SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(123rd Edition)

MARCH 2023

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 Twenty Third** 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 **March, 2023**. 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

11. Aspect and orientation-based sentiment analysis of customer feedback using mathematical optimizationmodels, ***6.Neha Punetha*** and ***3.Goonjan Jain***, Applied Mathematics, DTU

12. Assessment of groundwater quality and human health risks of nitrate and fluoride contamination in a rapidly urbanizing region of India, ***6.Riki Sarma*** and ***3.Santosh Kumar Singh*** Environmental, DTU

13. A Survey on Smart Parking Management System, ***7.Rohit Ramchandani*** and ***6.Anjali Bansal***, CSE, DTU

14. Blockchain Driven Access control architecture for the internet of things, ***6.Rajiv K. Mishra***, ***3.Rajesh K. Yadav*** and Prem Nath, CSE, DTU

15. Computational optimization of engine performance and emission responses for dual fuel CI engine powered with biogas and Co3O4 nanoparticles doped biodiesel, ***7.S. Lalhriatpuia*** and ***3.Amit Pal***, Mechanical, DTU

16. Deformation response of Twin Tunnels under the effect of static loading conditions, ***6.Parvesh Kumar*** and ***3.Amit Kumar Shrivastava***, Civil, DTU

17. Design and Computational Analysis of an MMP9 Inhibitor in Hypoxia-Induced Glioblastoma Multiforme, ***6.Smita Kumari*** and ***3.Pravir Kumar***, Biotechnology, DTU

18. Doped graphene characterized via Raman spectroscopy and magneto-transport measurements, Marie-Luise Braatz, Nils-Eike Weber, ***3.Barthi Singh***, Klaus Müllen, Xinliang Feng, Mathias Kläui and Martin Gradhand, Applied Physics, DTU

19. Dynamic Combined Economic Emission Load Dispatch using Perfectly Convergent Particle Swarm Optimization, Devinder Kumar, ***3.Narender kumar Jain*** and ***3.Nangia Uma***, Electrical, DTU

20. Effect of air-fuel ratio and pressure ratio on the exergetic performance of combined cycle gas turbine plant components, ***7.Sandeep Kumar*** and ***6.Ashutosh Mishra***, Mechanical, DTU

21. Effect of the inlet-to-outlet key width ratio of Piano Key Weir on its hydraulic behavior, ***6.Deepak Singh*** and ***3.Munendra Kumar***, Civil, DTU

22. Effectual seizure detection using MBBF-GPSO with CNN network, ***6.Dinesh Kumar Atal*** and ***3.Mukhtiar Singh***, Electrical, DTU

23. Experimental and Simulation Study of the Latest HFC/HFO and Blend of Refrigerants in Vapour Compression Refrigeration System as an Alternative of R134a, Uma Shankar Prasad, ***3.Radhey Shyam Mishra***, Ranadip Kumar Das and Hargovind Soni, Mechanical, DTU

24. Exploring Vision Transformer model for detecting Lithography Hotspots, ***6.Sumedha*** and ***3.Rajesh Rohilla***, Electronics, DTU

25. Measuring Influence of Indices in DN Planning, ***8.Shubham Gupta***, ***3.Vinod Kumar Yadav*** and ***3.Madhusudan Singh***, Electrical, DTU

26. Monitoring and sensing of glucose molecule by micropillar coated electrochemical biosensor via CuO/[Fe(CN)6]3 and its applications, ***7.Purva Duhan***, ***7.Deepak Kumar***, ***7.Mukta Sharma***, ***3.Deenan Santhiya*** and ***3.Vinod Singh***, Applied Chemistry and Applied Physics, DTU

27. Multi-view Multi-modal Approach Based on 5S-CNN and BiLSTM Using Skeleton, Depth and RGB Data for Human Activity Recognition, ***8.Rahul Kumar*** and ***3.Shailender Kumar***, CSE, DTU

28. Numerical modeling of a dielectric modulated surrounding-triple-gate germanium-source MOSFET (DM-STGGS-MOSFET)-based biosensor, Amit Das, ***3.Sonam Rewari***, Binod Kumar Kanaujia, S. S. Deswal and R. S. Gupta, Electronics and Communication, DTU

29. Optimization of Biodiesel Parameters Using Response Surface Methodology and Production of Biodiesel, ***Y. K. Singh***, Biotechnology, DTU

30. Optimized ensemble-classification for prediction of soil liquefaction with improved features, ***6.Nerusupalli Dinesh Kumar Reddy***, ***3.Ashok Kumar Gupta*** and ***3.Anil Kumar Sahu***, Civil, DTU

31. OptNet-Fake: Fake News Detection in Socio-Cyber Platforms Using Grasshopper Optimization and Deep Neural Network, ***3.Sanjay Kumar***, Akshi Kumar, ***8.Abhishek Mallik***, and Rishi Ranjan Singh, CSE, DTU

32. Perfectly Convergent Particle Swarm Optimization for Solving Combined Economic Emission Dispatch Problems with and without Valve Loading Effects, Devinder Kumar, ***3.Narender kumar Jain*** and ***3.Nangia Uma***, Electrical, DTU

33. Performance Analysis of Solar PV Modules with Dust Accumulation for Indian Scenario, ***7.Komal Singh*** and ***3.M. Rizwan***, Electrical, DTU

34. Probabilistic intuitionistic fuzzy c-means algorithm with spatial constraint for human brain MRI segmentation, ***6.Rinki Solanki*** and ***3.Dhirendra Kumar***, Applied Mathematics, DTU

35. Reversible data hiding with high visual quality using pairwise PVO and PEE, Neeraj Kumar, ***3.Rajeev Kumar***, Aruna Malik, Samayveer Singh & Ki-Hyun Jung, CSE, DTU

36. Scheduling of Energy Storage System (ESS) for Electricity Distribution Companies (DISCOMs), ***6.Chetan Gusain***, ***3.Madan Mohan Tripathi*** and ***3.Uma Nangia***, Electrical, DTU

37. Settlement in Geosynthetic Reinforced Square Footing over Plastic Soil, Ankur Mudgal, Bibek Jha, ***3.Raju Sarkar***, ***3.Amit Kumar Srivastava***, Akshit Mittal and Nehal Jain, Civil, DTU

38. Stability Analysis of Rainfall-Induced Landslide Using Numerical Modelling, ***7.Akash Bhardwaj*** and ***3.Amit Kumar Shrivastava***, Civil, DTU

39. Structural, thermal, and luminescence kinetics of Sr4Nb2O9 phosphor doped with Dy3+ ions for cool w-LED applications, Ravina Lohan, A. Kumar, Mukesh K. Sahu, ***6.Anu Mor***, V. Kumar, Nisha Deopa and ***3.A. S. Rao***, Applied Physics, DTU

40. Sustainable Green Approach of Silica Nanoparticle Synthesis Using an Agro-waste Rice Husk, Mikhlesh Kumari, Kulbir Singh, Paramjeet Dhull, Rajesh Kumar Lohchab and ***3.A. K. Haritash***, Environmental, DTU

41. Temporal variation and source identification of carbonaceous aerosols in Monrovia, Liberia, ***7.Emmanuel Juah Dunbar*** and ***3.Lovleen Gupta***, Environmental, DTU

42. The Awkward World of Python and C++, ***8.Manasvi Goyal***, Ianna Osborne and Jim Pivarski, Production and Industrial, DTU

43. Thermal-hydraulic Behaviour of Corrugated Pipe Configurations, ***8.Aryan Tyagi***, ***6.Gaurav Kumar*** and ***3.Raj Kumar Singh***, Mechanical, DTU

44. Using tropical cyclone characteristics and considering local factors, the radius of maximum wind over the North Indian Basin is evaluated, ***6.Monu Yadav*** and ***3.Laxminarayan Das***, Applied Mathematics, 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*

A Comparative Study of CMOS Transimpedance Amplifier (TIA)

Priya Singh, Vandana Niranjan, Ashwni Kumar

Abstract: In this paper a comparative study of different CMOS transimpedance amplifier has been presented. Standard device parameters of transimpedance amplifier such as gain, input refereed noise, power dissipation and group delay are studied and compared. Here the transimpedance amplifier is divided on the basis of its topology and device technology used and performance is summarized to get the overview. Most of the analysis taken are performed on 0.18 µm technology and some are implemented using 45nm, 0.13µm, 65nm, and 90nm.

Keywords: CMOS, Transimpedance, Amplifier (TIA), Technology

I. INTRODUCTION

In today's intelligence technology various sensors are used to convey the information signal. Most of the sensors convert this information containing signal to current signal but these signals cannot be further processed since most of the subsequent blocks in a system work in voltage mode. Therefore, it is mandatory to convert this current signal to amplified voltage signal. This conversion can be carried out using transimpedance amplifier (TIA). Transimpedance amplifier can be designed using active and passive circuit elements. Passive circuit elements have many limitations in VLSI designing such as larger chip area, higher power dissipation and so on thus active elements are used more often such as BJT, MOSFET or an operational amplifier. While designing a Practical Transimpedance Amplifier Main Characteristics of concern are low input impedance, low output impedance, high transimpedance gain, high bandwidth and low noise. In this paper main concern is on CMOS transimpedance amplifier. Further section II presents different topologies of CMOS transimpedance amplifier and section III includes comparative study analysis of CMOS existing transimpedance amplifier on the basis of topology and device technology used.

Manuscript received on 12 January 2023 | Revised Manuscript received on 25 February 2023 | Manuscript Accepted on 15 March 2023 | Manuscript published on 30 March 2023.

* Correspondence Author

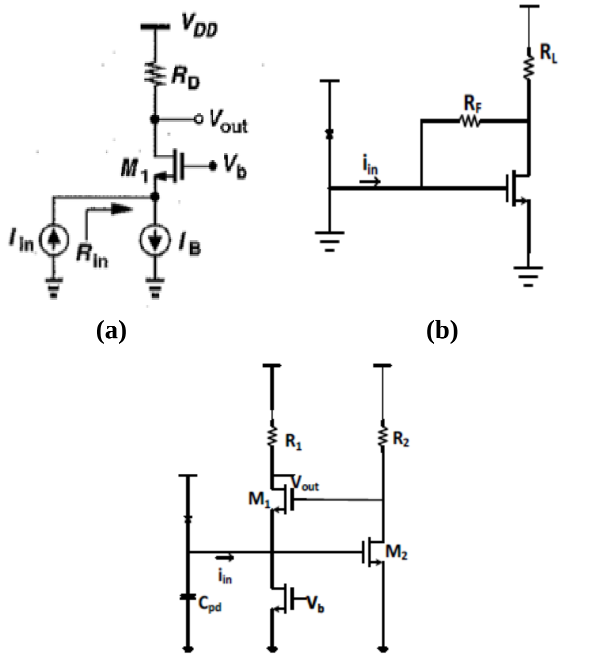
Priya Singh*, Research Scholar, Indira Gandhi Delhi Technical University for Women, Kashmeri Gate, Delhi, India. E-mail: priya004phd17@igdtuw.ac.in, ORCID ID: https://orcid.org/0000-0003-0716-764X

Dr. Vandana Niranjan, Professor, Indira Gandhi Delhi Technical University for Women, Kashmeri Gate, Delhi, India. E-mail: vandana_niranjan@yahoo.com

Prof. Ashwni Kumar, Professor, Indira Gandhi Delhi Technical University for Women, Kashmeri Gate, Delhi, India. E-mail: drashwnikumar@gmail.com

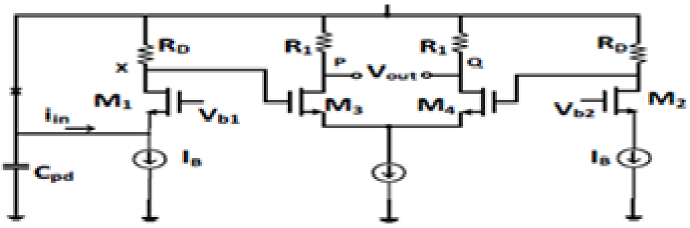
II. TOPOLOGIES OF CMOS TRANSIMPEDANCE AMPLIFIER

A number of research papers published related to CMOS transimpedance amplifier were studied. [1-40] Based on the literature review most of the reported CMOS transimpedance amplifier can be broadly classified in to four types- common gate topology, common source topology, regulated cascode topology and differential topology. Most of the transimpedance amplifier used these basic topologies or hybridized form are used for further improvement. In common gate topology input impedance is low as compared to other topology due to proper bias current. Due to this high bandwidth can be achieved. Noise is the major issue of common gate topology when used in low supply voltage. Common source topology have high driving capability but as the device size minimize transistors breakdown voltage also reduces. Due to this tradeoff amid transimpedance gain and voltage headroom is very crucial for common source topology. Effect of parasitic capacitance on bandwidth can be reduced by cascode structure in regulated cascode topology but due to severe trade off in all characteristics more improvement in circuit is required. Differential structure results in less offset noise and cannot be used alone due to deferential input a converter is always needed that makes circuit more complex.



(a) **(b)**

(c)



(d)

Figure 1. (a) Common gate TIA topology [36] (b) Common source TIA topology [13] (c) Regulated Cascode TIA topology [33] (d) Differential TIA topology [36]

III. COMPARATIVE STUDY OF CMOS TIA

Comparison of some of the transimpedance amplifier designed with best characteristic achieved so far is refereed in table 1 in ascending order of publication year. Best result obtained among them for each characteristics are highlighted. In table 2 the latest transimpedance amplifier based on device technology used are compared. CMOS technology was always in lime light due to its low cost, low power dissipation and integrity property, but due to its parasitic properties and degraded noise performance design of TIA becomes challenging. Various research gaps had been identified based on this study. Common gate topology of CMOS TIA cannot work at low supply voltage efficiently because of degraded noise performance. Nano scale application cannot use common source TIA due to less voltage headroom thus high bit error rate. Regulated cascode topology provides best isolation from parasitic capacitance but severe tradeoff between gain and bandwidth reduces the efficiency.

Table 1. Comparison of Various TIA Designed

Ref. No.	Year	Topology /technique	Power supply (volts)	Gain D Bohms	BW	Input Referred Noise (pA/√Hz)	Power dissipation M W
[8]	1991	GaAs FET	-	-	100 MHz	-	-
[12]	1999	C peaking technique	-	0.95	2.3 GHz	-	-
[14]	2002	hybrid topology	-	20	1 Ghz	-	27
[9]	2004	Regulated cascade	5	58	950 MHz	6.3	85
[21]	2004	Interstage matching network technique	2.5	54	9.2 GHz	-	-
[10]	2006	Regulated cascade	2	52	7.6GHz		34
[27]	2007	Regulated cascade	1.8	53	8GHz	18	13.5
[5]	2007	Common base-BiCMOS		65.2	7.2GHz	17.7	56.6
[29]	2010	Common gate with active feedback	1.8	54.6	7GHz	17.5	18.6
[32]	2012	Regulated Cascode		52	35GHz	14	-
[34]	2012	SOI technology	-	55	33GHz	20.47	-
[39]	2014	Common gate	-	104.2	19MHz	-	-
[19]	2015	Differential		87.8	1.4GHz	-	8.1
[3]	2015	Regulated Cascode	3.3	61	15GHz	-	32
[24]	2015	Regulated cascade	-	-	9GHZ	-	-
[18]	2015	Regulated cascade	1.5	50	7GHz	31	-
[17]	2016	Differential with Regulated Cascode input stage	1.8	18	870 MHz	6	27
[15]	2017	Common gate using FGMOS	1	37.7	13.5GHz	-	1.1
[40]	2018	BiCMOS tech.	-	67	28GHz	10	95

Table 2. Performance Comparison of Latest Existing TIA Using Different Device Technology

Technology And devices	Paper no.	Bandwidth (GHz)	Input referred noise (pA/√Hz)	Gain (Z_T) dBΩ	Power Dissipation mw
Bi CMOS 130 nm	IEICE transaction 2018[40]	28	10	67	95
FGMOS 0.18µm	Integration: the VLSI journal: ELSEVIER 2018[15]	13.5	-	37.7	1.1
CMOS 65nm	ISCAS 2012[35]	35	14	52	168
SOI CMOS 45nm	JSSC 2012[34]	33	20.47	55	9

IV. CONCLUSION

Apart from topology device technology is playing major issue in improving the characteristics especially when low voltage and low power is concerned. FGMOS as active feedback provides best result in terms of low power supply and power dissipation. BiCMOS gives most consistent result in terms of all characteristics. All topologies are improvised and further improvement can be done to attain the best characteristics by introducing several techniques of bandwidth enhancement and hybridizing these topologies to take the advantage of each. These improved TIA can enhance the performance of intelligent system and application in receiving and transmission of various information carrying signal.

FUTURE WORK

Design of transimpedance amplifier necessitates efficient tradeoff and improvisation of each characteristic. Low voltage circuit techniques such as FGMOS, QFGMOS, bulk driven, DTMOS etc can be used for low voltage application and reducing the power consumption. Various noise reduction techniques could be used in the TIA circuit to optimize the noise performance. Further cascading of different stages will improve the gain as well as improve the parasitic capacitance effect to improve the bandwidth requirement.

DECLARATION

Funding/ Grants/ Financial Support	No, I did not receive.
Conflicts of Interest/ Competing Interests	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material/ Data Access Statement	Not relevant.
Authors Contributions	All authors have contributed equally.

REFERENCE

1. X Huang, Cha, D Zhao, B Guo, M Je, H Yu "Transimpedance Amplifier for Integrated 3D Ultrasound Biomicroscope Applications", , *World Academy of Science, Engineering and Technology International Journal of Electronics and Communication Engineering* , Vol:6, No:9, 2012
2. Praveen Dwivedi, Amit Kumar Singh and R.G. Sangeetha "A Comparative Analysis of Transimpedance Amplifier in Giga-bit Optical Communication "*Research Journal of Engineering Sciences* ISSN 2278 – 9472 Vol. 3(3), 6-9, March (2014)
3. Q Song, LMao "Wideband sige BICMOS transimpedance amplifier foe 20 Gb/s optical links", *IEICE electronics express*, Vol. 12, No. 13, july 2015 , 1-8 [CrossRef]
4. T. Ridder, P. Ossieur, X. Yin, B. Baekelandt, C. Me´lange, J. Bauwelinck, X.Z. Qiu and J. Vandewege , "Bicmos variable gain transimpedance amplifier for automotive applications" *The Institution of Engineering and Technology* 2008 28 October 2007 Electronics Letters online no: 20083101 [CrossRef]
5. F. Touati and M. Loulou, "High-Performance bicmos Transimpedanc Amplifiers for Fiber-Optic Receivers", *The Journal of Engineering Research* Vol. 4, No.1 (2007) 69-74 [CrossRef]
6. Douglas Bespalko ,"Transimpedance Amplifier Design using 0.18 µm CMOS Technology" ,*thesis Queen's University Kingston*, Ontario, canadajuly 2007
7. Taghavi, , Belostotski, James W. Haslett, Ahmadi , "10-Gb/s 0.13-µm CMOS Inductorless Modified-RGC Transimpedance Amplifier" , *IEEE Transactions On Circuits And Systems*—I: Regular Papers, Vol. 62, No. 8, August 2015 [CrossRef]
8. N Scheinberg, , R.J. Bayruns, and T.M. Laverick , "Monolithic GaAs Transimpedance Amplifiers for Fiber-optic Receivers", *IEEE JOURNAL OF SOLID-STATE CIRCUITS*, VOL. 26, NO. 12, DECEMBER 1991 [CrossRef]
9. M.Park, , and H.J Yoo, "1.25-Gb/s Regulated cascodecmostransimpedance Amplifier for Gigabit Ethernet Applications ", *IEEE Journal Of Solid-State Circuits*, Vol. 39, No. 1, January 2004 [CrossRef]
10. H.Y.Hwang, J.C. Chien , "A CMOS Tunable Transimpedance Amplifier" , *IEEE Microwave And Wireless Components Letters*, Vol. 16, No. 12, December 2006 [CrossRef]
11. S Goswami, J Silver, T Copani, W Chen, Hugh J. Barnaby, Bert Vermeire, Sayfe Kiaei , "A 14mw 5Gb/s CMOS TIA with Gain-Reuse Regulated Cascode Compensation for Parallel Optical Interconnects" , *IEEE International Solid-State Circuits Conference,* 2009 [CrossRef]
12. F.T. Chien and Y.C.Chan , "Bandwidth Enhancement of Transimpedance Amplifier by a Capacitive-Peaking Design ", *IEEE Journal Of Solid-State Circuits*, Vol. 34, No. 8, August 1999 [CrossRef]
13. M. Ahmed, "Transimpedance Amplifier (TIA) Design for 400 Gb/s Optical Fiber Communications " . *thesis for M.S ,Virginia Polytechnic Institute and State University*, May 02, 2013 Blacksburg, VA
14. J Lee, S.J. Song, S. Park, C. Nam1,Y. Kwon and H. Yoo , "A Multichip on Oxide of 1Gb/s 80db Fully- Differential CMOS Transimpedance Amplifier for Optical Interconnect Applications", ISSCC 2002 / SESSION 4 / BACKPLANE INTERCONNECTED Ics / 4.7
15. U. Bansal, M. Gupta , "High bandwidth transimpedance amplifier using FGMOS for low voltage operation", *INTEGRATION, the VLSI journal, ELSEVEIR*, 2017 [CrossRef]
16. Claudio Talarico · G. Agrawal ·J. -Roveda · H. ,Design "Optimization of a Transimpedance Amplifier for a Fiber Optic" *Springer Science+Business Media* New York 2015 [CrossRef]
17. G. Royo, C. Sánchez-Azqueta, C. Aldea, and S. Celma , "CMOS Transimpedance Amplifier with Controllable Gain for RF Overlay", *2016 IEEE journal.* [CrossRef]
18. M. Seifouri n, parvizamiri,majidrakide ,"Design of broadband transimpedance amplifier for optical communication systems" , 2015 *ELSEVIER- optik LTD* [CrossRef]
19. Liu, Jiao Zo, N. Maa, Z. Zhu, Y. Yang , "A CMOS transimpedance amplifier with high gain and wide dynamicrange for optical fiber sensing system ", *Optik 126 (2015) 1389–1393 , ELSEVIER* [CrossRef]
20. Z. Lu, K. Seng Yeo, J. Ma, , M. Do, , Wei Meng Lim, Xueying Chen "Broad-Band Design Techniques for transimpedance amplifier", *IEEE Transactions On Circuits And Systems*: Regular Papers, Vol. 54, No. 3,March 2007 [CrossRef]
21. B. Analui, and A. Hajimiri, "Bandwidth Enhancement for Transimpedance Amplifiers" , *IEEE Journal Of Solid-State Circuits,* Vol. 39, No. 8, August 2004 [CrossRef]
22. Z Lu, Kiat S Yeo, W Lim, M Do "Gm-boosted differential transimpedance amplifier architecture" *IEICE Electronics Express · August 2007*
23. Z Lu, K Seng Yeo, W Lim, M Do "Design of a CMOS broadband transimpedance Amplifier With Active Feedback", *IEEE Transactions On Very Large Scale Integration (VLSI) Systems*, Vol. 18, No. 3, March 2010 [CrossRef]
24. T.-C. Chen, C Chan, and RSheen "Transimpedance Limit Exploration and Inductor-Less Bandwidth Extension for Designing Wideband Amplifiers" *IEEE Transactions On Very Large Scale Integration (VLSI) Systems*, Syst. 24.1 (2016) 348–352. [CrossRef]
25. B. Analui and A. Hajimiri, "Bandwidth Enhancement for Transimpedance Amplifiers*",* IEEE Journal of Solid-State Circuits, vol. 39, no. 8,pp. 1263-1270, August 2004. [CrossRef]
26. C.T. Chan and O. T. C. Chen, "Inductor-less 10Gb/s CMOS Transimpedance Amplifier Using Source-follower Regulated Cascode and Double Three-order Active Feedback", International Symposium on Circuits and Systems (ISCAS), pp.-5487-5490, May 2006

