diagrams of the building show the formation of a vortex on the rear side of the structure, which grows in size as the distance between the interference and reference buildings grows.
- The influence of interference on the instrumentation building was assessed. Values less than one showed that the observing building was shielded, whilst values greater than one indicated higher loading.
- The highest shielding effect is seen when our instrumental building is positioned at 50mm and with a 0 degree angle.
- $C_{FX}$ has the same pattern as $C_{FY}$, but $C_{FY}$ exhibits an unusual tendency. When we increased the distance from 50mm to 100mm, $C_{FY}$ increased in all orientations except 30 degrees due to the formation of a vortex. The case with 200mm spacing and a 30 degree orientation produced the greatest $C_{FY}$ force, whereas the case with 50mm spacing and a 60 degree angle produced the least.

**References**


1. Gu M, Quan Y. Across-wind loads and effects of super-tall buildings and structures. Science China Technological Sciences. 2011 Oct;54(10):2531-41.
2. Kumar A, Raj R. Study of Pressure Distribution on an Irregular Octagonal Plan Oval-Shape Building Using CFD. Civil Engineering Journal. 2021 Oct 1;7(10):1787-805.
3. Bhattacharyya B, Dalui SK. Investigation of mean wind pressures on 'E'plan shaped tall building. Wind and structures. 2018 Feb 1;26(2):99-114.
4. Dagnew AK, Bitsuamalk GT, Merrick R. Computational evaluation of wind pressures on tall buildings. In11th American conference on Wind Engineering. San Juan, Puerto Rico 2009 Jun 22.
5. Bhattacharyya B, Dalui SK. Along and across wind effects on irregular plan shaped tall building. InAdvances in Structural Engineering 2015 (pp. 1445-1460). Springer, New Delhi.
6. "IS (Indian Standard). 2015. Indian standard code of practice for design wind load on building and structures. IS 875 (Part 3). New Delhi, India: IS."
7. Mukherjee S, Chakraborty S, Dalui SK, Ahuja AK. Wind induced pressure on' Y' plan shape tall building. Wind & structures. 2014;19(5):523-40.
8. Sanyal P, Dalui SK. Effects of side ratio for 'Y'plan shaped tall building under wind load. In Building Simulation 2021 Aug (Vol. 14, No. 4, pp. 1221-1236). Tsinghua University Press.
9. Pal S, Raj R. Evaluation of wind induced interference effects on shape remodeled tall buildings. Arabian Journal for Science and Engineering. 2021 Nov;46(11):11425-45.
10. Jones WP, Launder BE. The prediction of laminarization with a two-equation model of turbulence. International journal of heat and mass transfer. 1972 Feb 1;15(2):301-14.
11. Menter FR. Two-equation eddy-viscosity turbulence models for engineering applications. AIAA journal. 1994 Aug;32(8):1598-605.
12. Jensen G, Franke J, Hirsch C, Schatzmann M, Sathopoulos T, Wisse JA, Wright N. Impact of Wind and Storm on City and Built Environment–Working Group 2 CFD Techniques– Computational Wind Engineering. Inconference; COST Action C14, Impact of Wind and Storm on City Life and Built Environment; 2004-05-05; 2004-05-07 2004.
13. Pal S, Raj R, Anbukumar S. Bilateral interference of wind loads induced on duplicate building models of various shapes. Latin American Journal of Solids and Structures. 2021 Jul 26;18.
14. Lin N, Letchford C, Tamura Y, Liang B, Nakamura O. Characteristics of wind forces acting on tall buildings. Journal of Wind Engineering and Industrial Aerodynamics. 2005 Mar 1;93(3):217-42.