27. Z. Lu, K. S. Yeo and J. Ma, "Broad-Band Design Techniques for Transimpedance Amplifiers", *IEEE Transactions on Circuits and Systems—I: Regular Papers*, vol. 54, no. 3, pp.- 590– 600, March 2007. [CrossRef]
28. T. H. Ngo, T. W. Lee and H. H. Park, "Design of Transimpedance Amplifier for Optical Receivers in 0.13 µm CMOS", *International Conference on Optical Internet (COIN)*, pp.- 1-3, July 2010.
29. Z. Lu, K. S. Yeo, W. M. Lim and M. A. Do, "Design of a CMOS Broadband Transimpedance Amplifier With Active Feedback*", IEEE Transactions on VLSI Systems*, vol. 18, no. 3, pp.- 1964 – 1972, March 2010. [CrossRef]
30. J. D. Jin and Shawn S. H. Hsu , "A 40-Gb/s Transimpedance Amplifier in 0.18-µm CMOS Technology" , *IEEE Journal of Solid-State Circuits*, vol. 43, no. 6, pp. 520-5123, June 2008. [CrossRef]
31. C. F. Liao and S. L. Liu, "40 Gb/s Transimpedance-AGC Amplifier and CDR Circuit for Broadband Data Receivers in 90 nm CMOS", *IEEE Journal of Solid-State Circuits*, vol. 43, no. 3, pp. 642 - 655 March 2008. [CrossRef]
32. J. Kim and J. F. Buckwalter, "Bandwidth Enhancement with Low Group-Delay Variation for a 40-Gb/s Transimpedance Amplifier", *IEEE Transactions on Circuits and Systems—I: Regular Papers,* vol. 57, no. 8, pp.- 1964 – 1972, August 2010. [CrossRef]
33. S. Bashiri, C. Plett, J. Aguirre and P. Schvan "A 40 Gb/s Transimpedance Amplifier in 65 nm CMOS", *International Symposium on Circuits and Systems (ISCAS)*, pp- 757-760, May 2010. [CrossRef]
34. J. Kim and J. F. Buckwalter "A 40-Gb/s Optical Transceiver Front-End in 45 nm SOI CMOS", *IEEE Journal of Solid-State Circuits*, vol. 47, no. 3, pp. 1-4, March 2012. [CrossRef]
35. S. T. Chou, S. H. Huang, Z. H. Hong, and W. Z. Chen "A 40 Gbps Optical Receiver Analog Front-End in 65 nm CMOS", *International Symposium on Circuits and Systems (ISCAS)*, pp.-1736-1739, May 2012.
36. Behzad Razavi-"Design of Integrated Circuits for Optical Communications".
37. S. H. Huang, W.Z. Chen, Y.W. Chang, and Y.T. Huang, "A 10-Gbps CMOS Single Chip Optical Receiver with 2-D Meshed Spatially-Modulated Light Detector," IEEE J. Solid-State Circuits , pp. 1158-1169, vol. 46, NO. 5, May, 2011. [CrossRef]
38. M jalai, Mohammed k. , "Gm boosted differential transimpedance amplifier architecture", *Electronics Express, IEICE, 2007* [CrossRef]
39. A Chaddad, C Tanougast, "low- noise amplifier dedicated to biomedical devices", *IEEE conference* , 2014
40. X LUO, Y Chang, "A44Gbit/sec wide dynamic range and high linearity transimpedance amplifier 130nm bicmos technology", *IEICE TRANSACTIONS on Fundamentals of Electronics Communications and Computer Sciences* Vol.E101-A No.2 pp.438-440, feb 2018 [CrossRef]

AUTHORS PROFILE

Priya Singh had done B.E from Gyan Ganga College of Technology Jabalpur in electronics and communication and M.tech from Banasthali University Jaipur, in VLSI Design. She have teaching experience in Rajasthan Technical University and Abdul Kalam Technical University formerly known as UPTU. She had served as registered evaluator of the U.P state technical university for two years and had worked with Bhabha Atomic Research Centre, Mumbai as project intern in microelectronics domain for one year. She had given many publications in the field of hardware implementation of adhoc protocol in international conference and journals. She had worked in the field of ASIC designing and channel implementation using FPGA and published few research in international journals. Presently she is perusing Phd in analog IC designing from Indira Gandhi Delhi technical university for women, Delhi.

Dr. Vandana Niranjan is working as an Professor in Department of Electronics and Communication Engineering at Indira Gandhi Delhi Technical University Delhi. She received her B.E. degree in Electronics and Communication Engineering in the year 2000 from Government Engineering College (now University Institute of Technology of Rajiv Gandhi Proudyogiki Vishwavidyalaya) Bhopal. She received her M. Tech degree in the year 2002 from the Department of Electronics and Communication Engineering at Indian Institute of Technology Roorkee with VLSI Design as specialization. She pursued her research interest and was awarded her Ph.D degree in the area of Low Voltage VLSI Design from University School of Engineering & Technology, GGSIP University Delhi in the year 2015. She has a teaching and research experience of approximately 15 years at Indira Gandhi Delhi Technical University Delhi. She is member of I.E.T.E India, Institution of Engineers India and IEEE & Women in Engineering USA. She has published over 15 papers in reputed International Journals. She has also published over 35 papers in IEEE International/national conferences. *(https://www.researchgate.net/profile/Dr_Vandana_Niranjan/publications)*

Prof. Ashwni Kumar did his Ph.D (ECE) from Delhi University, M.E. (ECE) from Delhi College of Engineering, B.E. (ECE) from Delhi College of Engineering and MBA from FMS, Delhi University. He is at IGDTUW from the last two years. He has 16 Years of Research and Development experience in C-DOT, Research & development Center of Ministry of communication, Govt. of India.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Lattice Science Publication (LSP)/ journal and/ or the editor(s). The Lattice Science Publication (LSP)/ journal and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT) | 978-1-6654-7451-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/IDCIoT56793.2023.10053428

A Review of Convolutional Neural Network-based Approaches for Disease Detection in Plants

1st Barsha Biswas
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
barshabiswas2599@gmail.com

2nd Rajesh Kumar Yadav
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
rkyadav@dtu.ac.in

***Abstract*—Around 60.3% of land in India is used for agricultural purposes and the whole population depends on agriculture. That's why crop yield is very crucial to get high agricultural output. The economical loss will be very high if the agricultural output is low. So, that's why the diagnosis of disease in plants is very important. And the detection should be in the early stage not in a later stage. Using Deep Learning (DL) i.e. a branch of Artificial Intelligence (AI), a farmer can detect plant diseases very easily. In Deep Learning(DL), Convolutional Neural Networks (CNNs) are a cutting-edge method for image classification tasks. And Plant Disease Detection is an image classification task in which image is given as input and a class of plant disease is obtained as an output. This research study reviews the CNN-based approaches that are used to detect various diseases in plants.**

***Index Terms*—Artificial Intelligence(AI), Convolutional Neural Network(CNN), Deep Learning(DL), Machine Learning(ML), Plant Disease Detection, Transfer Learning**

I. Introduction

Agriculture is the practice of cultivating soil and raising crops. It is extremely important for a country's economic progress. In agriculture, excellent production quality and crop output are critical. This helps to avoid a massive economic loss. Making agriculture more profitable requires the early diagnosis of plant diseases. Pathogens such as viruses, fungi, and bacteria as well as climate changes can cause plant diseases.

Infections can be spotted by carefully examining the plant leaves as it's more common for the symptoms to appear at the leaves early than other parts of the plants.

Symptoms may appear in leaves like discoloration of leaves, black spots in leaves, etc. Next, there is a chance that other symptoms could appear in other parts of the plants. Farmers use to hire an expert for the early detection of plant disease but it's very expensive as well as time-consuming. And Some farmers try to detect the plant disease using the naked eye which is very expensive, takes too much time, and sometimes it could lead to a false prediction.

The purpose of DL[1] is to diagnose plant diseases quickly and at a minimal cost. DL[1]has a revolutionary algorithm, CNN, that consistently outperforms other algorithms. Fig. 1 shows the major phases in DL[1], which are Feature Extraction & Classification. A Spatial Filter is generally used for feature extraction in DL[1] as well. It is applied directly to the pixels of an image. It is typically considered that a mask has a specific center pixel that is added to its size. To move this mask to cover all pixels in the image, its center is moved across the image.

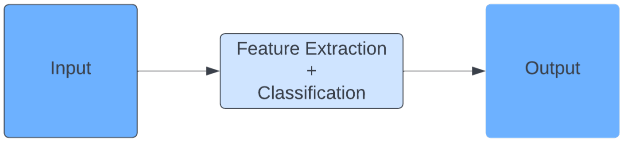


Fig. 1. Structure of Deep Learning

An overview of Convolutional Neural Network is presented in Section 3, followed by an analysis of types of plant diseases in Section 2. In Section 4, the Literature Survey is discussed. In Section 5, the Conclusion is discussed, followed by a list of references.

II. Types of Plant Disease

Diseases in plants are caused by climatic change or by pathogens like bacteria, fungi, and viruses. Diseases caused by pathogens are infectious plant diseases and this can damage the mild leaf or fruit or can lead to death. Diseases caused by some external factors like climatic change are the non-infectious disease.

There are two types of Plant Diseases: Infectious and Non-Infectious.

Mineral toxicity, acidity in the soil, nutrient deficiency, as well as other factors lead to non-infectious diseases.

Infectious Plant Disease is further divided into three types: Bacterial Disease, Fungal Disease, and Viral Disease.

A. Bacterial Disease

Internally, it damages the plant without presenting any visible signs. Leaf spot with a Crown gall, Fruit spot, Canker, yellow halo, etc.. are some of the symptoms.

B. Fungal Disease

More than 80% of plant diseases are caused by fungi. It infects by killing the cells of that plant. It is caused by infected soil, workers, machinery, tools, animals, etc. Symptoms

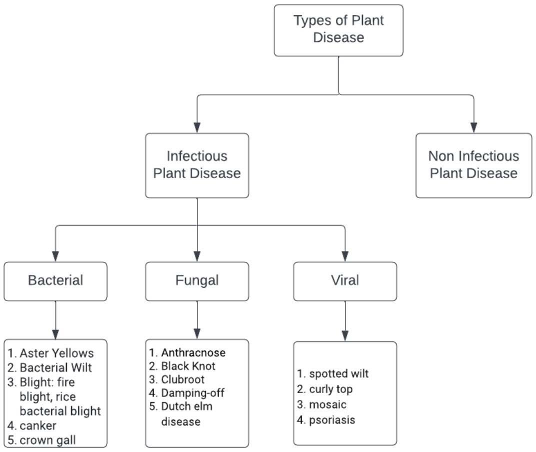


Fig. 2. Types of Plant Disease

include Damping-off of seedlings (phytophthora), Birds-eye spot (anthracnose), Leaf spot (septoria brown spot), chlorosis (yellowing of leaves), etc...

C. Viral Disease

The viral disease is difficult to detect in plant leaves. Because most viruses are hidden in nature, we can't detect viral disease till a particular stage. An abnormal growth pattern, necrotic spots, unusual light and dark green blending of leaves, stunting, ring patterns on foliage, and distorted flower color and formation are some of the signs of Viral Disease.

III. Convolutional Neural Network

To detect plant diseases, CNN, also called the CovNet algorithm, comes under DL[1].So, Deep Learning, abbreviated as DL[1], is a branch of AI in which a system or a machine mimics human behavior, and how humans acquire knowledge.

It's just like brain mapping. And in agriculture, DL[1] is used to check the water level of the crops and also used for monitoring the temperature. Farmers can observe their fields anywhere in the world. And the main advantage of using deep learning in agriculture is it saves time and too much effort.

For this task i.e. diagnosing the plant disease, due to its higher accuracy, CNN is considered as a highly advanced algorithm. There are ML algorithms like Support Vector Machine, and K-Nearest Neighbours as well but ML algorithms don't give good accuracy in the task of the detection of plant disease. So, that's why we use DL[1] algorithms instead of ML. And one such algorithm is CNN.

CNN consists of several layers. Different features of the input image are detected by each layer. Various kernels or filters are applied to an image to produce better and more detailed results. Filters can start as simple features in the lower layers. To further identify features that uniquely represent the input object, the filters become more complex. After each layer is partially recognized, the output of each convolved image becomes the input for the next layer. A Fully Connected layer recognizes the image or object in the last layer. In the case of Plant Disease detection, the image input is the image of a plant leaf because leaf shows the symptoms earlier than any other part of a plant. To detect plant disease, CNN uses texture, colors, and features to classify them.

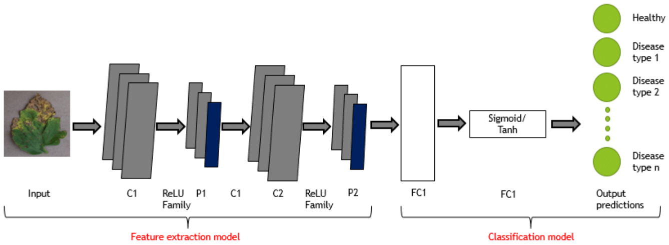


Fig. 3. Convolutional Neural Network(CNN) schematic design [3]

Convolutional Neural Networks (CovNets) are portrayed schematically in Fig.2.

In DL[1], We can use CNN in the recognition of patterns in the images. It was developed in the year 1980. Using CNN, the pixel value of an image serves as the input, the model is trained, and the features are then extracted for classification.

In CNN, there are three layers, each of which has different parameters and performs different functions on the input data:

A. Convolution layer

This layer applies filters to the original input image. The number and size of kernels are the most important parameters. The input image is transformed by a convolution layer to extract features.

B. Pooling layer

It works similarly to the convolution layer, but it focuses on specific tasks like average pooling and max-pooling. The highest value is taken in a certain area of a filter in max pooling, whereas the average value is taken in a certain area of a filter in average pooling. Its main purpose is to reduce the image's dimensionality.

C. Fully Connected layer

It's somewhat similar to the Multilayer Perceptron(MLP) which is used to compact the result before the classification.

And the main thing in CNN is Activation Function. An activation function is a node inserted between or end of a neural network. Neurons rely on them to decide whether to fire or not. There is a total of 6 types of Activation functions - ReLU, Leaky ReLU, tanh, Maxout, Sigmoid, and ELU. In CNN, ReLU is used extensively.

IV. Literature Survey

Detecting and classifying plant disease is crucial in the agriculture sector, and deep learning plays a key part in this process. The following are some deep-learning approaches that have been proposed in recent years:

Ferentinos[3] proposed a CNN architecture, trained on an open dataset called PlantVillage[4], which contains 87484

TABLE I
TABLE OF COMPARISON.

Author	Year of Publication	Proposed CNN Model	Pros	Cons
Ferentinos [3]	2018	Deep CNN	Computational power is low	Model cannot predict severity so that farmers can estimate how much pesticide they need, which is harmful to the environment.
Singh et al.[5]	2019	Multilayer CNN	Simple and computionally efficient	Accuracy is lower than other models
Sachdeva et al. [6]	2021	Deep CNN using Bayesian Learning	feature learning is efficient	High computational time
Agarwal et al. [7]	2020	A nine-layer Deep CNN	Memory Efficient	Low testing Accuracy
Mohanty et al.[8]	2016	Deep CNN	works reasonably well	Model training takes a lot of time
Akhtar et al.[9]	2021	CNN model	Less computational efforts	Trained with less number of epochs. Can train a model for multiple epochs
Goncharov et al.[10]	2021	Deep Siamese CNN	Computationally Efficient	low testing accuracy
Jiang et al[11]	2020	CNN+SVM	High accuracy	Takes too much time to train
Tiwari et al.[12]	2021	DenseNet201	solves the vanishing-gradient problem, improves feature propagation, promotes feature reuse, reduction on the number of parameters	Excessive connections not only reduce the computation and parameter efficiency of networks, but also make networks more prone to overfitting.
Geetharamani et al.[14]	2019	A nine-layer Deep CNN	faster training speed, computationally less expensive	Prone to overfitting
Chen et al.[15]	2020	INC-VGG	Higher accuracy than other conventional CNN	Takes more time for training than conventional CNN due to more layers.
Bedi et al.[16]	2021	CAE + CNN	Image recognition problems are very accurate, Automatically detects the important features without any human intervention, and Having a less dimensional input image requires less training time	Object positions and orientations are not encoded and Lack of spatial invariance to input data
Liu et al. [17]	2018	Cascading of AlexNet-precursor network and Inception network.	number of parameters is reduced and faster convergence rate is also there	Prone to underfitting
Atila et al. [18]	2021	EfficientNet	The CNN model performed better than other models that received higher resolution inputs	Computationally not efficient
Brahimi et al. [19]	2018	CNN with fine tuning of AlexNet and GoogleNet	Eliminates the need to extract features from images in order to train models	High Computational Time

plant leaf images divided into 58 classes. 99.53% accuracy was achieved in this work.

To determine if a plant is healthy or unhealthy, Singh et al.[5] suggested a multilayer Convolutional Neural Network having 6 convolution layers and 3 pooling layers. This model is applied to the images of the leaves of mango. Images are gathered from both the PlantVillage[4] Dataset and the field. The accuracy of this work is 97.13%.

Sachdeva et al.[6] proposed a Bayesian Learning based deep convolutional neural network, which implements Bayesian learning on top of the residual network. The PlantVillage[4] dataset was used to train in this work, which contains 20,639 images of healthy and unhealthy tomato, potato, and pepper bell leaf images. It has a 98.9% accuracy rate and no evidence of overfitting.

The CNN model proposed by Agarwal et al. [7] consists of three convolutional layers, three max-pooling layers, and two fully connected layers. Nine different types of tomato plant disease are identified using this model. It had a 91.2% in a test set.

To distinguish thirteen different crop disease types, Mohanty et al. [8] employed a Deep Convolutional Neural Network. They used 53,000 images of healthy and unhealthy plant leaves from the PlantVillage[4] Dataset. It had a 99.36% accuracy rate.

Deep CNN was utilized by Akhtar et al. [9] to determine whether the plant is healthy or unhealthy. They captured images of various plant diseases using an optical camera and other similar equipment. To enhance the dataset size, they used data augmentation. The preprocessing step is followed by data augmentation. After training for up to 45 epochs, the Model achieves the best accuracy of 97.6%.

Goncharov et al. [10] proposed a Deep Siamese Convolutional Neural Network to handle the challenge of a tiny dataset of plant leaves. They gathered photographs of grape leaves and sorted them into four groups. This work achieves a level of accuracy of over 90%.

Jiang et al. [11] suggested a model in which the Mean Shift Algorithm is used for segmentation, artificial computation is used for shape feature extraction, Color features are extracted using CNN, and plant diseases are identified using a Support Vector Machine(SVM) Classifier. The suggested model identifies four types of disease in rice plants. This work had a 96.8% accuracy rate.

To classify plant diseases, Tiwari et al. [12] proposed the DenseNet201 architecture, a deep neural network architecture. PlantVillage[4], the iBeanLeaf dataset[13], Citrus Leaf Images, and Rice Leaf Images were used. It had a 99.20% accuracy rate.

A nine-layer deep conventional neural network (DCNN) was proposed by Geetharamani et al. [14]. PlantVillage[4], an open dataset with 54,305 photos classified into 13 plant diseases, was used to train it. Transformation (Data Augmentation) is performed before training to expand the dataset size. Image flipping, noise injection, gamma correction, and so forth are examples of transformations. A Deep CNN is composed of three convolutional layers, three max-pooling layers, one flatten layer, and two fully connected layers. The accuracy of this work is 96.46%.

Chen et al. [15] studied the use of a deep convolutional neural network for plant disease detection via transfer learning. The INC-VGG model, which combines VGG19 with InceptionV3, was demonstrated using transfer learning. The bottom layers of VGG19 models were used, then the 3X3 ConvBN layer, then two InceptionV3 layers, and finally the Softmax layer. This work has a 91.83% accuracy rate.

Bedi et al. [16] proposed a model which is a combination of the Convolutional Neural Network (CNN) and the Convolutional AutoEncoder (CAE). Training accuracy of 99.35% and testing accuracy of 98.38% was achieved in this study.

The AlexNet-precursor network and an Inception network are cascaded in a novel Convolutional Neural Network structure developed by Liu et al.[17]. The fully connected layers of the AlexNet Model are replaced by the Inception network. To improve convergence speed, Nesterov's accelerated gradient (NAG) optimization technique is used instead of the stochastic gradient descent (SGD) algorithm. In this proposed study, the number of trainable parameters was reduced, which helped to reduce storage requirements. And the accuracy in this work is 97.62%.

For the classification of plant diseases, Atila et al.[18] introduced the EfficientNet architecture. Transfer Learning is used for training. The PlantVillage[4] dataset, which was used in this study, has 55,448 images divided into three categories: apple, grape, and potato. It had a 98.42% accuracy rate.

Brahimi et al.[19] proposed a Convolutional Neural Network (CNN) architecture for identifying the nine types of diseases in tomato plant leaves. They used images of tomato plant leaves from the PlantVillage[4] dataset, which is openly accessible. They developed a classifier using conventional architecture such as AlexNet and GoogleNet. After training, they found that training a model with fine-tuning results in higher accuracy than training a model without fine-tuning. With fine-tuning, GoogleNet's accuracy increased from 97.71% to 99.18%, while AlexNet's accuracy increased from 97.35% to 98.66%.

Even [20] compares the ML algorithms such as SVM, Random Forest classification, and Stochastic Gradient Descendent and DL algorithms such as AlexNet, GoogleNet, LeNet, etc. for the detection of plant disease and the author found that DL algorithms give more accuracy as compared to ML algorithms.

Conclusion

This study looked at various DL algorithms that have been proposed in recent years. In DL, Convolutional Neural Network, or CovNet, are the state-of-the-art algorithm providing the highest degree of accuracy.

The openly available dataset called PlantVillage was used mostly for research purposes. In this dataset, 54,303 images are categorized by species and disease into 38 categories.

DL approaches which are used in the research work saves a lot of time and also a farmer can monitor their fields anywhere in the world. And also the accuracy to detect a plant disease is higher than in other experimental Methods.

References

[1] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553. Nature Publishing Group, pp. 436–444, May 27, 2015. doi: 10.1038/nature14539.

[2] "Example of CNN architecture used in plant disease detection — Download Scientific Diagram." https://www.researchgate.net/figure/Example-of-CNN-architecture-used-in-plant-disease-detection_fig1_353527437 (accessed Nov. 19, 2022).

[3] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," Comput Electron Agric, vol. 145, no. September 2017, pp. 311–318, 2018, doi: 10.1016/j.compag.2018.01.009.

[4] "GitHub - spMohanty/PlantVillage-Dataset: Dataset of diseased plant leaf images and corresponding labels." https://github.com/spMohanty/PlantVillage-Dataset (accessed May 25, 2022).

[5] U. P. Singh, S. S. Chouhan, S. Jain, and S. Jain, "Multilayer Convolution Neural Network for the Classification of Mango Leaves Infected by Anthracnose Disease," IEEE Access, vol. 7, pp. 43721–43729, 2019, doi: 10.1109/ACCESS.2019.2907383.

[6] G. Sachdeva, P. Singh, and P. Kaur, "Plant leaf disease classification using deep Convolutional neural network with Bayesian learning," Mater Today Proc, vol. 45, pp. 5584–5590, 2021, doi: 10.1016/j.matpr.2021.02.312.

[7] M. Agarwal, A. Singh, S. Arjaria, A. Sinha, and S. Gupta, "ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network," Procedia Comput Sci, vol. 167, no. 2019, pp. 293–301, 2020, doi: 10.1016/j.procs.2020.03.225.

[8] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Front Plant Sci, vol. 7, no. September, pp. 1–10, 2016, doi: 10.3389/fpls.2016.01419.

[9] F. Akhtar, N. Partheeban, A. Daniel, S. Sriramulu, S. Mehra, and N. Gupta, "Plant Disease Detection based on Deep Learning Approach," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2021, pp. 74–77, Mar. 2021, doi: 10.1109/ICACITE51222.2021.9404647.

[10] A. Smetanin, A. Uzhinskiy, G. Ososkov, P. Goncharov, and A. Nechaevskiy, "Deep learning methods for the plant disease detection platform," AIP Conf Proc, vol. 2377, no. October, 2021, doi: 10.1063/5.0068797.

[11] F. Jiang, Y. Lu, Y. Chen, D. Cai, and G. Li, "Image recognition of four rice leaf diseases based on deep learning and support vector machine," Comput Electron Agric, vol. 179, no. August, p. 105824, 2020, doi: 10.1016/j.compag.2020.105824.

[12] V. Tiwari, R. C. Joshi, and M. K. Dutta, "Dense convolutional neural networks based multiclass plant disease detection and classification using leaf images," Ecol Inform, vol. 63, no. March, p. 101289, 2021, doi: 10.1016/j.ecoinf.2021.101289.

[13] "beans — TensorFlow Datasets." https://www.tensorflow.org/datasets/catalog/beans (accessed May 26, 2022).

[14] G. Geetharamani and A. P. J., "Identification of plant leaf diseases using a nine-layer deep convolutional neural network," Computers and Electrical Engineering, vol. 76, pp. 323–338, Jun. 2019, doi: 10.1016/j.compeleceng.2019.04.011.

[15] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, "Using deep transfer learning for image-based plant disease identification," Comput Electron Agric, vol. 173, p. 105393, Jun. 2020, doi: 10.1016/J.COMPAG.2020.105393.

[16] P. Bedi and P. Gole, "Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network," Artificial Intelligence in Agriculture, vol. 5, pp. 90–101, 2021, doi: 10.1016/j.aiia.2021.05.002.

[17] B. Liu, Y. Zhang, D. J. He, and Y. Li, "Identification of apple leaf diseases based on deep convolutional neural networks," Symmetry (Basel), vol. 10, no. 1, 2018, doi: 10.3390/sym10010011.

[18] Ü. Atila, M. Uçar, K. Akyol, and E. Uçar, "Plant leaf disease classification using EfficientNet deep learning model," Ecol Inform, vol. 61, p. 101182, 2021, doi: 10.1016/j.ecoinf.2020.101182.

[19] M. Brahimi, M. Arsenovic, S. Laraba, S. Sladojevic, K. Boukhalfa, and A. Moussaoui, "Deep Learning for Plant Diseases: Detection and Saliency Map Visualisation," no. June, pp. 93–117, 2018, doi: 10.1007/978-3-319-90403-0_6.

[20] A. Sungheetha, "State of Art Survey on Plant Leaf Disease Detection," Journal of Innovative Image Processing, vol. 4, no. 2, pp. 93–102, Jul. 2022, doi: 10.36548/jiip.2022.2.004.

A Rubik's Cube Cryptosystem based Authentication and Session Key generation model driven in Blockchain environment for IoT Security

Ankit Attkan
Department of Computer Engineering, NIT Kurukshetra, Kurukshetra, 136119, India, Email: ankit_62000070@nitkkr.ac.in

Virender Ranga
Department of Information Technology, Delhi Technological University, Delhi, 110042, India, Email: virenderranga@dtu.ac.in

Priyanka Ahlawat
Department of Computer Engineering, NIT Kurukshetra, Kurukshetra,136119, India, Email: priyankaahlawat@mail.nitkkr.ac.in

Over the past decade, IoT has gained huge momentum in terms of technological exploration, integration and its various applications even after having a resource-bound architecture. It is challenging to run any high-end security protocol(s) on Edge devices. These devices are highly vulnerable towards numerous cyber-attacks. IoT network nodes need peer-to-peer security which is possible if there exists proper mutual authentication among network devices. A secure session key needs to be established among source and destination nodes before sending the sensitive data. To generate these session keys, a strong cryptosystem is required to share parameters securely over a wireless network. In this article, we utilize a Rubik's cube puzzle based cryptosystem to exchange parameters among peers and generate session key(s). Blockchain technology is incorporated in the proposed model to provide anonymity of token transactions, on the basis of which the network devices exchange services. A session key pool randomizer is used to avoid network probabilistic attacks. Our hybrid model is capable of generating secure session keys that can be used for mutual authentication and reliable data transferring tasks. Cyber-attacks resistance and performance results were verified using standard tools, which gave industry level promising results in terms of efficiency, light-weightedness and practical applications.

CCS Concepts. •Security and privacy~Security services~Authentication~Multi-factor authentication •Security and privacy~Cryptography ~Symmetric cryptography and hash functions •Security and privacy~Cryptography~Public key (asymmetric) techniques~Digital signatures •Computing methodologies~Modeling and simulation~Simulation theory~ Network science •Network~Network types~Cyber-physical networks •Computer systems organization~Embedded and cyber-physical systems~Sensor networks •Mathematics of computing~Probability and statistics~ Probabilistic algorithms •Computing methodologies~Concurrent computing methodologies~Concurrent algorithms

Additional Keywords and Phrases: Rubik's Cube Cryptosystem, Mutual Authentication, Internet of Things, Secret Session Keys, Session Key Randomizer, Blockchain, Token (zvma), Cyber-security, IoT Security, Data Privacy, Chaotic maps, Image encryption-decryption, number theory, RCC-Block-MASSK Model, Light-weight

1 INTRODUCTION

The Internet of things comprises "things" that are technologically smart enough to collect data and communicate with each other. There are multiple devices and components of an IoT network such as sensing devices[1], RFIDs[2], routers, gateways, servers, cloud servers etc. Edge layer of IoT network deals with the IoT edge devices that have data collecting and environment sensing capabilities. Using these capabilities, these edge devices collect sensitive data on which the IoT network operates. Data being collected can be healthcare[3] data, smart home[4] applications, tracking gadgets[5], sensing detectors or even biometric[6] security equipment etc. All such devices connected in a network require security solutions in order to prevent attackers from manipulating, fetching or stealing sensitive/private data from IoT network components. These security solutions are mostly heavy in computational applications such as large key sized cryptosystems[7], which on the contrary are not supported by these resource constrained devices. Some devices that are capable of executing such heavy security oriented protocols or cryptosystems run out of battery power due to continuous rigorous computation and simultaneous data transmission. Data transferring from one node to another node is highly risky as IoT networks are vulnerable towards numerous cyber-attacks[8]. To cope with such issues, there is a need for a secure secret session key to establish a secure communication connection between any two peers. This secret session key is supposed to be valid for a particular session between specific devices that generate this common secret session key after exchanging key generating parameters using a secure channel.However, two or more IoT network devices can not trust each other unless they mutually authenticate[9] each other. Authenticating each other before sending data helps in developing a TRUST [10] metric among technological devices that lack human understanding. A centralized Internet of Things network is not able to cope with single point of node failure and suffers from unwanted service interruptions if the single point node fails. It also lacks the anonymity and transactions based data privacy which gets compromised due to shortcomings in authentication mechanisms. If the peers in an IoT network are not able to timely authenticate each other due to an inefficient authentication mechanism, there will be delay introduced in an IoT network resulting in slow scaling of the network[11]. Network congestions are observed in

2691-1914/2023/1-ART1 $15.00
http://dx.doi.org/10.1145/3586578

centralized IoT networks [12]. Such IoT networks face serious vulnerability induced issues such as Man-in-the-middle attack[13], DOS attack[14], packet sniffing[15,16], data-deduplication[17], resource(battery) exhaustion etc. To provide high-end security based on efficient data privacy maintaining mechanisms, Blockchain technology is being utilized to support the security, network peer scaling and distributed service interaction among IoT network components. Blockchain technology is a fruitful integration to the IoT cyber-space as it provides optimal practical solutions to enhance the network security[18]. Blockchain exploits the digital signatures technique in generating the NFTs-Non Fungible Tokens. It uses the lightweight hashing mechanisms[19] for associating relatable data points in a large sized data configuration. Consensus algorithm[20] and transaction handling methodology provides the base of any Blockchain. IoT networks now can be easily integrated with light-weighted blockchains[21,22] in order to serve high-end security solutions. In the co-ordinated run time environment of Blockchain, every network node needs to authenticate themselves among peers and gateways, in order to join the network cluster designated to a specific gateway. Nodes that prefer network blockchain mining[23] are selected on a random basis. Meanwhile in a non coordinated run time environment, any given device can join the IoT network and perform several transactions without proper authorization (lacking proper mutual authentication). Inclusion of Blockchain in any IoT network removes the requirement of third party "TRUST" metrics. In a blockchain enabled IoT network, a concrete security consensus algorithm is required where various heterogeneous technological devices and sensors come to a common agreement point and provide their mutual services as service-balanced-peers. This incorporation of heterogeneous technologies boosts the IoT network security[24]. The goal of Blockchain in sync with the IoT network is to bring a common point to agree on for consensus algorithms to work in communicating and authenticating IoT network devices and provide services to each other or to the upper gateways and servers as per data service demands. Inclusion of Blockchain technology into the Internet of Things world is itself a challenge as Blockchain requires a large number of resources due to the resource exploitative behaviour of the consensus algorithms. IoT networks generate huge streams of data periodically which collectively becomes a very large sized data gathering at the server end. So, there are data storage and network scalability related challenges in practical incorporation of blockchain into the IoT networks. IoT network security requires the protocols and frameworks to be light-weight, while maintaining data privacy. This can be effectively achieved by using efficient and strong mechanisms such as hashing and network security using cryptography respectively. This paper provides such a hybrid solution to multiple challenges addressed earlier. It utilizes the randomness of a real life physical puzzle, the Rubik's Cube[25] and maps it into the cryptographic applications where security relies on the number of permutations and combinations being higher, making it difficult for the attacker to decode the encrypted data being sent over a communication channel[26]. Our Rubik's cube based cryptosystem supports data privacy. Using this cryptosystem we generate a specific secret session key per communication for a sender and receiver pair of nodes. Further using this secretly generated session key, a secure communication session can be created among peers. Multiple connections require multiple secret session keys that can also be used to mutually authenticate one device with another device. We introduce a new crypto-token based blockchain transaction handling mechanism that drives a balanced connective service among IoT network components. Users are responsible for maintaining the account balances for their cluster devices such that their cluster's IoT edge devices can keep receiving as well as keep providing data exchange services. Users need to timely recharge the token balances of their devices in order to keep them in running state.

1.1 Our Contribution

Major contributions of our research are:

- We introduce a new mutual authentication and secret session key agreement mechanism for IoT networks, known as "*MASSK*". Parameters shared among IoT network components are sent using our newly designed Rubik's cube based cryptosystem. IoT services are driven based on Rubik's cube based place shifting blocks in Blockchain, so we call our proposed model, "*RCC-Block-MASSK*" model. These session keys generated are also stored in a dynamic environment in a similar Rubik's Cube logic based session key randomizer (RCR_{KEY}) on a server.
- A new consensus algorithm supports the agreement handshake and block creation mechanism in Blockchain whereas a modified transaction handling algorithm ensures light-weight and smooth updation of IoT service based transactions.
- We propose an efficient Proof-of-Lightweight-Work and Secure Crypto-Transaction (*POLWSCT*) consensus algorithm based on novel "*zvma*" token. Proposed *POLWSCT* consensus algorithm uses lesser amounts of energy quantums, gives lesser executional lags and reduced amount of time for "zvma" token mining and sending updated transactions to other blockchain nodes.
- *RCC-Block-MASSK* model's security is verified against some of the severe IoT network attacks. Security verification and scalability were analyzed using both formal and informal logical mathematical tools.
- NS-3 simulator[27] was used to estimate the run time time delays and cryptographic parameters generation and their average time consumption complexity.
- Formal security verification for security breach attacks were successfully prevented and the reachability of the SSK_{KEY} (Shared Secret Session Key) was proved to be strong.
- A Blockchain integrated IoT network is implemented in order to compute numerous block transactions and also keeping the account of the number of verified blocks mined per standard time interval.
- Proposed POLWSCT consensus algorithm is simulated for resource constrained IoT network nodes by utilizing the Contiki Cooja simulator[28], to compute and analyze run time complexity and efficiency.

1.2 Related Work and Motivation

Internet of Things is a global research area with numerous applications, dependency benefits along with their associated vulnerabilities towards hardware and software attacks. It is still struggling to accommodate various heterogeneous technologies while growing in size over cyberspace. We say it is struggling because as the size scale of the IoT network increases, the scope of being vulnerable towards the security based attacks increases with the increase in difficulty in managing the multiple technologies(with their own security flaws). This not only makes IoT flexible in terms of connectivity but also exposes it towards attack vulnerabilities of the respective technologies being entertained in the IoT network. As dependency of humans grows on the technology, ensuring data privacy and security standards is a necessary requirement for ensuring multiple users with "TRUST" that their data is being handled with proper secrecy and care. Blockchain is a rising and likewise trending technology that can provide better security solutions if the problem of running consensus in an energy efficient manner is taken care of properly. Philip used the blockchain technology to manage digital finance in privately running networks[29]. As for rising technologies, 5G and 6G networking were applied to the IoT network environment by Stergiou[30,31,32] while aiming for secure data transmission. Joboury[33] deduced a solution to incorporate cooperative synchronization in Fog computing networks. Biswas[34] proposed an incentive driven mechanism for dynamic motion capable networks such as vehicular synchronization in modern traffic based on blockchain token based service exchange. Heavier nodes in an IoT network that can run consensus based architecture can deduce a cooperative way of debating and celebrating a transaction in the ledger(we refer to these devices as "resource capable devices), but for those network components that at the edge layer of the IoT network which "resource constrained devices", there is a need to generate "Trust" among any two more communicating devices. Consensus algorithms execute only to generate agreement among devices without the need of generating any third party "Trust" parameters. It is done based on various modes of proofing based consensus mechanisms such as PoW[35], PoS[36], PoAu[37], PoA[38] etc. We propose a different lightweight consensus algorithm considering the heterogeneity of network components and applicability in real-time as well as simulated environments. However, the edge nodes do not have sufficient batteries to sustain the blockchain computation. So, instead of using consensus algorithms, there is a need for interoperability among the edge nodes with a session key agreement protocol, which can side by side authenticate the device among each other. We derived such a mechanism for the edge and gateway node environment during cluster communication. Iorga [39] discussed the scalability of such scalable resources using the distributed logic between the user and the gateway nodes. Al-Ali[40] proposed that the handshake mechanism can be treated as a light-weight mode of interaction among the low computation devices in IoT networks. AK Das [41] provided a different 3-factor based mutual authentication mechanism among resource constrained devices. Fan[42] provided a session key agreement protocol for an even greater number of nodes while avoiding generic attacks based on advanced hashing(lightweight) and concatenating mechanism for parameter exchange. Kasyoka[43] gave a mutual authentication security framework that operates without any digital certificate. It rather favours digital fingerprinting and signature mechanisms to create "Trust" factor among data exchanging devices. Tan[44] also highlighted the use of PUF based security and trusted parameter passing mechanism. All such mechanisms[45,46,47] motivated us to utilize a light-weight but secure enough authentication mechanism for cluster based communication among IoT_{EDGE} nodes and the cluster Gateway nodes (GW_{NODE_N}). However, heavier nodes are managed by the proposed *POLWSCT* consensus driven blockchain which provides "*zvma*" token as an incentive to the service providing or mining nodes. The same token is also utilized to monitor the barter exchange of services, incentive and data on the cost of same "zvma" token. To gain strength against the strong attacks from the adversaries, we incorporate a newly designed Rubik;s cube based cryptosystem for encryption and decryption purposes. We got motivation for such physical puzzle based mechanism to be utilized in digital security from various such inventories [48,49,50,51,52,53,54,55,56,57,58,59,60]. The proposed *POLWSCT* consensus, RCC-cryptosystem, session key generation and authentication mechanism all are discussed in detail in various sections ahead.

1.3 Organization of the article

This article aims at providing solutions to multiple issues with IoT network security by introducing a technologically boosted hybrid model that utilizes Blockchain integration with IoT networks while enhancing security using mutual authentication, secure session key generation, crypto-system utilization and a light-weight consensus algorithm for efficient transaction updation. Section 1 provides a brief overview of the IoT network security, issues, challenges and vulnerability factors. It also enlightens the problem statements which are addressed by the major contributions of our work in the field of IoT security. Section 2 discusses the blockchain environment, transaction handling algorithm and IoT integration with the advanced *POLWSCT* consensus algorithm. It discusses the block mining technique and "zvma" token exchange mechanism that drives the IoT service barter system. Section 3 illustrates our newly introduced Rubik's cube based developed cryptosystem to exchange secret parameters and its working mechanism. Section 4 combines the mutual authentication mechanism with the secret session key generation and management protocol among communicating IoT network peer components. Section 5 discusses the secure storing of multiple session keys on a server using a Rubik's cube logic based session key randomizer(RCR_{KEY}). This randomizer mixes up multiple session keys in a complex manner such that the attacker is not able to reveal the inner components(the secret session keys) to compromise the communication channel. Section 6 highlights the security analysis and verification of the proposed RCC-Block-MASSK model using multiple standard security verification tools. Section 7 provides Future scope and conclusion of this paper followed by significant references. Table 1 illustrates some of the significant abbreviations and terminologies used throughout the paper:

Table 1. Major Abbreviations and Terminologies used in this research

Serial No.	Major Symbols /Acronyms Used in the Article		Meaning Represented by the Symbol
1.	*MASSK*	:	mutual authentication and secret session key agreement mechanism
2.	*RCC-Block-MASSK*	:	Rubik's Cube Cryptosystem based Authentication and Session Key generation model driven in Blockchain environment for IoT Security
3.	RCR_{KEY}	:	session key randomizer
4.	*POLWSCT*	:	Proof-of-Lightweight-Work and Secure Crypto-Transaction
5.	zvma	:	Zonal Verification and Mutual authentication token of service
6.	G_{NODE} and GW_{NODE_N}	:	Gateway node used in blockchain and mutual authentication algorithms respectively
7.	RC_{PUZZLE})	:	Rubik's Cube puzzle
8.	RP_{+ve}	:	randomly chosen positive prime integers
9.	B_{MAX},	:	Maximum number of blocks allowed in a blockchain at a given instance
10.	$P_{NODE/MINER}$	:	Network Peer nodes (miners/nodes)
11.	$USER_P$ and $USER_Q$	:	IoT Edge device users "P" and "Q" participating in a transactional update in B-IoT
12.	S_{PAY}	:	the sum to be transacted between $USER_P$ and $USER_Q$
13.	"P_{XC}" and "Q_{XC}"	:	denote the increment and decrement of user's account balances
14.	G_{ZP^*}	:	defines the cyclic generator
15.	$1P_{NODE/MINER}$ and $2P_{NODE/MINER}$	:	Two miner/network node(s) about to share sensitive information among one another
16.	"A_X"	:	arbitrarily selected number
17.	$CP\gamma \in [0, 1]$	:	It is the controlling parameter.
18.	Z_{i+1}	:	chaotic maps for diffusion on encryption key image with the data-image
19.	*βeta*	:	Standard encryption diffusion threshold parameter *used for adjusting diffusion deviation*
20.	$SnapS_1$ and $SnapS_2$	:	$SnapS_1$ *is* the encryption image for the data to be sent to the node $2P_{NODE/MINER}$ *and* $SnapS_1$ is used to diffuse it with the data image $SnapS_2$
21.	IMG_{CIPHER}	:	The cipher image to be generated after successful encryption process.
22.	CT_{Admin}	:	Control Tower (administrator)
23.	$Digi\text{-}Cert\text{-}S_{SERVER}$	:	digitally verifiable certificate for S_{SERVER}
24.	SP_{KEY}, PK_{KEY}	:	for low levelled edge devices, these small private key and public key respectively
25.	$PR_{GENERATOR}$	:	It is the ambiguity generator used to mix the secret parameters
26.	PSS_{KEY}	:	public server station key
27.	SSP_{KEY}	:	server station private key
28.	$T_{GW\text{-}REG}$	:	It is the instantaneous time at which the gateway node got registered by the CT_{Admin}
29.	$GW\text{-}ID_{NODE\_N}$	:	It is the distinctive singular identification assigned number
30.	$GW\text{-}ID_{PSEUDO}$	:	pseudo-ID for gateway(s)
31.	$GWR_{SS} \in Z\,q^*$	:	arbitrarily a nonce for gateway node
32.	$Digi\text{-}Cert\text{-}\ GW_{NODE\_N}$	:	the digitally verifiable certificate for GW_{NODE_N}
33.	$Digi\text{-}Cert\text{-}\ Dyn\text{-}IoT_{EDGE}$	:	uniquely identifiable digital certificate, for $Dyn\text{-}IoT_{EDGE}$
34.	$GWSP_{KEY}$)	:	GW_{NODE_N} secretly creates its individual private key, *GW_Pkey* such that $GWSP_{KEY} \in (Zq^*)$
35.	$a_{EDGE(i)} \in (Zq^*)$	:	Any device, operating at the edge layer creates an arbitrarily chosen number, $a_{EDGE(i)}$
36.	TS_{EDGE1}	:	for global clock synchronization
37.	SSK_{KEY} *(EDGE, GW)*	:	Secure session key established between edge layered device and the corresponding gateway node after successful mutual authentication
38.	$new\text{-}TID^*_{EDGE}$	:	An arbitrary ID for IoT_{EDGE} *during secure parameter transmission among network peers*
39.	EC_Q	:	Energy Consumed in quantums

2 THE RCC-BLOCK-MASSK MODEL (PHASE 1) : PROPOSED POLWSCT CONSENSUS ALGORITHM FOR BLOCKCHAIN-IOT ENVIRONMENT

IoT networks are abundant in technological heterogeneity of various devices with different functionalities interacting with each other. This large network is divided into multiple clusters which are handled by their administrator assigned gateways respectively. Gateway node acts as the cluster head which administers the data packets being sent within the cluster among peers and outside the cluster towards other foreign cluster(s). Blockchain-IoT integrated environment is discussed ahead:

2.1 IoT Environment and System Architecture

Sensing devices collect data in the form of periodical data streams and send these streams to their cluster handling gateway, which further pushes it towards the servers handling the data [61][62] in orderly form. IoT edge nodes, gateways(cluster heads) and intermediate heterogeneous data packet passing devices such as routers etc [63][64]. All are GPS[65] enabled and geo-location active devices that are timely synchronized with each other. Gateway nodes of their respective assigned clusters are self-sufficient in terms of power and computation to support at least a light-weighted Blockchain network. Lower levelled edge devices are not resource-sufficient in order to keep the track of storage and the maintenance of the blockchain ledger. A gateway node is capable of handling such light-weighted blockchain service requests. Multiple gateway nodes interact with each other with the involvement of servers to collectively maintain the blockchain network. Numerous on-demand requests that maintain the run-time blockchain ledger are gathered into a common block which is then multicasted to the other peers in order to trigger the block-updation procedure. Lower level peers execute a consensus algorithm and transmit the newly created transactional requests to their handling gateway(s). Consensus algorithm is constructed to be light-weighted in the earlier developmental phase.

2.2 POLWSCT Consensus Mechanism and Transaction Updation Methodology

Our proposed consensus establishing algorithm is defined on the following major observations:

- Proof-of-Work and Proof-of-Token based consensus techniques are highly battery consuming for even the gateway nodes. These algorithms require complete attention till their single iteration completion which drains out rapidly the IoT edge sensing devices. Hence, Traditional consensus algorithms are not reliable.
- Blockchain is to be decentralized in order to avoid single point failure, relying on only one gateway for block updation is out of option, so, transactional updates are handled over multiple gateway IoT nodes such that request load is equally divided and does not puts strain on any single gateway (G_{NODE}).
- Energy and services are graded and provided on the basis of a "*zvma*" token based mechanism that ensures that every block miner or User's sensing nodes(hired) receive a certain token of prize for completing a Blockchain-IoT relevant task. A user recharges its hired cluster of IoT node device's account balance(s) in order to keep the services running whereas the IoT edge devices spend these "*zvma*" tokens in mutual exchange of information constituting the block transactions.

Each and every networking component of an IoT network is required to have some base for parallel synching the service related tasks such as message passing coordination, updating the dataset at the data collecting high-end servers etc. So, time-synchronization among network peers on every level is clocked in sync with the *GC*-Global Clock. *POLWSCT* is a global time-clocked algorithm for maintaining consensus goals. It is based on the *NxNxN* dimensional Rubik's cube oriented NP-Hard problems mapped to computational devices based on their resource configuration. Dealing with randomly picked very large prime numbers based NP-Hard problems are densely computation demanding and can not be considered as light-weight, so we moved to a 3-Dimensional Rubik's puzzle mechanism that peaks its security strength on the basis of probabilistic randomness and difficulties in detecting the number of combinations in limited amount of time. It is to be noted that simple rubik's cube are easily solvable, but the induction of randomly fitted small prime numbers in *NxNxN* cubes working with various jumbled cube orientation takes the security level to such peaks, that the puzzle becomes NP-Hard for IoT edge devices having insufficient resources to ever solve such puzzle's. Here, "*N*" is the number of blocks in one row or column per dimension. Three techniques are merged into one "*NxNxN*" NP-Hard Rubik's cube, which are, Difficulty in finding the prime factors of an unknown prime number(randomly chosen),Parallel coping with the timely changing spatial positions of each block in the Rubik's cube puzzle where the best match is jumbled. It is to be noted that instead of going for large primes, the proposed consensus algorithm chooses smaller primes with collaboration of 3-dimensional puzzles. This makes the consensus algorithm light-weight in terms of computation and lesser resource hungry, and Decrypting the cryptographically linking blocks within a limited number of attempts and managing difficulties for devices(hacking or mining) to compute the number of permutations and combinations while keeping the track of cube's dynamically changing blocks.

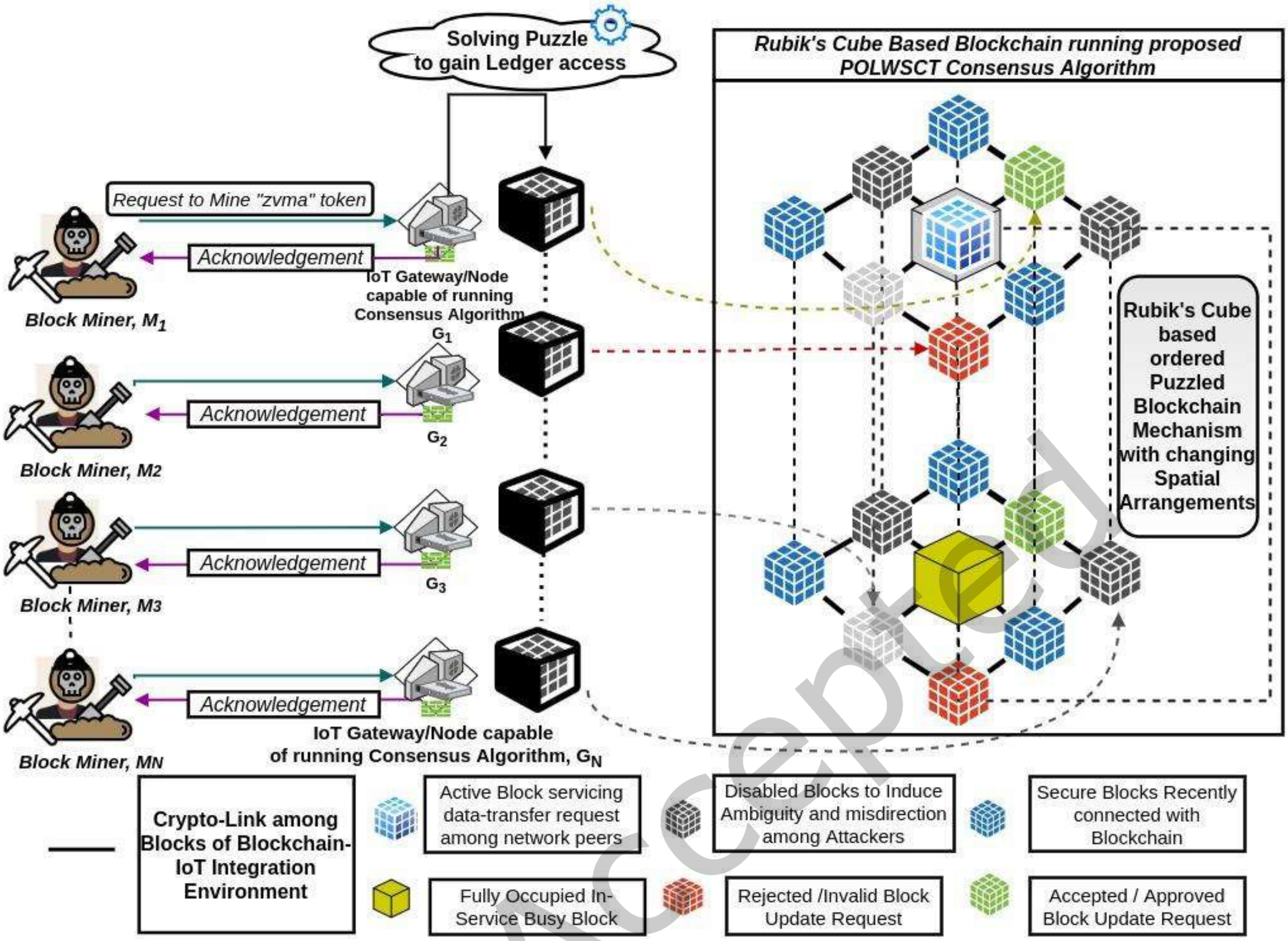

Solving Puzzle to gain Ledger access
Rubik's Cube Based Blockchain running proposed POLWSCT Consensus Algorithm
Request to Mine "zvma" token
Acknowledgement
IoT Gateway/Node capable of running Consensus Algorithm
G1
Block Miner, M1
Acknowledgement
G2
Block Miner, M2
Acknowledgement
G3
Block Miner, M3
Acknowledgement
Block Miner, MN
IoT Gateway/Node capable of running Consensus Algorithm, GN
Rubik's Cube based ordered Puzzled Blockchain Mechanism with changing Spatial Arrangements
Crypto-Link among Blocks of Blockchain-IoT Integration Environment
Active Block servicing data-transfer request among network peers
Disabled Blocks to Induce Ambiguity and misdirection among Attackers
Secure Blocks Recently connected with Blockchain
Fully Occupied In-Service Busy Block
Rejected /Invalid Block Update Request
Accepted / Approved Block Update Request

Figure 1 Shows the network peer/miner interaction with the Blockchain in Aim to gain "zvma" tokens, consensus control and transactional block updation rights

For example, If we have a "*8x8x8*" rubik's cube puzzle with some dummy blocks "B_{DUMMY}" (induced for confusion creation among attackers), some valid blocks "B_{VALID}" along with integer and prime values in a scrambled order. The spatial arrangements of each dummy or valid block keeps changing according to the global time-synchronization. Whether it is an attacker "A_{ADV}", or a Mining node "ND_{MINER}", to reveal the contents of the block, firstly the node will have to decrypt the crypto-link of the blocks while the spatial arrangement of the rubik's cube puzzle is dynamically changing, then after penetrating into the block puzzle, the node needs to guess the random number that is the suitable positive prime for that puzzle. This prime number acts as the "*key*" to disclose the contents of the Rubik's block, only if fed as input to the Puzzle at a specific cube block orientation. There are exhaustive number of combinations to do such matching which makes it even harder to crack. On the other hand, the nodes which successfully hit all the three requirements receive incentive in the form of blockchain consensus token reward-"*zvma*" token. The goal of *POLWSCT* based rubik's cube blockchain puzzle is to find a constant positive integer using an efficient hash function. The IoT edge devices while mining randomly pick up a NP-Hard cube puzzle and try fitting their randomly chosen positive prime integers, "RP_{+ve}" into the Rubik's cube puzzle (RC_{PUZZLE}) such that the internal contents form a perfect prime factorization pair. The proposed consensus algorithm ensures that every participating IoT network node /miner is assigned a limited amount of time for finding a solution to the randomly chosen puzzle. As the time limit is exceeded , the output, whether successful or not is multicasted to all the other nodes/miners that participated. This multicasting of the results ensures the consistency and integrity in the Blockchain while maintaining a transactional balance on the basis of defined consensus. The miner with the maximum number of decrypted blocks/mined blocks from the randomly chosen Rubik's cube puzzle gains the authentic right to update the blockchain. If there is a tie among two or more nodes/miners where the maximum number of blocks, B_{MAX}, is same, then to avoid the forking chances, we define a new proof-of-light-weight-work-and-secure-crypto-transaction, where the network peers having a higher ID is assigned the priority to hand-over the transaction updation control. The transaction updation, mining and *POLWSCT* consensus algorithm, all work in a global clock mapped environment to ensure the data consistency, integrity and reliability. Following are the major phases in mining a block using POLWSCT consensus algorithm:

Phase 1: All the blockchain comprising and handling network peers, *"$P_{NODE/MINER}$"* are allotted a limited time interval within which the *RC_{PUZZLE} is to be solved.* *RC_{PUZZLE}* is randomly chosen by the miner. The POLWSCT consensus algorithm assigns the randomly chosen puzzle to the miner depending on numerous parameters in terms of resources and computation capability of the device. The amount of battery time available on a given "*$P_{NODE/MINER}$*" also determines the type of puzzle, which can be assigned during runtime. The chances of tie-ups are managed within the mechanism via exceptions as provided below in the pseudocode:

2.3 The pseudocode for POLWSCT consensus algorithm is given below:

Algorithm: POLWSCT-Proof of Light-weight work and secure crypto-transaction

Output of POLWSCT module: Balanced Blockchain network peers with updated transactional records

```
Step:1    :  Initial  Setup( )
             {   Assign all unique desirable ID(s) to peer nodes/miners $P_{NODE/MINER}$
                 Set Global synchronization time "$G_{TIME}$"→"0" for all $P_{NODE/MINER}$
                 Transmit "$G_{TIME}$" →$P_{NODE/MINER}$ to time align every device participating in
                 Blockchain consensus for services or mining purposes                 }

Step: 2   :  (i) Main( ) {  For i=0 to N,
                   { Call (Compute_Resource_Power( $P_i$ ));
                     Maintain computation power ad resource record for each Blockchain peer $P_{NODE/MINER}$ in List[ $P_i$ ];
                                                                                  }      //End of For loop

ii.       :  Randomly choose $RC_{PUZZLE}$ [i] for each IoT-Blockchain integrated network miner/peer(s),

iii.      :  Assign Puzzles to nodes, For i=0 to N,
                          { List[ $P_i$ ]→$RC_{PUZZLE}$ [i]  }                          //End of For loop

iv.       :  For each Miner node, Call (Solver   $RC_{PUZZLE}$( $P_{NODE/MINER}$ ) ) ;
             Return (No. of puzzles solved per Miner record);

v.        :  Collect results from each $P_{NODE/MINER}$ and store the result in List_Miner [ $P_i$ ],

vi.       :  For each node from i= to N, Prepare Mining ranking, $M_{rank}$[P],

vii.      :  do

viii.     :  { Initialize counter, int count=0;

ix.       :  Call function Compare_Rank();

x.        :  }  while ($M_{rank}$[$P_i$] && count <=0);        // If there's a match in rank, tie-breaker exception handler is executed

xi.       :  If ( count <=0) , Choose Max( $M_{rank}$[$P_i$]),
             else( Choose the $P_{NODE/MINER}$ , $P_i$ , with the Higher ID as Tie-Breaker among multiple
             candidates
xii.  :  Return ( highest Ranking Node/Peer) assigned with transactional rights

xiii. :  Call (Update_Transaction_Ledger( ));                    // Discussed in Step 3 of this Algorithm
         Manage Ledger till key is assigned for a specific interval of time, then release the Ledger key,

xiv.  :  Max ( $M_{rank}[P_i]$)                                  // Function Definition and execution methodology
         { a. If greater two or more $M_{rank}[P_i]$ match,
           b. Choose $R_{ND}$ from Tie-exhibiting Miner Rank Values,
           c. Compute Hash for each Block, $B_{DUMMY}$ or $B_{VALID}$, i=0 to N,
           d. Sort these h( $M_{rank}$ ) in Descending Order, Return (Maximum Value);  }

xv.   :  Repeat Step 2; }                                        // End of Main( ) function
```

Step: 3. In order to update the Blockchain Ledger transactional records and broadcast it among all the IoT peers participating in a mining contest for record consistencies. We assume that a IoT edge device User, "*$USER_P$*" initiates a transaction to another IoT network peer (say) "*$USER_Q$*" and the sum to be transacted is "S_{PAY}". Assuming the instantaneous balance during the transaction for *$USER_P$* and $USER_Q$ are "*P*" and "*Q*" respectively (in terms of "zvma" token). The Instantaneous balances of their respective USER accounts are encrypted using the Rubik's cube based cryptosystem defined in Section 3. The encrypted User accounts are denoted as "*P_X*" and "*Q_X*". Let "*P_{XC}*" and "*Q_{XC}*" denote the increment and decrement of user's account balances respectively.
Calling function *Update_Transaction_Ledger( )*:

→Update_Transaction_Ledger( ) {

```
1. :  The newer crypto-secure balance of $USER_P$ is $P_X$'.
2. :  Arbitrarily produce a highly unlikely unpredictable random number, $A_R$,
      If ($A_R$' is greater than ( $P_{Z*}$ -1 ) ), then, $P_X = A_R$' and $Q_X$ = ( $P_{Z*}$ -1 ), where "$P_{Z*}$" is a small prime but large enough to challenge resource constrained IoT network devices
      else, $P_X$ = ( $P_{Z*}$ -1 ) and $Q_X = A_R$'
3. :  Calculate $S_{PAY}$= [ $P_X$ % $Q_X$ ]
4. :  If ($S_{PAY}$==0), Jump to (3), else, Jump to (1)
5. :  If ($Q_X$ == 1), then assign $A_R = A_R$' , Jump to (5), else, Jump to (1),
6. :  Calculate $P_{XC}1$= ( $G_{ZP*}$ ) $A_R$ modulo ( $P_{Z*}$), where $G_{ZP*}$ defines the cyclic generator
```

```
7.  :  Estimate $P_Y$= ( $G_{ZP*}$ )( $P_{RANDOM}$) mod $P_{Z*}$, Where $P_{RANDOM}$ denotes $RP_{+ve}$( Randomly chosen positive real number (used as private key ))
8.  :  Calculate $P_{XC}2$= ($P_Y$)($A_R$)($S_{PAY}$)mod P, where $P_Y$ is the public crypto-key
9.  :  Calculate $P_X$'($P_X1$', $P_X2$' )=$P_X$($P_X1$, $P_X2$) - $P_{XC}$($P_{XC}1$, $P_{XC}2$),
10. :  Assuming the new crypto-protected account balance of the $USER_O$ is $Q_X$,
11. :  Arbitrarily produce a highly unlikely unpredictable random number, $A_R$',
       If ($A_R$'' is greater than ( $P_{Z*}$ -1 ) ), then, $P_X$ = $A_R$'' and $Q_X$ = ( $P_{Z*}$ -1 ), where "$P_{Z*}$" is a small prime but large enough to challenge resource constrained IoT network devices
       else, $P_X$ = ( $P_{Z*}$ -1 ) and $Q_X$= $A_R$"
12. :  Calculate $S_{PAY}$= [ $P_X$ % $Q_X$ ],
13. :  If ($S_{PAY}$==0), Jump to (14), else, $P_X$=$Q_X$, $Q_X$= $S_{PAY}$ and then, Jump to (12).
14. :  If ($Q_X$ == 1), then assign $A_R$= $A_R$'' , Jump to (15), else, Jump to (11),
15. :  Calculate $Q_{XC}1$= ( $G_{ZP*}$ ) $A_R$ modulo ( $Q_{Z*}$), where $G_{ZP*}$ defines the cyclic generator
16. :  Estimate $Q_Y$= ( $G_{ZP*}$ )( $Q_{RANDOM}$) mod $Q_{Z*}$, Where $Q_{RANDOM}$ denotes $RP_{+ve}$(Randomly chosen positive real number(used as private key ))
17. :  Calculate $Q_{XC}2$= ($Q_Y$)($A_R$)($S_{PAY}$)mod $Q_{Z*}$, where $Q_Y$ is the public crypto-key
18. :  Calculate $Q_X$'($Q_X1$', $Q_X2$' )=$Q_X$($Q_X1$, $Q_X2$) - $Q_{XC}$( $Q_{XC}1$, $Q_{XC}2$),
19. :  Final consistent Account transaction results are reflected as New $P_X$ and $Q_X$ respectively.
       Return (Px, Qx);                    } // End of function
```

The transaction update in the maintained ledger is multicasted to the other blockchain consensus participating nodes in IoT network. This maintains an updated record at every mining node to maintain consistency and data integrity. Section 3 ahead defines the Rubik's cube based crypto-system on the basis of which the secret parameters are passed on among IoT network peers to generate (private key $X_{PRIVATE}$, public key Y_{PUBLIC}) key pair for creating a secure communication mechanism and to authenticate the peers among each other in an IoT-Blockchain environment. Initiating a secure secret session key generation depends on the synchronized working of blockchain consensus algorithm with the IoT run-time environment. Rubik's cube based cryptosystem, RCC_{CRYPT} provides the secure communication mechanism using secure image encryption mechanism with private and public key pairs. The proposed *RCC-Block-MASSK* model exploits the complex security provided by the cube randomness as discussed ahead.

3 THE RCC-BLOCK-MASSK MODEL (PHASE 2): PROPOSED RUBIK'S CUBE BASED CRYPTOSYSTEM FOR SECURE PARAMETER EXCHANGE

Rubik's cube was invented by Ernő Rubik in 1974 [66]. It is a classic 3x3x3 cube with 6 fixed center face pieces while the side and corner pieces are allowed to move freely. This is an old puzzle where each face of the cube is of different colours and when scrambled, there are certain moves possible to bring back the cube to its original form order. The scrambled form can have multiple combinations and permutations of scrambling and coming back to the original structure by algorithmic reordering respectively.

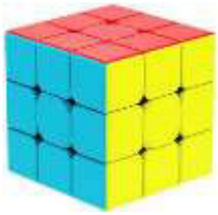


Figure 2.1. Correct Order of a Classic 3x3x3 Rubik's Cube

Figure 2.2. A Scrambled Rubik's Cube after certain rotation Operation (s)

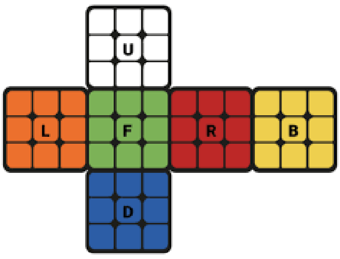


Figure 2.3. A 3-Dimension to 2-Dimension conversion of Rubik's Cube. (This concept is used to convert cube data in 3-D to 2-D for Cipher image generation

The number of possible permutations and combinatorics is so high that its predictability is highly random and difficult to predict by mere guessing. Although there are multiple algorithmic solutions available to the Rubik's puzzle [67,68] which can solve it within a limited time interval, We use the idea of Rubik's cube to extend it further up to *NxNxN* ordered cube where "N" is sufficiently large to avoid any brute force[69] and other generic predictability based attacks[70]. The value of "N" is not that large that it can not be handled by the small computation powered IoT edge device(s). This proposed encryption- decryption cryptosystem is designed keeping in mind the resource scarcity and computation limitation of IoT edge devices. The high randomness of the NxNxN increases the chances of being secured against multiple network attacks that are possible during a data packet transmission and it decreases the chances of (private key, public key) being compromised by the attacker. The original structure, scrambled structure and 3-D to 2-D conversion of cube structure are shown in Figure 2.1,Figure 2.2 and Figure 2.3 respectively.A classic 3-ordered Rubik's cube was originally believed to possess nearly 3 billion combinations, which were corrected and found to be even higher in number, depending on from which piece of the cube the counting initiation is exhibited. There are 40,320 possible outcomes of spatially arranging the corner eight pieces of a Rubik's cube. Edges of the cube can be differently ordered in half of 12! ways, which is 239,500,800 possible scrambling arrangement outcomes. Corner pieces have 3^7 possible permutations. Of the twelve edges, eleven edges can be individually rotated to multiple directions whereas the twelfth one is mutually dependent on the previous eleven edge rotations, giving 2^{11} number of possible arrangements. So, in total there are,

$$2^{11} \text{x } 3^7 \text{ x } 8! \text{ x } 12!/2 = 43252003274489856000,$$

which is 43 Quintillion ways of possible arrangements. For an *NxNxN* order rubik's cube, the number of arrangements is given in Eq(1),

$$N = \begin{cases} \dfrac{8! \times 3^7 \times 12! \times 2^{10} \times (24!)^{\frac{(n-3)(n+1)}{4}}}{24^{\frac{3[(n-2)^2-1]}{2}}}, & n = 2k+1, k \geq 1, k \in Z \\ \dfrac{7! \times 3^6 \times (24!)^{\frac{n(n-2)}{4}}}{24^{\frac{3(n-2)^2}{2}}}, & n = 2k, k \geq 1, k \in Z \end{cases} \quad \text{..............}Eq\ (1)$$

Clearly, there are so many combinations for just a small value of "N" which induces the required unpredictability factor among the secrecy mechanism of resource constrained IoT edge devices. Using our proposed Rubik's cube cryptosystem, the network nodes in an IoT network can share secret parameters securely while avoiding passive snooping[71] or network sniffing[72] attacks. Our model provides two possible ways of mutual authentication among network peers and gateways, First, by using the encryption-decryption key pair of the proposed Rubik's cube based cryptosystem, and, Second, by using the Secret Session key, which is generated using the mechanism discussed in of proposed *RCC-Block-MASSK* model (phase 3) in section 3. Following is the pseudocode for generating encryption-decryption key pair for N-order Rubik's cube based crypto-system:

3.1 The Algorithm for N-Order Rubik's Cube based encryption/decryption key pair generation algorithm is given below:

Algorithm: *RCC-Rubik's Cube Crypto-system*
Output of RCCmodule: *Private and Public key pair(s) for secure parameter exchange among network peers IoT.*
Initial Setup: Suppose two IoT network nodes, $1P_{NODE/MINER}$ and $2P_{NODE/MINER}$ need to share sensitive information between each other, then, $1P_{NODE/MINER}$ chooses a randomly picked positive integer *"RP_{+VE}" such that* RP_{+VE} >=3 (going for higher order for security reasons) , and convert it into the stream of binary bits. Each block of Rubik's cube stores 1 byte of data, i.e, 8-bits. By Default, the initial cube dimension is *"3x3x3"*. Open the 3-order rubik's cube into a 2-Dimensional plane as shown in Figure 2.3. Choose arbitrarily a starting point and start filling the bits in an ascending sequence. Fill the necessary empty blocks with dummy zeroes *(0's)*.

- Scramble the cube into a jumbled order using *Scramble_Cube( )* function defined in section 3.3. and append the "starting point of filling bits of randomly chosen number [RP_{+VE}]*"* at the end of the scrambled cube in the dummy block section. Valid IoT devices are allotted the de-scrambling process algorithm by the control room/Base station.
- Send the scrambled cube with the appended starting point to the receiving $2P_{NODE/MINER}$. The receiving node $2P_{NODE/MINER}$, de-scrambles the cube and reads the randomly chosen RP_{+VE}, and sends the receiving acknowledgement*(Ack).* Now, both of the sending and receiving nodes have a common number *"RP_{+VE}".* From this point onwards, each of the nodes share secret parameters in the form of a *"RP_{+VE} -Order Cube".*

Note: Till now, Actual data is not sent among sending and receiving nodes, $1P_{NODE/MINER}$ and $2P_{NODE/MINER}$. The two nodes have secretly shared the randomly chosen "Order" of the cube, both parties will be using it from this point of time to share the data which will be encrypted and decrypted. It is necessary in order to induce confusion among attackers passively listening to the data transmission.

3.1.1 Encryption Phase:

Step 1. Create a RP_{+VE} -Ordered cube and enter a newly arbitrarily selected number *"A_X"*, in the form of bits just like we did in the initial setup phase of 3.1. Add padding zeros in the cube and append the bits-filling starting point at the end of the cube.

Step 2. Call function *Scramble_Cube( ),* which returns the cube in scrambled form. Convert the 3-Dimensional cube into a 2-Dimensional plane of bit sequences as in Figure 2.3. Call this cube as 1st_RP$_{\text{CUBE}}$. Take a snapshot of the 2-D plane according to the blockchain synchronized global clock defined in section 2 of this paper. This snapshot/image, "*SnapS1 " acts* as the encryption image for the data to be sent to the node *$2P_{NODE/MINER}$*. Scrambling algorithm already has a list of pre-defined valid block arrangements for any N-ordered cube that can be descrambled out of multiple combinations possible.

Step 3. Take the data to be sent, convert it into a stream of bits and fill the newly created *RP_{+VE}* -Ordered cube, name it 2nd_RP$_{\text{CUBE}}$ with the data bits. Keep at least 1/5th of the total blocks to be "Dummy Blocks" in an *RP_{+VE} -Ordered* cube for security reasons.

Step 4. Take the 2nd_RP$_{\text{CUBE,}}$ which holds the data to be sent, call *Scramble_Cube( ),* which returns the scrambled data in the form of a cube, convert this cube from 3-Dimension to an open 2-Dimensional plane, do the same as in Step 2. Take the snapshot of this 2-D plane, name it as "*$SnapS_2$*"

Step 5. Use the encryption image, *$SnapS_1$* and diffuse it with the data image *$SnapS_2$*. By utilizing the Chaos theory, we defined chaotic maps for diffusion on encryption key image with the data-image are described by equations Eq(2) and Eq(3) :

$$Z_{i+1} = \cos\ [\ (\pi\ (2CP\gamma * Z_i + 4(1 - CP\gamma)\, Z_i\, (1 - Z_i) + \beta eta))\ ],\ Z_i < (1/2) \ldots\ldots\ldots\ldots\ldots\ldots \text{Eq(2)}$$

$$Z_{i+1} = \cos\ [\ (\pi\ (2\ CP\gamma\ (1 - Z_i) + 4(1 - CP\gamma)Z_i\, (1 - Z_i) + \beta eta))\ ],\ Z_i > (1/2) \ldots\ldots\ldots\ldots\ldots\ldots \text{Eq(3)}$$

Where βeta is the standard encryption diffusion threshold parameter used for adjusting diffusion deviation, CPγ ∈ [0, 1] is the controlling parameter. Generate the Cipher Image, *IMG_{CIPHER}*, using following row and column diffusion rule for every pixel in the 2-Dimensional Image,

For Rows pixel diffusion, as depicted in Eq(4), Eq(5), Eq(6) and Eq(7), *Create Pixel matrix IMG_{CIPHER} (u,v),* where, *"u"* denotes row number and *"v"* denotes *column number,*

$$IMG_{CIPHER}\,(u,1) = mod(SnapS_1(u,1) + SnapS_2(i,\ RP_{+VE}) + SnapS_1\ (u,\ RP_{+VE} - 1) + SnapS_2{}^{*}(u, RP_{+VE} - 2) + floor\ (Row_1\ (u,\ 1) \times 2^{14}), A_X), v=1] \quad \text{Eq(4)}$$

$$IMG_{CIPHER}\,(u,2) = mod(SnapS_1\ (u,\ 2) + IMG_{CIPHER}\,(u,1) + SnapS_1\ (u,\ RP_{+VE} - 1) + SnapS_2{}^{*}(u, RP_{+VE} - 1) + floor\ (Row_1\ (u,\ 2) \times 2^{14})\ ,\ A_X), v=2] \quad \text{Eq(5)}$$

$$IMG_{CIPHER}\ (u,3) = mod(\ SnapS_1\ (u,\ 3) + IMG_{CIPHER}\,(i,\ 2) + IMG_{CIPHER}\,(u,\ 1) + SnapS_2\ {}^{*}(u,\ RP_{+VE}\ ) + floor\ (Row_1\ (u,\ 2) \times 2^{14}\ ),\ A_X), v=3] \quad \text{Eq(6)}$$

$$IMG_{CIPHER}\,(u,v) = mod(SnapS_1\ (u,\ v) + IMG_{CIPHER}(i,\ RP_{+VE} - 1) + IMG_{CIPHER}\ (u,\ RP_{+VE} - 2) + IMG_{CIPHER}\ {}^{*}(u,\ RP_{+VE} - 3) \ldots\ldots IMG_{CIPHER}\ {}^{*}(u,\ RP_{+VE} - N) + floor\ (Row_1\ (u,\ 1 \times 2^{14}\ ),\ A_X),\ v = RP_{+VE} - 1\ \ ] \quad \text{Eq(7)}$$

→*For Columns pixel diffusionas depicted in* Eq(8), Eq(9), Eq(10) and Eq(11), *Create Pixel matrix IMG_{CIPHER}' (u,v),* where, *"u"* denotes row number and *"v"* denotes *column number,*

$$IMG_{CIPHER}'(1,v) = mod\ (IMG_{CIPHER}\,(1,v) + IMG_{CIPHER}\,(RP_{+VE} - 1, v) + IMG_{CIPHER}(RP_{+VE} - 2\ , v) + IMG_{CIPHER}{}^{*}(RP_{+VE} - 3, v) + floor(Row_2(1,v) \times 2^{14}), A_X), v=1] \quad \text{Eq (8)}$$

$$IMG_{CIPHER}'(2,v)\ = mod\ (\ IMG_{CIPHER0}\,(2,v) + IMG_{CIPHER}'(\ 1,v\ ) + IMG_{CIPHER}\,(\ RP_{+VE} - 2,v\ ) + IMG_{CIPHER}{}^{*}(RP_{+VE} - 3,v\ ) + floor\ (Row_2(2,v) \times 2^{14}),\ A_X), v=2] \quad \text{Eq(9)}$$

$$IMG_{CIPHER}'(3,v) = mod\ (\ IMG_{CIPHER}\ (3,v) + IMG_{CIPHER}'(2,v\ ) + IMG_{CIPHER}'\ (1,v) + IMG_{CIPHER}{}^{*}(\ RP_{+VE} - 3,v\ ) + floor\ (\ Row_2(3,v\ ) \times 2^{14}),\ A_X), v\ = 3] \quad \text{Eq(10)}$$

$$IMG_{CIPHER}'(u,v) = mod(IMG_{CIPHER}\ (u,\ v) + IMG_{CIPHER}'\ (\ u-1,v) + IMG_{CIPHER}'\ (u-2,v) + \ldots\ldots. + IMG_{CIPHER}'(u-2,v) + IMG_{CIPHER}\ {}^{*}(\ RP_{+VE} - 3,v) + floor\ (Row_2\ \ (u,\ v) \times 2^{14}\ ),\ A_X),\ v >= 4] \quad \text{Eq(11)}$$

Step 6. Send this Encrypted image to the recieving IoT network node, *$2P_{NODE/MINER}$*. This is the encryption process from encrypting the data and sending the obtained cipher-image to the other receiving node.

3.1.2 Decryption Phase

Inversing the diffusion process to retrieve the isolated Data and encryption Images in 2-D form, Reversing the conversion and obtaining 3-D cube from this 2-D cube. Later, after obtaining the 3-Dimensional cube with the data in scrambled form, Reverse the scrambling process using the pre-defined, de-scrambling algorithms. De-scrambling the cube gives the data in the actual form which can be read easily using the starting point "A_X", which was defined earlier in step 1 of encryption phase of section 3.1.1.

Step 1. : Inverse the column and row diffusion steps on the Cipher Image, *IMG_{CIPHER}'* as follows *in* Eq(12), Eq(13), Eq(14) and Eq(15) below:

$$IMG_{CIPHER}'(u,v)=mod\ (\ IMG_{CIPHER}(u,v)+IMG_{CIPHER}'(u-1,v)+IMG_{CIPHER}'\ (u-2,\ v)+....+IMG_{CIPHER}'\ (u-2,\ v)+IMG_{CIPHER}{}^{*}(RP_{+VE}-3,v)+floor(Row_2\ \ (u,\ v)\times 2^{14}\ ),\ A_X),\ v>=4\ \ ] \qquad Eq(12)$$

$$IMG_{CIPHER}'(\ 3,v\ )=mod\ (\ IMG_{CIPHER}\ (3,v\ )+\ IMG_{CIPHER}'(2,v)+\ IMG_{CIPHER}'\ (1,v)+IMG_{CIPHER}{}^{*}(RP_{+VE}-3,\ v\ )+\ floor(\ Row_2\ (3,\ v)\times 2^{14}),\ A_X\ ),v=3] \qquad Eq(13)$$

$$IMG_{CIPHER}'(2,v)=mod\ (\ IMG_{CIPHER}\ (2,v)+IMG_{CIPHER}'\ (1,v\ )+IMG_{CIPHER}\ (RP_{+VE}-2,v)+IMG_{CIPHER}{}^{*}(RP_{+VE}-3,\ v)+floor(Row_2\ (2,\ v)\times 2^{14}),\ A_X),\ v=2\ \ ] \qquad Eq(14)$$

$$IMG_{CIPHER}'(1,v)=mod\ (\ IMG_{CIPHER}(1,v)+IMG_{CIPHER}(RP_{+VE}-1,v)+IMG_{CIPHER}(RP_{+VE}-2,v)+\ IMG_{CIPHER}{}^{*}(RP_{+VE}-3,v)+\ floor\ (Row_2\ (1,v)\times 2^{14}),\ A_X),\ v=1\ \ ] \qquad Eq(15)$$

→Similarly, Inverse the diffusion for the Rows as well using Eq(16), Eq(17), Eq(18) and Eq(19) ,

$$IMG_{CIPHER}(u,v)=\ mod(SnapS_1(u,\ v)+\ IMG_{CIPHER}(i,\ RP_{+VE}-1)+IMG_{CIPHER}(u,\ RP_{+VE}-2)+\ IMG_{CIPHER}{}^{*}(u,\ RP_{+VE}-3)\ ....IMG_{CIPHER}{}^{*}(u,\ \ RP_{+VE}-N)+\ floor\ (Row_1\ (u,\ 1\times 2^{14}\ \ ),\ A_X),\ v=RP_{+VE}-1\ \ ] \qquad Eq(16)$$

$$IMG_{CIPHER}\ (u,3)=\ mod(\ SnapS_1\ (u,\ 3)+\ IMG_{CIPHER}(i,\ 2)+\ IMG_{CIPHER}\ (u,\ 1)+\ SnapS_2{}^{*}(u,\ RP_{+VE}\ )+\ floor\ (Row_1\ (u,\ 2)\times 2^{14}\ ),\ A_X),\ v=\ 3\ ] \qquad Eq(17)$$

$$IMG_{CIPHER}(u,2)=mod(SnapS_1(u,\ 2)+\ IMG_{CIPHER}\ (u,1)+SnapS_1\ (u,RP_{+VE}-\ 1)+SnapS_2{}^{*}(u,\ RP_{+VE}-1)+floor(Row_1(u,\ 2)\times\ 2^{14}\ )\ ,\ A_X),\ v=2\ ] \qquad Eq(18)$$

$$IMG_{CIPHER}\ (u,1)=mod(SnapS_1(u,\ 1)+SnapS_2(i,\ RP_{+VE})+SnapS_1(u,RP_{+VE}-1)+SnapS_2{}^{*}(u,RP_{+VE}-2)+floor\ (Row_1\ (u,\ 1)\times\ 2^{14}),\ A_X),v=1] \qquad Eq(19)$$

Step 2. : Obtain the isolated pixel matrices as, encryption image, *$SnapS_1$* and data image *$SnapS_2$*, Convert these images, which are in 2-Dimensional plane into their previous-3-Dimensional Cube forms.

Step 3. : De-scramble the cube using the de-scrambling algorithm and obtain the original cubes as, *$1st\_RP_{CUBE}$* and *$2nd\_RP_{CUBE}$*.

Step 4. : Read the data from the *$2nd\_RP_{CUBE}$* using the starting point, A_X and re-order the data into a stream of bits.

Step 5. : Convert the stream of bits back to User(s) readable format. Send the acknowledgement to the cipher-Image (*IMG_{CIPHER}') S*ender node "*$1P_{NODE/MINER}$*".

Step 6. : Terminate Peer/Node connection from one IoT edge device(*$1P_{NODE/MINER}$*) with the receiving IoT edge device(*$2P_{NODE/MINER}$*) after successful completion of data transfer.

NOTE: Scrambling and Descrambling algorithms are user-defined and purely depends on the programmer(s), so as which out of 2^{11}x 3^7 x 8! x 12!/2= 43252003274489856000 combinations, the programmer selects certain block arrangements, which will be treated as valid and prepare its reversible order for the descrambling algorithm.

4 THE RCC-BLOCK-MASSK MODEL (PHASE 3) : PROPOSED SECURE SESSION KEY GENERATION MECHANISM FOR IOT PEER(S) AND THEIR MUTUAL AUTHENTICATION

In our proposed model, Internet of things's integration with smart heterogeneous technologies and its network driving dependencies over blockchain not only reduces the risk of single point of failure but also increases the security strength and understanding among IoT network nodes/peers to mutually authenticate each other. In our "*RCC-Block_MASSK model*", Blockchain(consensus algorithm) can be executed over the "resource-capable" devices like the heavier gateways(*GW_{NODE}*), servers and the intermediate cluster handling workstation(s). Parameters can be passed from one peer-2-peer securely using the Rubik's cube based crypto-system but the IoT network can also have certain devices with minimal resources and computation power. It becomes really difficult to execute even the proposed light-weight cryptosystem in section 2 of this paper. Edge devices such as basic sensing devices or the radio-frequency operable RFID tags, motion or heat sensors etc. Such edge layer devices are found to be consuming heavy battery power leaving no room for crypt-system based "authentication techniques" [73,73,75] to work on such low levelled devices, which were feasible on the comparatively heavier nodes like Arduino UNO[76], Raspberry Pie 4 or Higher[77] and Nvidia Jetson Xavier/Nano boards[78]. No matter how lightweight a crypto-system is, there are always certain issues on their practical real time im-plementation on such low levelled edge devices. Authentication is a necessary condition for data privacy because after successfully authenticating the devices in a network, then only the “Authorization, access and Data rights” can be assigned to these network nodes. So, clearly there is a need for authenticating the devices[79] in order to create “Trust” factor among IoT network peers before they can actually send the data back and forth. To do so, we propose a lightweight session key generation[section 4] and key management framework for low level IoT edge devices, which is already incorporated in our RCC-Block_MASSK model as the (Phase 4) [section 5]. Figure 3 shows the major steps and process flow of IoT device authentication among low levelled edge devices.

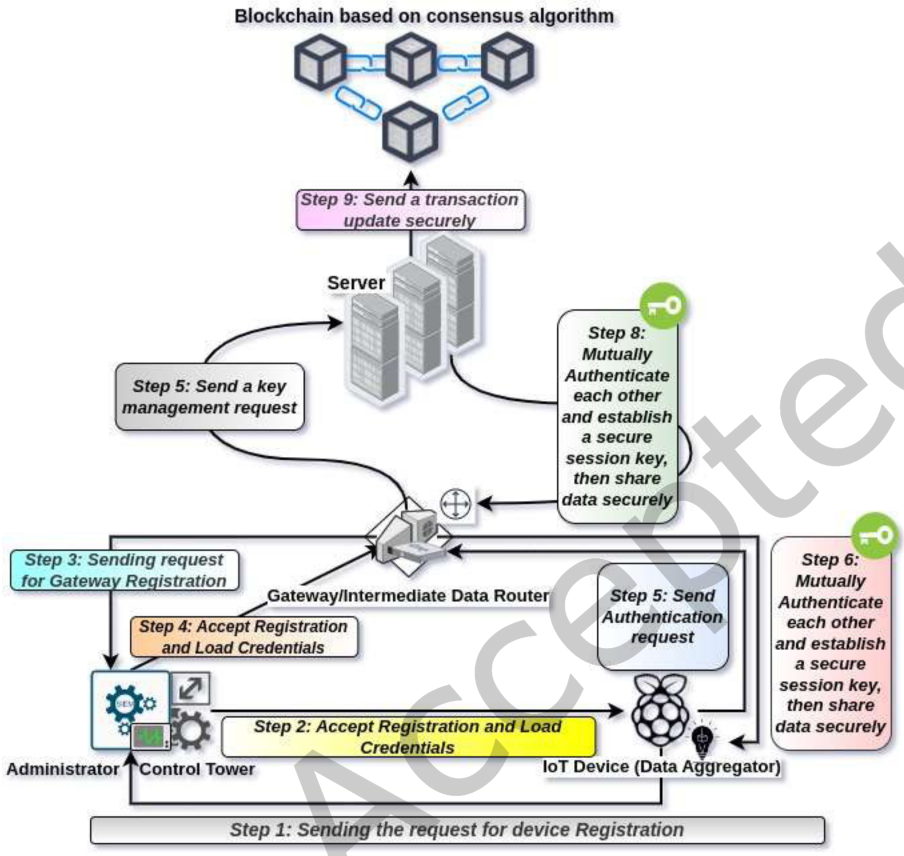

Blockchain based on consensus algorithm
Step 9: Send a transaction update securely
Server
Step 8: Mutually Authenticate each other and establish a secure session key, then share data securely
Step 5: Send a key management request
Step 3: Sending request for Gateway Registration
Gateway/Intermediate Data Router
Step 5: Send Authentication request
Step 6: Mutually Authenticate each other and establish a secure session key, then share data securely
Step 4: Accept Registration and Load Credentials
Step 2: Accept Registration and Load Credentials
Administrator
Control Tower
IoT Device (Data Aggregator)
Step 1: Sending the request for device Registration

Figure 3. Major Process Flow steps of IoT device authentication among low levelled edge devices

The edge device, IoT_{EDGE}, initially sends the request for its registration to the administrative control tower, CT_{Admin} . The control tower(CT_{Admin}) verifies and accepts the registration request, executes the process and loads the credentials back to the device, IoT_{EDGE}. GW_{NODE} also needs to send a registration request to the administrative control room/tower/station, CT_{Admin} and in similar fashion, the request is verified and the credentials are sent to the respective requesting node using secure RCC-Rubik’s cube cryptosystem. With gateways, it is possible to use cryptosystems but for low level devices, the authentication mechanism is quite different. The IoT Edge devices are called the data aggregators which collect continuous data periodically and send this data to the data storing servers using the intermediate, data passing gateways. For this to work, IoT_{EDGE} establishes a secure communication session after mutually authenticating each other using a secure session key(meant for low levelled edge devices) with the GW_{NODE}. If the mutual authentication is authentic and valid, the secure session is established successfully and data can be transferred from peer-to-peer(s) or peer-to-gateway(s). Further, the data is collected from all the gateway nodes, GW_{NODE1}, GW_{NODE2}, GW_{NODE3}.......GW_{NODEN} and stored at a backup server. Blockchain on their hand parallelly synchronizes as per the global clock to deliver on-time services and updated transactions in the ledger. The Secure Session Key generation process is complex and it is time saving if the pre-generated session keys are recycled and stored at the server end. We utilize the Rubik’s cube mechanism again to store the session keys in a N-Order cube at the server that keeps the session keys in the scrambled form, safe-guarding it from the possible data theft and SQL-injection[80] attacks. Rubik’s Cube logic based session key randomizer (RCR_{KEY}) is discussed in section 4. Blockchain has the property of immutability that ensures that no contents of the blocks can be tempered by any means by the attacker(s). This property ensures the guarantee and security needed to store the credentials in the blockchain alongwith the transactional records. Whether there are credentials or data, everything is verified using the *POLWSCT* consensus algorithm before adding any of it to the Blockchain's blocks. Following setup and steps illustrate the secure session key generation and key management process:

4.1 Initialization and Environment setup:

The command control, CT_{Admin} selects all the necessary parameters as follows:

		Initializing setup:
Step 1.	:	CT_{Admin} Randomly chooses a number "RP_{+VE}", converts it into binary form and stores it into the RP_{+VE} *-Order* Cube, calls the scrambling algorithm, *Prime_Scramble_Cube( )* and reads the bits again. *Prime_Scramble_Cube( )* function scrambles the bits in such a way that only a prime number is generated after scrambling. This alters the bits orientation and if we read it now, it will result in a new prime number, $new\_\ RP_{+VE}$. For example, if bits were, "00100101" in binary, which means "37" decimal system, then after calling function, *Prime_Scramble_Cube( )*, say, we receive "10000011", which corresponds to "131", then the newly randomly picked prime number will be "131", in decimal system. It is to be noted here that the cube mechanism is being used as a bit-shuffler and not for cryptographic key generation. Here, $RP_{+VE} \in Z\, q_{+} = \{1, 2, \ldots N\}$. RP_{+VE} is a small prime but sufficiently large number.
Step 2.	:	The CT_{Admin} selects a lesser interference and collision based hashing algorithm, *h( )*, such that $h(\,) \in SHA\text{-}128$ bit length algorithm.
Step 3.	:	The CT_{Admin} creates a small private key, SP_{KEY} for low levelled edge devices such that, the respective public key, $PK_{KEY} = SP_{KEY} * new\_\ RP_{+VE} * PR_{GENERATOR}$, where $PR_{GENERATOR}$ is the ambiguity generator used to mix the secret parameters. CT_{Admin} stores "$SP_{KEY} * new\_\ RP_{+VE}$" as its private key, and broadcasts publicly Vector, $V = [\, PK_{KEY}, h(\,)\ ]$ is shared pubicly over a wireless network and even if the attacker captures the vector, it would not be able to figure out the parameters on the basis of which the public key was computed, which are SP_{KEY}, $new\_\ RP_{+VE}$ and $PR_{GENERATOR}$.

4.2 IoT Network Component(s) Registration Procedure

4.2.1 This process is exhibited at the administrative control tower/station CT_{Admin} in order to register the server station, S_{SERVER} storing the data and session keys, gateway(s) GW_{NODEN}, even the edge devices, IoT_{EDGE}. Figure 4. Describes the parameter exchange among CT_{Admin} and S_{SERVER},

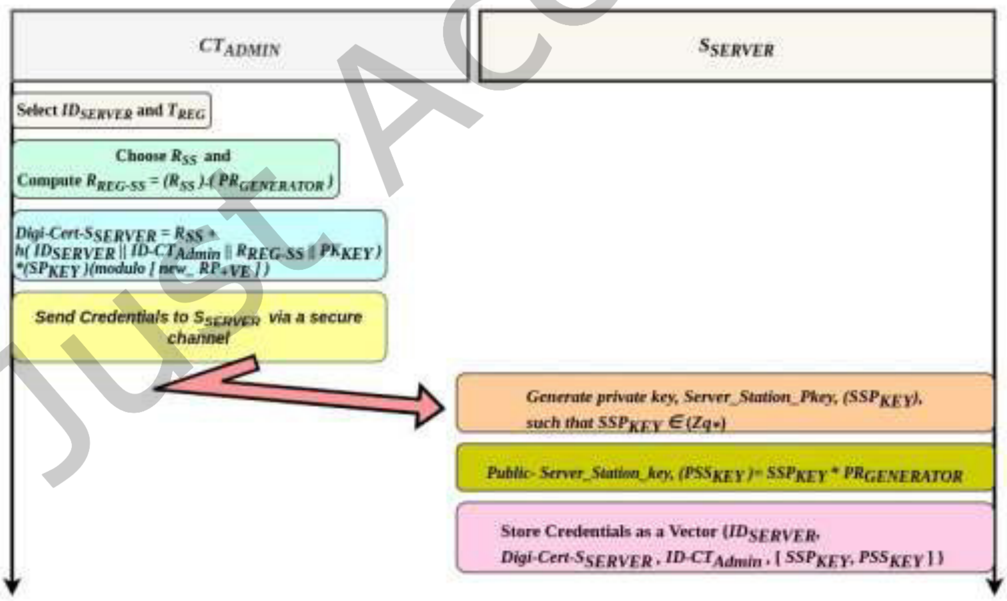

CT_ADMIN
S_SERVER
Select ID_SERVER and T_REG
Choose R_SS and
Compute R_REG-SS = (R_SS).(PR_GENERATOR)
Digi-Cert-S_SERVER = R_SS +
h(ID_SERVER || ID-CT_Admin || R_REG-SS || PK_KEY)
*(SP_KEY)(modulo [new_ RP_+VE])
Send Credentials to S_SERVER via a secure channel
Generate private key, Server_Station_Pkey, (SSP_KEY),
such that SSP_KEY ∈(Zq*)
Public- Server_Station_key, (PSS_KEY)= SSP_KEY * PR_GENERATOR
Store Credentials as a Vector (ID_SERVER,
Digi-Cert-S_SERVER , ID-CT_Admin , [SSP_KEY, PSS_KEY])

Figure 4. Parameter exchange among Administrative Control Tower/Station, CT_{Admin} and Server Station, S_{SERVER}

Step 1. CT_{Admin} Selects ID_{SERVER} and T_{REG} for the server station, where ID_{SERVER} is the distinctive singular identification assigned number whereas "T_{REG}" is the instantaneous time at which the server station got registered by the CT_{Admin}.

Step 2. CT_{Admin} produces arbitrarily a nonce as "R_{SS}" $\in Z\ q_{*}$ and also create the alternate $R_{REG\text{-}SS} = (R_{SS}).(PR_{GENERATOR})$ for the S_{SERVER}-Server station. Let ID for Control Tower/Station be "$ID\text{-}CT_{Admin}$". For authentication and verification purposes, CT_{Admin} constitutes the digitally verifiable certificate for S_{SERVER}, namely, $Digi\text{-}Cert\text{-}S_{SERVER} = R_{SS} + h(\, ID_{SERVER} \,||\, ID\text{-}CT_{Admin} \,||\, R_{REG\text{-}SS} \,||\, PK_{KEY}\,) * (SP_{KEY})(modulo\ [\ new\_\ RP_{+VE}\ ]\,)$.

Step 3. On receiving the credentials generated after successful registration process from CT_{Admin}, *the* S_{SERVER} secretly creates its individual private key, *Server_Station_Pkey,* (SSP_{KEY}), such that $SSP_{KEY} \in (Zq_{*})$. Its respective key for public use by other IoT edge devices for cipher-packet decryption purposes is computed as, *Public-Server_Station_key,* $(PSS_{KEY}) = SSP_{KEY} * PR_{GENERATOR}$. After the keys are computed successfully, the registration affiliated credentials $\{ID_{SERVER}, Digi\text{-}Cert\text{-}S_{SERVER}, ID\text{-}CT_{Admin}, [SSP_{KEY}, PSS_{KEY}]\}$ are sent and stored in the memory of the S_{SERVER}.

4.2.2 Low Levelled IoT Edge Device Registration Phase

Before deploying the edge devices to the physical surrounding for data gathering or analysis purposes, these IoT edge devices are registered at the CT_{Admin} using following:

Step 1. CT_{Admin} generates the distinct and unique $ID\text{-}IoT_{EDGE(i)}$ for every ith low levelled resource constrained IoT edge device(s) trying to register at a particular point of time. It further computes the pseudo-ID as $PID_{EDGE(i)} = h(ID\text{-}IoT_{EDGE(i)} \| SP_{KEY} \| T_{REG\text{-}EDGE(i)})$ and arbitrarily *ID* as $A_{ID\text{-}EDGE(i)}$, for each ith edge device being registered at time $T_{REG\text{-}EDGE(i)}$.

Step 2. CT_{Admin} produces a small private key as $SP_{IoT\text{-}EDGE(i)}$ and the required key for public distribution is calculated as, $PK_{IoT\text{-}EDGE(i)} = SP_{KEY} * PR_{GENERATOR}$, and distribute this public key across the networking cluster by directed multicasting or broadcasting.

Step 3. CT_{Admin} feeds the secure parameters created in steps 1 and 2 as vectors, $V_1 = \{A_{ID\text{-}EDGE(i)}, PID_{EDGE(i)}\}$, $V_2 = \{SP_{IoT\text{-}EDGE(i)}, PK_{IoT\text{-}EDGE(i)}\}$. Send these vectors to the assigned cluster head(s) for monitoring and record updation purposes. Figure 4 shows the parameter and vector exchange among CT_{Admin} and IoT_{EDGE} components of an IoT network.

4.2.3 Gateway Node(s) and Dynamic IoT (Dyn-IoT$_{EDGE}$) device Registration Phase

In order to assign service(s) and cluster managing responsibilities to the respective gateway(s) GW_{NODE_N} for the *Nth* cluster in an IoT network,

Step 1. CT_{Admin} Selects gateway unique *ID*, $GW\text{-}ID_{NODE\_N}$ and $T_{GW\text{-}REG}$, where $GW\text{-}ID_{NODE\_N}$ is the distinctive singular identification assigned number whereas "$T_{GW\text{-}REG}$" is the instantaneous time at which the gateway node got registered by the CT_{Admin} . For dynamic edge nodes which are motion capable while staying connected with the cluster wirelessly, CT_{Admin} generates $Dyn\text{-}ID_{IoT\text{-}EDGE}$ and its corresponding timestamp, $T_{REG\text{-}DYN}$. CT_{Admin} generates the pseudo-ID for gateway(s), as $GW\text{-}ID_{PSEUDO} = h(GW\text{-}ID_{NODE\_N} \| SP_{KEY} \| T_{GW\text{-}REG})$, and for Dynamic IoT edge nodes as , $Dyn\text{-}ID_{PSEUDO} = h(Dyn\text{-}ID_{IoT\text{-}EDGE} \| SP_{KEY} \| T_{REG\text{-}DYN})$. Here, "$SP_{KEY}$" is the small sized private key generated by the administrative control tower/station in Step 3 of section 4.1. earlier. CT_{Admin} also picks randomly the temporary IDs for gateways and dynamic nodes as, *TG-ID* and *Tdyn-ID* respectively.

Step 2. ***(i)*** CT_{Admin} produces arbitrarily a nonce for gateway node as "GWR_{SS}" $\in Z\, q_{*}$ and also create the alternate $GWR_{REG\text{-}SS} = (GWR_{SS}).(PR_{GENERATOR})$ for the secure communication of the gateway nodes. For authentication and verification purposes, CT_{Admin} constitutes the digitally verifiable certificate for GW_{NODE_N}, namely, $Digi\text{-}Cert\text{-}GW_{NODE\_N} = GWR_{SS} + h(GW\text{-}ID_{NODE\_N} \| ID\text{-}CT_{Admin} \| GWR_{REG\text{-}SS} \| PK_{KEY}) * (SP_{KEY})(modulo\ [new\_\ RP_{+VE}]\ )$, where, PK_{KEY}, SP_{KEY} and $new\_\ RP_{+VE}$ are already defined in section 4.1.

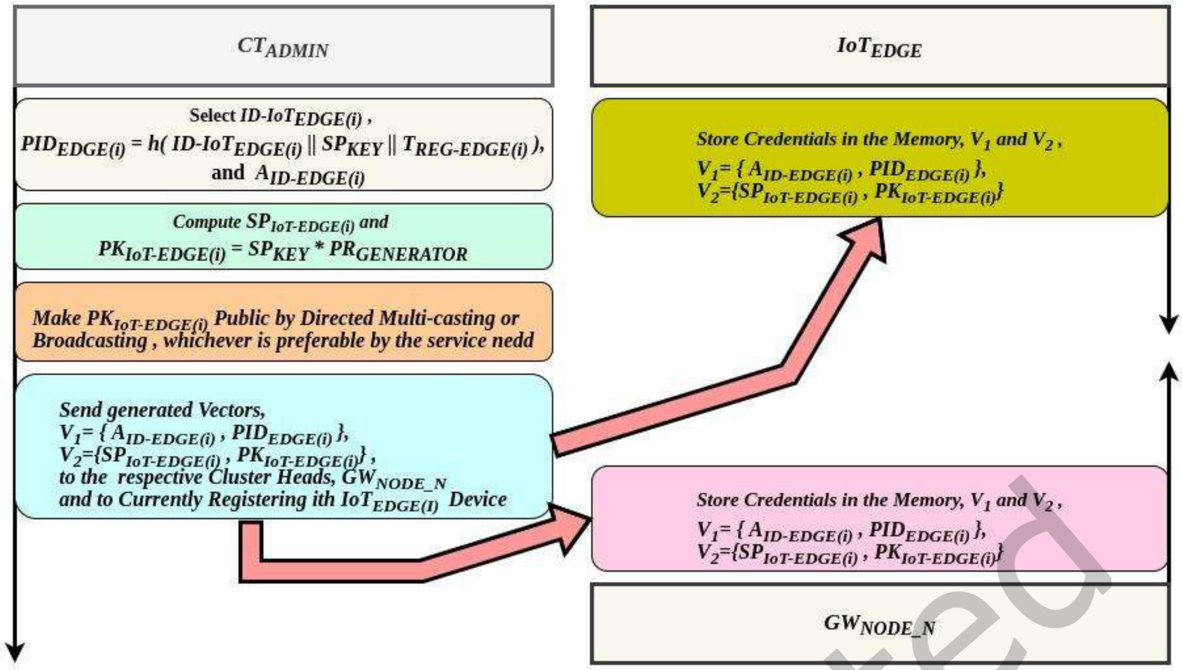

CT_{ADMIN}
IoT_{EDGE}
Select $ID\text{-}IoT_{EDGE(i)}$, $PID_{EDGE(i)} = h(ID\text{-}IoT_{EDGE(i)} \| SP_{KEY} \| T_{REG\text{-}EDGE(i)})$, and $A_{ID\text{-}EDGE(i)}$
Compute $SP_{IoT\text{-}EDGE(i)}$ and $PK_{IoT\text{-}EDGE(i)} = SP_{KEY} * PR_{GENERATOR}$
Make $PK_{IoT\text{-}EDGE(i)}$ Public by Directed Multi-casting or Broadcasting , whichever is preferable by the service nedd
Send generated Vectors, $V_1 = \{ A_{ID\text{-}EDGE(i)} , PID_{EDGE(i)} \}$, $V_2 = \{SP_{IoT\text{-}EDGE(i)} , PK_{IoT\text{-}EDGE(i)}\}$, to the respective Cluster Heads, GW_{NODE_N} and to Currently Registering ith $IoT_{EDGE(I)}$ Device
Store Credentials in the Memory, V_1 and V_2 , $V_1 = \{ A_{ID\text{-}EDGE(i)} , PID_{EDGE(i)} \}$, $V_2 = \{SP_{IoT\text{-}EDGE(i)} , PK_{IoT\text{-}EDGE(i)}\}$
Store Credentials in the Memory, V_1 and V_2 , $V_1 = \{ A_{ID\text{-}EDGE(i)} , PID_{EDGE(i)} \}$, $V_2 = \{SP_{IoT\text{-}EDGE(i)} , PK_{IoT\text{-}EDGE(i)}\}$
GW_{NODE_N}

Figure 5. Parameter exchange among Administrative Control Tower/Station CT_{Admin} , IoT_{EDGE} devices and GW_{NODE_N}

Step 2. ***(ii)*** Similar process is repeated for Low levelled IoT edge devices and respective IDs and private public keys are generated as, CT_{Admin} produces arbitrarily a nonce for the dynamically motion capable Low levelled edge node as "$Dyn\text{-}R_{SS}$" $\in Z\ q_*$ and also create the alternate $DynR_{REG\text{-}SS}$ = ($Dyn\text{-}R_{SS}$).($PR_{GENERATOR}$) for the secure communication of the resource constrained edge nodes. For Mutual-Authentication and verification purposes, CT_{Admin} produces a uniquely identifiable digital certificate, for $Dyn\text{-}IoT_{EDGE}$, namely, $Digi\text{-}Cert\text{-}Dyn\text{-}IoT_{EDGE} = Dyn\text{-}R_{SS} + h(Dyn\text{-}ID_{IoT\text{-}EDGE} \| ID\text{-}CT_{Admin} \| DynR_{REG\text{-}SS} \| PK_{KEY})^*(SP_{KEY})(modulo\ [\ new\_\ RP_{+VE}\ ]\ )$, where, PK_{KEY} , SP_{KEY} and $new\_\ RP_{+VE}$ are already defined in section 4.1.

Step 3. ***(i)*** On receiving the credentials generated after successful registration process from CT_{Admin}, the GW_{NODE_N} secretly creates its individual private key, GW_Pkey, $(GWSP_{KEY})$, such that $GWSP_{KEY} \in (Zq_*)$. Also for public key use by other IoT edge devices and other communicating gateways, for cipher-packet decryption purposes, corresponding public key is computed as, $Public\text{-}GWkey$, $(PubGW_{KEY}) = GWSP_{KEY} * PR_{GENERATOR}$. After the keys are computed successfully, the registration affiliated credentials $\{ GW\text{-}ID_{NODE\_N}, Digi\text{-}Cert\text{-}\ GW_{NODE\_N}$, $ID\text{-}CT_{Admin}$, $[\ GWSP_{KEY}, PubGW_{KEY}\ ]\ \}$ are sent and stored in the memory of the GW_{NODE_N}. Similarly, for dynamic IoT edge devices with low level resources, encryption and decryption key pair is generated through CT_{Admin} as well.

Step 3. ***(ii)*** CT_{Admin} sends the necessary parameters to the Dynamic device(s), $Dyn\text{-}IoT_{EDGE}$ and on receiving the credentials, $Dyn\text{-}IoT_{EDGE}$ produces a private key, $Dyn\text{-}P_{KEY}$ $\in (Zq_*)$. Dyn-IoT_{EDGE} device uses this private key for encrypting the data to be sent during information exchange. Also, for decrypting the data from the the cipher packet, $Dyn\text{-}IoT_{EDGE}$ generates public key, $Pub\text{-}Dyn\text{-}IoT_{EDGE} = Dyn\text{-}P_{KEY} * PR_{GENERATOR}$. After the keys are computed successfully, the registration affiliated credentials $\{ Dyn\text{-}ID_{IoT\text{-}EDGE}, Digi\text{-}Cert\text{-}\ Dyn\text{-}IoT_{EDGE}$, $ID\text{-}CT_{Admin}$, $[\ Dyn\text{-}P_{KEY}, Pub\text{-}Dyn\text{-}IoT_{EDGE}\ ]\ \}$ are sent and stored in the memory of the $Dyn\text{-}IoT_{EDGE}$. Figure 6 shows the parameter exchange among the Control Tower, Dynamic low levelled devices and the gateway cluster head.

4.2.4 Secret Session Key Generation and Mutual Authentication: For All Peer(s), Gateway(s) and other Components of IoT Network

The steps for mutual authentication and secret session key generation among communicating and data transferring components as ahead:

Step 1. There are broadly two variants of edge devices that operate at the edge layer which are responsible for data collection and transmission, Dynamic plus resource constrained $Dyn\text{-}IoT_{EDGE}$ and the other are simple $IoT_{EDGE(i)}$ devices. Any device, operating at the edge layer creates an arbitrarily chosen number, $a_{EDGE(i)} \in (Zq_*)$ and a needed timestamp, TS_{EDGE1} for global clock synchronization. Then it calculates for every ith node,

a. For Simple Gateway/Edge Device	:	$A_{EDGE(i)} = h\ (a_{EDGE(i)} \Vert TG\text{-}ID \Vert GW\text{-}ID_{PSEUDO} \Vert GWSP_{KEY} \Vert TS_{EDGE1}),$
b. For Dynamic device	:	$A_{EDGE(i)} = h\ (a_{EDGE(i)} \Vert Tdyn\text{-}ID \Vert Dyn\text{-}ID_{PSEUDO} \Vert Dyn\text{-}P_{KEY} \Vert TS_{EDGE1}),$

It further computes the digital signature over $a_{EDGE(i)}$ as,

c. For Simple Gateway/Edge Device:	:	$Sig_{EDGE(i)} = h(a_{EDGE(i)} \Vert TG\text{-}ID \Vert GW\text{-}ID_{PSEUDO} \Vert GWSP_{KEY} \Vert TS_{EDGE1}) + h(PK_{IoT\text{-}EDGE(i)} \Vert PubGW_{KEY} \Vert PSS_{KEY} \Vert TS_{EDGE1})^{*}(GWSP_{KEY}) \ (mod\ (new\_\ RP_{+VE})).$
d. For Dynamic device:	:	$Sig_{EDGE(i)} = h(a_{EDGE(i)} \Vert Tdyn\text{-}ID_{EDGE} \Vert Dyn\text{-}ID_{PSEUDO} \Vert Dyn\text{-}P_{KEY} \Vert TS_{EDGE1}) + h(Pub\text{-}Dyn\text{-}IoT_{EDGE}\ or\ PK_{IoT\text{-}EDGE(i)} \Vert PubGW_{KEY} \Vert PSS_{KEY} \Vert TS_{EDGE1})^{*}(Dyn\text{-}P_{KEY}) \ (mod\ (new\_\ RP_{+VE}))$

Note: $PubGW_{KEY}$ is utilized in both the equations because, no matter what the device is, it has to communicate with its assigned cluster head(Gateway node).

After computing the necessary parameters, $IoT_{EDGE(i)}$ / $Dyn\text{-}IoT_{EDGE}$, sends the message,

M_1	=	$\{ Tdyn\text{-}ID_{EDGE}\ or\ TG\text{-}ID , A_{EDGE(i)} , Sig_{EDGE(i)} , TS_{EDGE1} \},$ to the intermediate $GW_{NODE\ N(i)}$, through a public channel.

Step 2. After receiving the vector, Message, M$_1$, at a time TS_{EDGE2}, and verifies the time difference as, $| TS_{EDGE1} - TS_{EDGE2} | < \Delta TS_{EDGE}$. If a valid time difference is evaluated, then, verification of digital signature is done as,

$(Sig_{EDGE(i)}).(PR_{GENERATOR})$	=	$\mathrm{A}_{\mathrm{EDGE(i)}} + h(Pub\text{-}Dyn\text{-}IoT_{EDGE}\ or\ PK_{IoT\text{-}EDGE(i)} \Vert PubGW_{KEY} \Vert PSS_{KEY} \Vert TS_{EDGE1}).\ (Pub\text{-}Dyn\text{-}IoT_{EDGE})$

If it evaluates as valid, then GW_{NODE_N} creates an instantaneous time-stamp, TS_{GW1} and an arbitrarily chosen number, a_{GW} such that $a_{GW} \in (Zq_{*})$ and computes,

A_{GW}	=	$h\ (a_{GW} \Vert A_{ID\text{-}EDGE(i)} \Vert GW\text{-}ID_{PSEUDO} \Vert SP_{IoT\text{-}EDGE(i)} \Vert TS_{GW1}).(PR_{GENERATOR}),$

Further, GW_{NODE_N} computes mutual authentication parameter between gateway and IoT network node, $MA_{(EDGE, GW)}$, *as,*

From Edge device to Gateway/Dyn-IoT$_{Edge}$,

$MA_{(EDGE, GW)}$	=	$h\ (a_{GW} \Vert TG\text{-}ID \Vert GW\text{-}ID_{PSEUDO} \Vert GWSP_{KEY} \Vert TS_{GW1}).(A_{EDGE(i)})$
Also Secret Session Key, as SSK_{KEY} ,		
$SSK_{KEY}(EDGE, GW)$	=	$h\ (MA_{(EDGE, GW)} \Vert Sig_{EDGE(i)} \Vert TS_{EDGE1} \Vert TS_{GW1}) ,$

Thereafter, the gateway node GW_{NODE_N} and intermediate resource capable device, IoT_{EDGE} formulate the digital-signature on parameter a_{GW} and SSK_{KEY} *(EDGE, GW) as,*

For Simple Gateway/Edge Device:

Sig_{GW1}	=	$= h(a_{GW} \Vert TG\text{-}ID \Vert GW\text{-}ID_{PSEUDO} \Vert GWSP_{KEY} \Vert TS_{GW1}) + h(SSK_{KEY}(EDGE, GW) \Vert Pub\text{-}Dyn\text{-}IoT_{EDGE}\ or\ PK_{IoT\text{-}EDGE(i)}\ or\ PubGW_{KEY} \Vert PSS_{KEY} \Vert TS_{GW1})^{*}\ (SP_{IoT\text{-}EDGE(i)})(mod\ (new\_\ RP_{+VE}))$

For Dynamic device:

Sig_{GW1}	=	$= h(a_{GW} \Vert Tdyn\text{-}ID_{EDGE} \Vert Dyn\text{-}ID_{PSEUDO} \Vert Dyn\text{-}P_{KEY} \Vert TS_{GW1}) + h(SSK_{KEY}(EDGE, GW) \Vert Pub\text{-}Dyn\text{-}IoT_{EDGE}\ or\ PK_{IoT\text{-}EDGE(i)} \Vert PSS_{KEY} \Vert TS_{GW1})^{*}\ (Dyn\text{-}P_{KEY})(mod\ (new\_\ RP_{+VE}))$

Gateway node, GW_{NODE_N} computes a newer arbitrary ID for IoT_{EDGE} as "*new-TID$_{EDGE}$*", and further computes the parameter,

$new\text{-}TID^{*}_{EDGE}$	=	$new\text{-}TID_{EDGE} \oplus h(Tdyn\text{-}ID_{EDGE} \Vert SSK_{KEY}(EDGE, GW) \Vert Sig_{GW1} \Vert TS_{GW1}) ,$

GW_{NODE_N} transmits the newer message, $M_2 = \{ new\text{-}TID^*_{EDGE}, A_{GW}, Sig_{GW1}, TS_{GW1} \}$ through the public channel to IoT_{EDGE}.

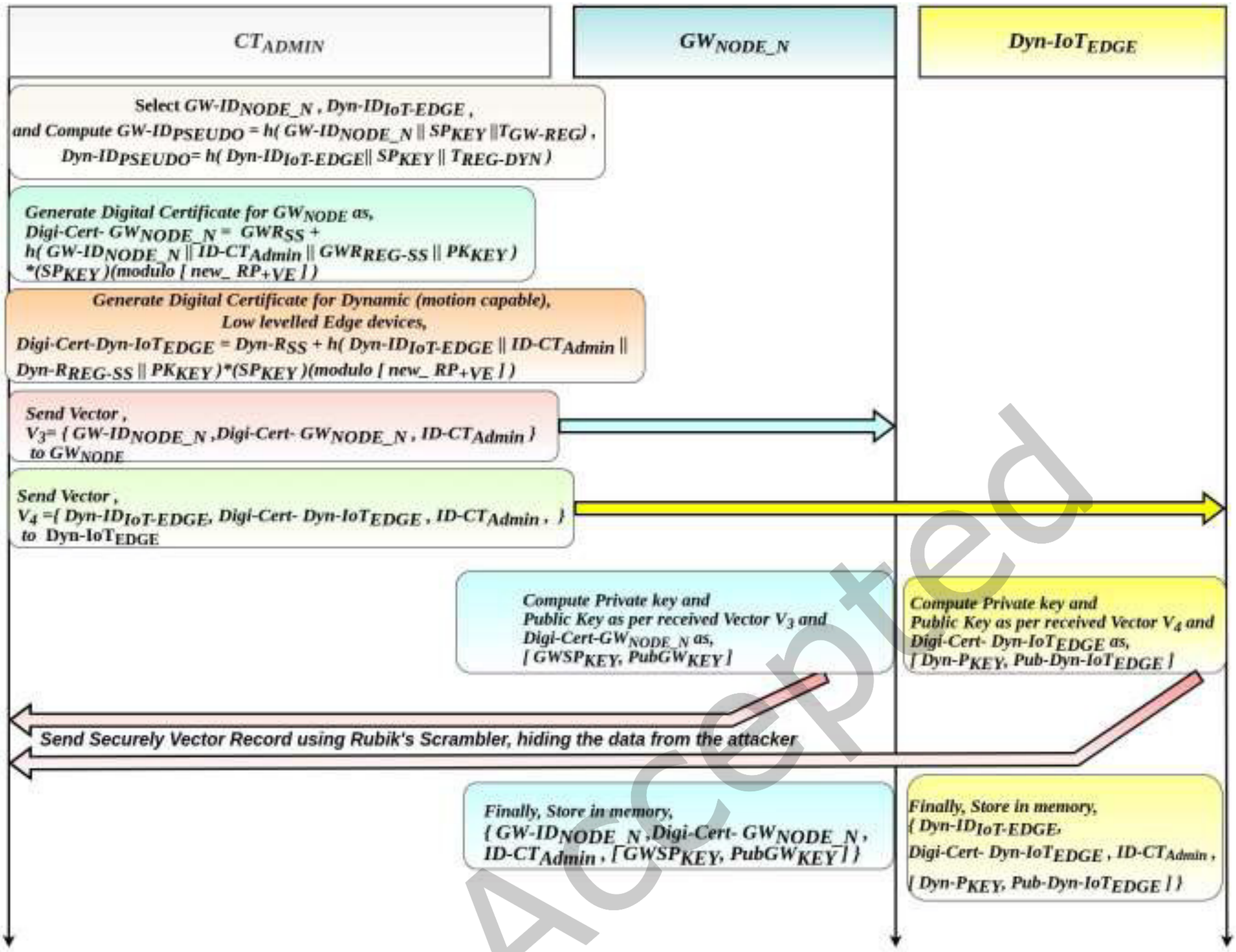

CT_ADMIN
GW_NODE_N
Dyn-IoT_EDGE
Select GW-ID_NODE_N , Dyn-ID_IoT-EDGE ,
and Compute GW-ID_PSEUDO = h(GW-ID_NODE_N || SP_KEY ||T_GW-REG) ,
Dyn-ID_PSEUDO= h(Dyn-ID_IoT-EDGE|| SP_KEY || T_REG-DYN)
Generate Digital Certificate for GW_NODE as,
Digi-Cert- GW_NODE_N = GWR_SS +
h(GW-ID_NODE_N || ID-CT_Admin || GWR_REG-SS || PK_KEY)
*(SP_KEY)(modulo [new_ RP+VE])
Generate Digital Certificate for Dynamic (motion capable),
Low levelled Edge devices,
Digi-Cert-Dyn-IoT_EDGE = Dyn-R_SS + h(Dyn-ID_IoT-EDGE || ID-CT_Admin ||
Dyn-R_REG-SS || PK_KEY)*(SP_KEY)(modulo [new_ RP+VE])
Send Vector ,
V3= { GW-ID_NODE_N ,Digi-Cert- GW_NODE_N , ID-CT_Admin }
to GW_NODE
Send Vector ,
V4 ={ Dyn-ID_IoT-EDGE, Digi-Cert- Dyn-IoT_EDGE , ID-CT_Admin , }
to Dyn-IoT_EDGE
Compute Private key and
Public Key as per received Vector V3 and
Digi-Cert-GW_NODE_N as,
[GWSP_KEY, PubGW_KEY]
Compute Private key and
Public Key as per received Vector V4 and
Digi-Cert- Dyn-IoT_EDGE as,
[Dyn-P_KEY, Pub-Dyn-IoT_EDGE]
Send Securely Vector Record using Rubik's Scrambler, hiding the data from the attacker
Finally, Store in memory,
{ GW-ID_NODE_N ,Digi-Cert- GW_NODE_N ,
ID-CT_Admin , [GWSP_KEY, PubGW_KEY] }
Finally, Store in memory,
{ Dyn-ID_IoT-EDGE,
Digi-Cert- Dyn-IoT_EDGE , ID-CT_Admin ,
[Dyn-P_KEY, Pub-Dyn-IoT_EDGE] }

Figure 6. Key Generation and Parameter exchange among CT_{Admin}, GW_{NODE_N} and $Dyn\text{-}IoT_{EDGE}$

Step 3. On receiving the message, M_2 at time TS_{GW2}, check whether the time difference is within the threshold limit, $| TS_{GW2} - TS_{GW1} | < \Delta TS_{GW}$. If it is successfully validated, then IoT_{EDGE} computes the mutual authentication parameter, $\boldsymbol{MA_{(EDGE, GW)}}$***, as,***

From Gateway/Dyn-IoT$_{Edge}$ to Edge device ,

$MA_{(GW, EDGE)}$	=	$h\,(a_{EDGE(i)} \,\|\, Tdyn\text{-}ID \,\|\, Dyn\text{-}ID_{PSEUDO} \,\|\, Dyn\text{-}P_{KEY} \,\|\, TS_{EDGE1}).(A_{GW})$

And the secret session key is computed as,

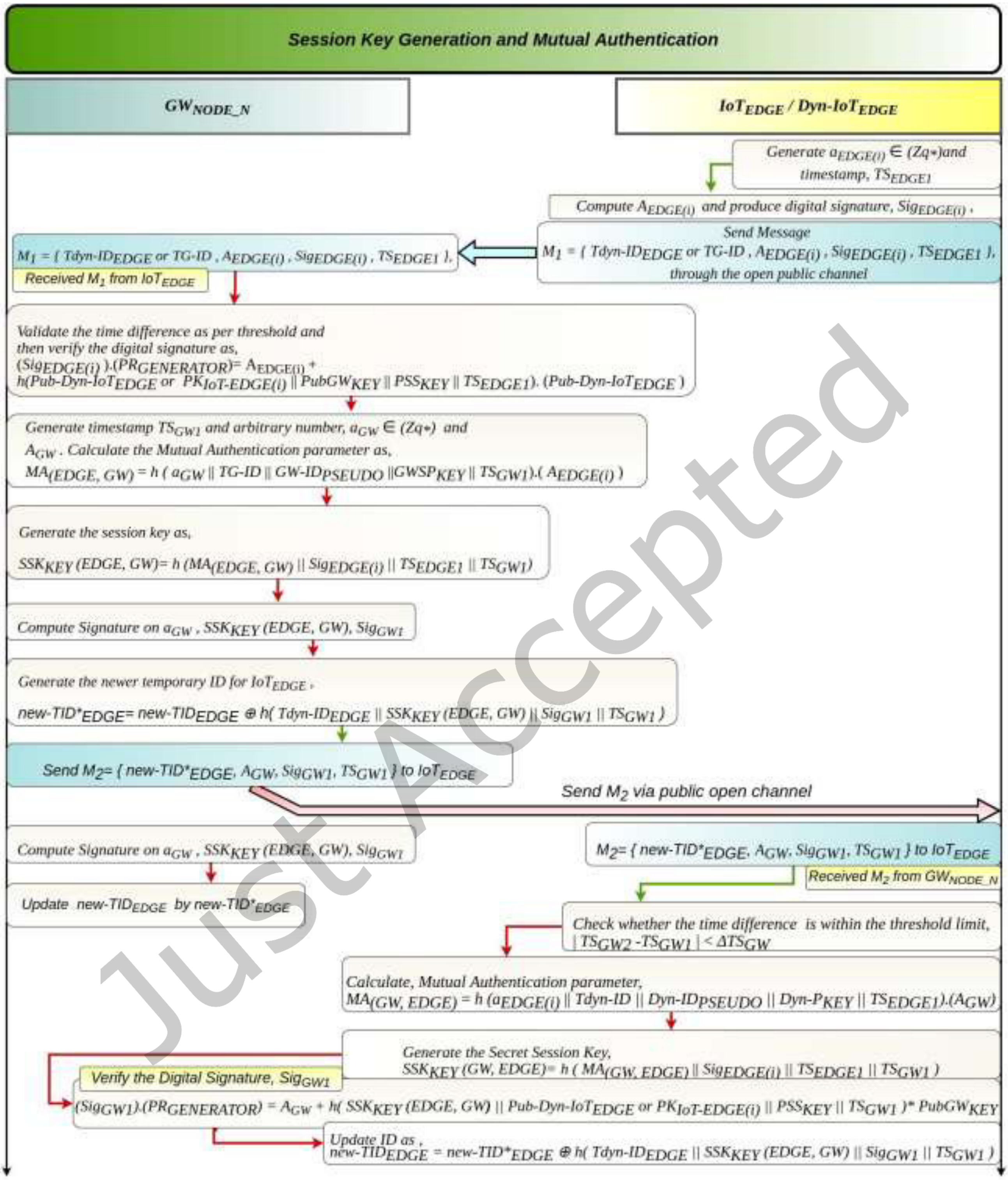
$SSK_{KEY}(GW, EDGE)$	=	$h\,(MA_{(GW, EDGE)} \,\|\, Sig_{EDGE(i)} \,\|\, TS_{EDGE1} \,\|\, TS_{GW1}),$

When the Secret Session Key is successfully generated, for a communicating pair of IoT network components, the digital signature verification is also necessary as,

$(Sig_{GW1}).(PR_{GENERATOR})$	=	$A_{GW} + h(SSK_{KEY}(EDGE, GW) \,\|\, Pub\text{-}Dyn\text{-}IoT_{EDGE}$ or $PK_{IoT\text{-}EDGE(i)} \,\|\, PSS_{KEY} \,\|\, TS_{GW1})^* PubGW_{KEY}.$

If the verification of the digital signature is successful, then the IoT_{EDGE} extracts $new\text{-}TID^*_{EDGE}$ from the received messages as,

new-TID$_{EDGE}$	=	*new-TID**$_{EDGE}$ ⊕ *h(Tdyn-ID*$_{EDGE}$ *‖ SSK*$_{KEY}$*(EDGE, GW) ‖ Sig*$_{GW1}$ *‖ TS*$_{GW1}$ *)*



Session Key Generation and Mutual Authentication
GW_NODE_N
IoT_EDGE / Dyn-IoT_EDGE
Generate a_EDGE(i) ∈ (Zq*)and timestamp, TS_EDGE1
Compute A_EDGE(i) and produce digital signature, Sig_EDGE(i) .
Send Message
M1 = { Tdyn-ID_EDGE or TG-ID , A_EDGE(i) , Sig_EDGE(i) , TS_EDGE1 },
through the open public channel
M1 = { Tdyn-ID_EDGE or TG-ID , A_EDGE(i) , Sig_EDGE(i) , TS_EDGE1 },
Received M1 from IoT_EDGE
Validate the time difference as per threshold and then verify the digital signature as,
(Sig_EDGE(i)).(PR_GENERATOR)= A_EDGE(i) + h(Pub-Dyn-IoT_EDGE or PK_IoT-EDGE(i) || PubGW_KEY || PSS_KEY || TS_EDGE1). (Pub-Dyn-IoT_EDGE)
Generate timestamp TS_GW1 and arbitrary number, a_GW ∈ (Zq*) and A_GW . Calculate the Mutual Authentication parameter as,
MA(EDGE, GW) = h (a_GW || TG-ID || GW-ID_PSEUDO ||GWSP_KEY || TS_GW1).(A_EDGE(i))
Generate the session key as,
SSK_KEY (EDGE, GW)= h (MA(EDGE, GW) || Sig_EDGE(i) || TS_EDGE1 || TS_GW1)
Compute Signature on a_GW , SSK_KEY (EDGE, GW), Sig_GW1
Generate the newer temporary ID for IoT_EDGE ,
new-TID*EDGE= new-TID_EDGE ⊕ h(Tdyn-ID_EDGE || SSK_KEY (EDGE, GW) || Sig_GW1 || TS_GW1)
Send M2= { new-TID*EDGE, A_GW, Sig_GW1, TS_GW1 } to IoT_EDGE
Send M2 via public open channel
Compute Signature on a_GW , SSK_KEY (EDGE, GW), Sig_GW1
Update new-TID_EDGE by new-TID*_EDGE
M2= { new-TID*EDGE, A_GW, Sig_GW1, TS_GW1 } to IoT_EDGE
Received M2 from GW_NODE_N
Check whether the time difference is within the threshold limit, | TS_GW2 -TS_GW1 | < ΔTS_GW
Calculate, Mutual Authentication parameter,
MA(GW, EDGE) = h (a_EDGE(i) || Tdyn-ID || Dyn-ID_PSEUDO || Dyn-P_KEY || TS_EDGE1).(A_GW)
Generate the Secret Session Key,
SSK_KEY (GW, EDGE)= h (MA(GW, EDGE) || Sig_EDGE(i) || TS_EDGE1 || TS_GW1)
Verify the Digital Signature, Sig_GW1
(Sig_GW1).(PR_GENERATOR) = A_GW + h(SSK_KEY (EDGE, GW) || Pub-Dyn-IoT_EDGE or PK_IoT-EDGE(i) || PSS_KEY || TS_GW1)* PubGW_KEY
Update ID as ,
new-TID_EDGE = new-TID*EDGE ⊕ h(Tdyn-ID_EDGE || SSK_KEY (EDGE, GW) || Sig_GW1 || TS_GW1)

Figure 7. The Secret Session Key generation and Mutual Authentication for all Peer(s), Gateway(s) and other components of IoT Network (GW$_{NODE\_N}$, IoT$_{EDGE}$, Dyn-IoT$_{EDGE}$)

Ultimately, the *IoT*$_{EDGE}$ device updates the { *Tdyn-ID or TG-ID* } *to new-TID*$_{EDGE}$ in the stored record in the device memory and the trust among each of the networking devices is developed after successful completion of Mutual Authentication. At Last, after all such steps are completed for every (u, v) peer pair of IoT networking devices, which intent to interact with each other or exchange the data between each other are able share data with each other using the common shared Secret Session Key *SSK*$_{KEY}$ as per the proposed authentication mechanism. Session Key generation and authentication are inter-related process which depend on one another for any two parties to share data securely. SSK_{KEY} is generated at the end of authentication and parameter passing based on digital signature verification. Figure 7 illustrates the secret session key generation and mutual authentication for all peer(s), gateway(s) and other components of IoT Network. The proposed RCC-Block_MASSK model is a three phase model, 4th phase is optional and storage oriented. First phase is where the system setup and initialization mode initiates the blockchain mechanism based on "*zvma*" token transaction system and *POLWSCT* consensus algorithm. 2nd phase establishes the proposed Rubik's cube based novel cryptosystem for higher security transmission of data among network peers. 3rd phase of the proposed model establishes the session key generation and mutual authentication mechanism for every possible device pair of an IoT network to develop a sense of "Trust" among each device. This helps in developing trust in order to share data. Our final phase of the model, 4th phase(optional because it depends on user to use it on the server or not) is the RCR_{KEY} session key store hub and randomizer to avoid the storage oriented attacks like, SQL-injection attacks[81], Information retrieval attacks[82,83], data duplicacy[84], data corruption[85], DOS and DDOS attacks[86][87] etc. Section 5 ahead discusses the algorithm of session key randomizer to avoid storage and guessing affiliated attacks.

5 RCC-BLOCK-MASSK MODEL (PHASE 4): SESSION KEY RANDOMIZER (RCR$_{KEY}$)

The session keys created in the phase 3 of the *RCC-Block_MASSK* model consume battery life and computation overhead on the network nodes. This battery loss cannot be restored on the same device which is already deployed at the data collecting site. The problem of rapid battery power drainage arises when the same device needs to generate another newer session key, SSK_{KEY}, every single time it needs to authenticate or transmit the data. To solve this problem in our proposed model, we introduce a session key recycler, which stores the pre-generated session keys along with their unique digital signatures at the server station, S_{SERVER}. RCR_{KEY} technique reduces time to create and alot a new session key to the requesting peer-pair *(u,v)* by allocating the already available session key stored in the key pool at the server, rather than creating a new one for the pair. Pre-created session keys can only be allotted to the requesting *(u,v)* peer pair, if there are some available session keys in the key pool stored at the server. In order to avoid SQL-injection, DOS-session keys and other information retrieval attacks, these session keys are stored in the scrambled form at the server. This ensures that the attacker is not able to directly read the session keys. RCR_{KEY} is a session key storing technique which breaks down the keys into pieces and scrambles it across a *Nth*-Order cube which also involves dummy cubes as well for generating the confusion factor C_F. Confusion factor, C_F is required to be high in order to evade Brute force attacks and key guessing attacks. Following is the pseudocode and diagramatic explaination in Figure 8 for how scrambling the stored keys at the server occurs:

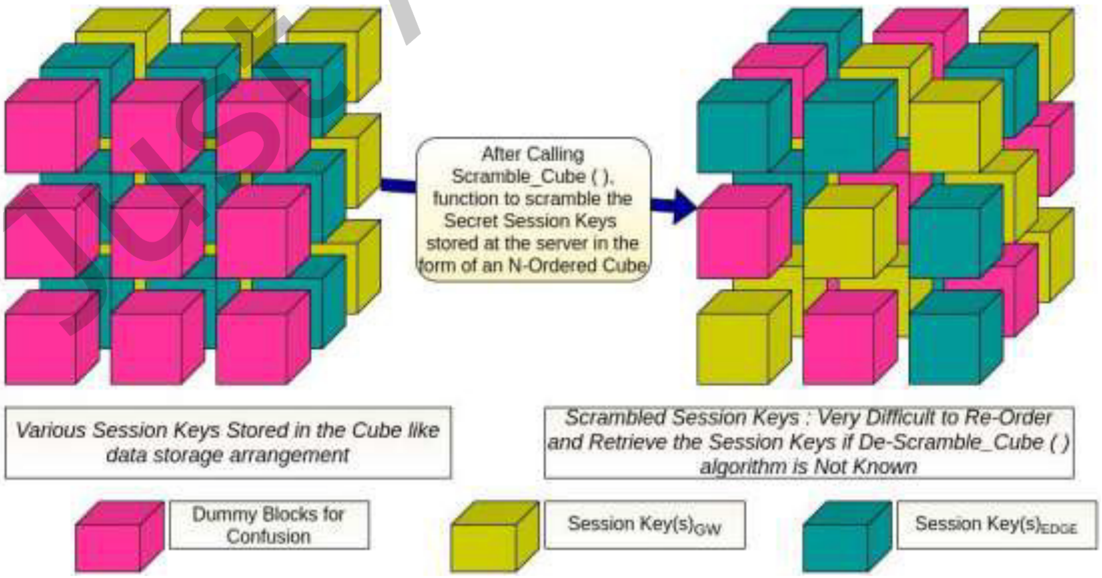

After Calling Scramble_Cube ( ), function to scramble the Secret Session Keys stored at the server in the form of an N-Ordered Cube
Various Session Keys Stored in the Cube like data storage arrangement
Scrambled Session Keys : Very Difficult to Re-Order and Retrieve the Session Keys if De-Scramble_Cube ( ) algorithm is Not Known
Dummy Blocks for Confusion
Session Key(s)GW
Session Key(s)EDGE

Figure 8. Session key Randomizer, RCR$_{KEY}$ for secure session key storage and data protection

```
Algorithm: RCC-Block-MASSK Model (Phase 4): Session Key Randomizer ($RCR_{KEY}$ )
Output of $RCR_{KEY}$ : Efficient Recycling of pre-generated  session keys and storage management
Step 1.  :  Arbitrarily select a small prime number, $sP_{+VE}$ , such that $sP_{+VE}$ <m where "m" is the number of session keys
            generated and m<N where the N is the order of the key randomizer.
Step 2.  :  Index each block in the storage cube at the server, $S_{SERVER}$ with a constant integer starting from "1".
Step 3.  :  Create two cubes, $C_1$ and $C_2$  such that $C_1$ is used to scramble the order of storing session keys, $SSK_{KEY1}$, $SSK_{KEY2}$,
            $SSK_{KEY3}$ ........$SSK_{KEY\ N}$. $C_2$ is used to finally store the converted bytes of the session keys and scramble them in the
            storage making it almost impossible for the attacker to retrieve the data without the de-scrambling algorithm stored
            at the $CT_{Admin}$.
Step 4.  :  For every $SSK_{KEY}$ , from i=1 to m,
                {   C1-Block [i]= $SSK_{KEY}$[i];
                    Add a Dummy Block, C1-Dummy[After every i=i+7 session key, $SSK_{KEY}$]
                    Call Scramble_Cube( )},      function after every entry of a session key, which increases its
                                                 unpredictability.} // End of For
Step 5.  :  Transfer the contents of cube $C_1$ to cube $C_2$, using matrix copy mechanism.
Step 6.  :  Scramble the cube finally at the server station, $S_{SERVER}$ one more time to produce the scrambled secure session key
            cube storage in $C_2$.
Step 7.  :  Delete Cube $C_1$, and send the acknowledgement, Ack to the device, that its session key is stored successfully at the
            server.                                        // End of $RCR_{KEY}$ algorithm
```

Figure 8. describes the session key randomizer, RCR_{KEY} for secure session key storage and data protection. Section 6. ahead discusses the results obtained from the proposed model implementation and simulations from a security point of view.

6 HARDWARE AND SOFTWARE REQUIREMENTS, RESULTS AND EXPERIMENTAL ANALYSIS

This section 6 discusses the basic system requirements for various stages of Blockchain-IoT environment simulations and security analysis.

6.1 Hardware and Software Requirements

The experiment simulations and security verifications were done in two mode:

1st Mode, M1: With few nodes and network miners working in sync with the edge IoT devices, drawing minimum battery and computation overhead in static environment simulation only with the session key agreement mechanism (Without RCR_{KEY} and RCC cryptosystem).

2nd Mode, M2: With maximum load possible in terms of number of blocks and IoT edge devices along with the RCC cryptosystem, session key agreement mechanism and RCR_{KEY} server storage simulation. M2-Mode draws maximum load from the blockchain-IoT setup to study the behaviour of the proposed model at peak conditions.

Hardware and Softwares required: A bare minimum workstation with Ryzen 9 processor with Turbo-Overclocking available up to *4.7 to 5 Ghz*, *64GB RAM(3200MHz)* with sufficient storage required for S_{SERVER} , SSD of minimum 2-TeraBytes with SATA support form block transmission to external workstation for extended intensive simulations. OS- Windows 10 or higher. GPU of minimum 8GB, Nvidia *RTX 3050Ti* or Higher, Preferably, Nvidia Jetson Xaviour with Nvidia *RTX 3080 (8GB)* in Linux Operating System environment. Contiki-Cooja Simulator and *NS-3* simulator for IoT environment integration with the workstation blockchain mechanism. Two mobile phones for Dynamic IoT Node real-time simulation on small scale for practical results and scaling purposes. Same mobile devices were used in 2-factor authentication during the session key agreement mechanism for device authentication. 4 *Raspberry Pie*(latest) with Raspbian (updated), 4 Arduino *UNO*, connecting cables and a wireless internet interface to provide wireless connectivity of internet. Each networking interface on the aforementioned devices having *WiFI6* is preferable. Formal cryptographic security was analyzed using Proverif protocol verifier (section 6.3). *NS-3* tool for cryptographic security verification of the proposed Rubik's cube based crypto-system and MIRACLE tool for time complexity analysis. We utilized the "RIME" mechanism for multicasting the "*zvma*"-token based block/service transaction on the Blockchain-IoT network. The variables to calculate the energy quantum consumption in (Joules) were, A_{ENERGY}: Amount of energy, I_A:Current, $E_{VOLTAGE}$: Electrostatic Potential difference/Voltage and Bit-PT(Bit processing time) and E-time(execution time) as shown in Eq(20).

Energy Consumed in quantums, $EC_Q = [(A_{ENERGY})*(I_A)*(E_{VOLTAGE})]/(Bit\text{-}PT)*(E\text{-}time)$ .... *Eq(20)*

6.2 Results and Experimental Analysis

The motes used in simulation were dynamic edge device/drone replicating motes and sensor type motes that may sense heat, moisture, motion with a certain number of *RFID* tag motes(virtual) to provide a real heterogeneous IoT like environment. Each node in the B-IoT network was assigned a bare minimum system configuration of *2.5Kbps* wireless transmission adapter (standard IEEE) with *16MHz* microcontroller (*20K* volatile memory and *56k* storage for edge data streaming to the gateway nodes). Wireless transmission distance for the simulation for dynamic and static motes in the process of deployment was set to *200* meters range. Bit-by-Bit energy consumption was estimated based on energy quantum EC_Q. Graphs plotted and comparative analysis was done considering the average of the numerous values obtained from a sufficient number of parameter variation, keeping in account the standard deviation and time-consumption during *POLWSCT* consensus algorithm. These average values take into account the standard deviation factor and minimize the range of optimal energy consumption for a given number of nodes in the *RCC-Block_MASSK* model. The experimental results and analysis of various algorithms proposed in our *RCC-Block-MASSK* model to estimate the comparative estimation among other present mechanisms. We evaluate the model in terms of:

- Mining time consumption for *POLWSCT* consensus algorithm,
- Energy consumption per number of mined transactions(energy depends on "*zvma*" token transactions being carried on for service exchange),
- Effect of increased number of blocks on the Blockchain driven IoT network,
- Effect of increased number of Transactions per block on the Blockchain-IoT network
- All the Effects of variations in parameters are analyzed on three major -scenarios:
  (i) Blockchain-IoT network model with the Rubik's cube cryptosystem,
  (ii) Blockchain-IoT network without the Rubik's cube cryptosystem, and
  (iii) Blockchain-IoT network with Rubik's cube cryptosystem and Session Key randomizer, RCR_{KEY}. Figure.9 describes the alterations observed in the consensus time of various statistical consensus algorithms, such as *PoS, PoW, PoA* etc against the proposed model. Consensus time difference analysis and their respective latencies depict that Proof-of-Work and Proof-of-Activity based Blockchain-IoT environments have higher values for time consumptions in processing a User's request because the time taken to process a request increases significantly as the mining rate and number of miners in a blockchain environment increases. Proof-of-Work and Proof-of-Activity are dense consensus algorithms that put a heavy toll on the miners. In the IoT network, this heavy toll is handled by the resource limited devices which results in the sudden increase in the number of frequent node replacements, replacement of the battery drained nodes with the new ones in order to keep the services running. These consensus algorithms may provide high security in some perspective but this security comes at the price of heavy computational load on the devices handling the consensus and mining tasks. Proof-of-Stake based blockchain consensus algorithm is a two part procedure in which life of a coin(token) and its blockchain stature (reputation) are taken into account for measuring the priority of the request whereas the second part involves ledger block verification and data renewal along with the ledger updation. Proof-of-Authentication works on the basis of virtual trust developed among 3-way communication among devices in order to authenticate each other.

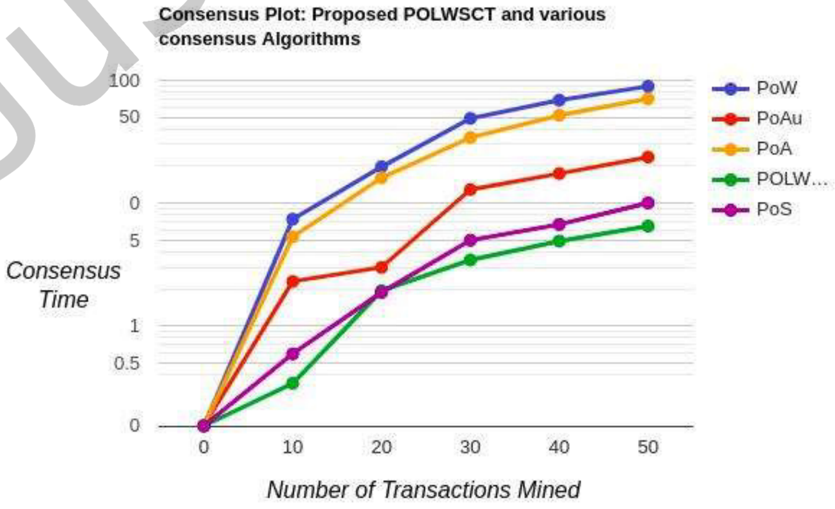


Figure 9. Comparative analysis: Performance of Proposed POLWSCTConsensus algorithm stacked against the other available probabilistic consensus algorithms.

Figure 9. Depicts the performance of *POLWSCT* consensus vs the other probabilistic mechanisms for establishing consensus in the blockchain network. Clearly, the average values in the plot illustrate that the proposed consensus takes lesser time quantums to service the consensus establishment process as compared to the other ones. Amount of time taken by the consensus mechanism used in our *RCC-Block-MASSK* model for blockchain driven IoT network for heterogeneous devices takes less time without compromising security and also while delivering the high end security from session key agreement perspective as well. Similarly, Figure 10. highlights the energy quantum consumption of various consensus based blockchain environments integrated with the IoT network, based on "EC_Q" parameter defined earlier. The amount of energy consumed for mining a certain number of transaction(s) rises as the number of transactions and block validation/updation requests increases. We started experimenting using 1 block for which results were almost the same, as we gradually increased the number of transactions being mined by the GW_{NODE_N} (gateway nodes) or "*zvma*"-token miners, the energy demands per miner increased significantly. Number of mined transactions for this case were tested up to *250* mined transactions per IoT edge cluster. As per simulations and real-time experiments, the time complexity and energy consumption results favors the proposed goal of the *RCC-Block-Model* to provide a light-weight solution to IoT security and authentication problem, providing with higher security and lesser energy dissipation during runtime. We tested and experimented the model in three **scenarios**, all with three major possible runtime **cases**:

Case 1: *POLWSCT consensus + Secret session key generation and authentication mechanism + Rubik's cube cryptosystem (without secret session key recycling,* RCR_{KEY}*)*

Case 2: *POLWSCT consensus + Secret session key generation and authentication mechanism + (with Rubik's cube cryptosystem and secret session key recycling,* RCR_{KEY}*)*

Case 3: *POLWSCT consensus + Secret session key generation and authentication mechanism (without Rubik's cube cryptosystem and session key recycling,* RCR_{KEY}*)*

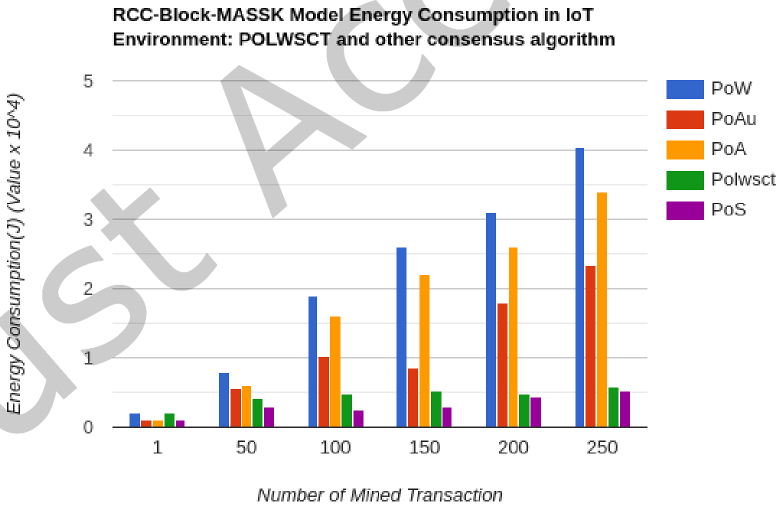


Figure 10. Illustrates that the proposed RCC-Block-MASSK model(marked green) operates at optimal amounts of energy quantums, making it more favourable for resource restricted IoT_{EDGE}.Dyn-IoT_{EDGE} nodes.

Scenario 1: This focuses on the number of blocks mined up to 3000 and measures the time taken for block processing as up to 200000(in ms). The simulation output average values plotted in Figure 11 describes that if the number of blocks being mined per chain assigned over an IoT_{EDGE} device cluster handled by GW_{NODE_N} is going to increase almost linearly with the total computation time. It means as the demand for processing more and more token mining or block processing requests increases, it is going to take a considerable amount of time to process such a high number of requests. For example, in Figure 11, at the maximum number of transactions mined, i.e, 3000, the proposed model assigned chain to one cluster "Cluster-1" of one simulation having (SSK_{KEY}+RCC+RCR_{KEY})[Marked with green plot line] performs significantly better than the "Cluster-2" of a chain running the same model without the session key recycler mechanism. The recycling of session keys plays a vital role in saving the energy quantums of

IoT_{EDGE} devices. However, the amount of time taken by the BIoT chain to process transaction requests is greater than a model with just the session key agreement phase.

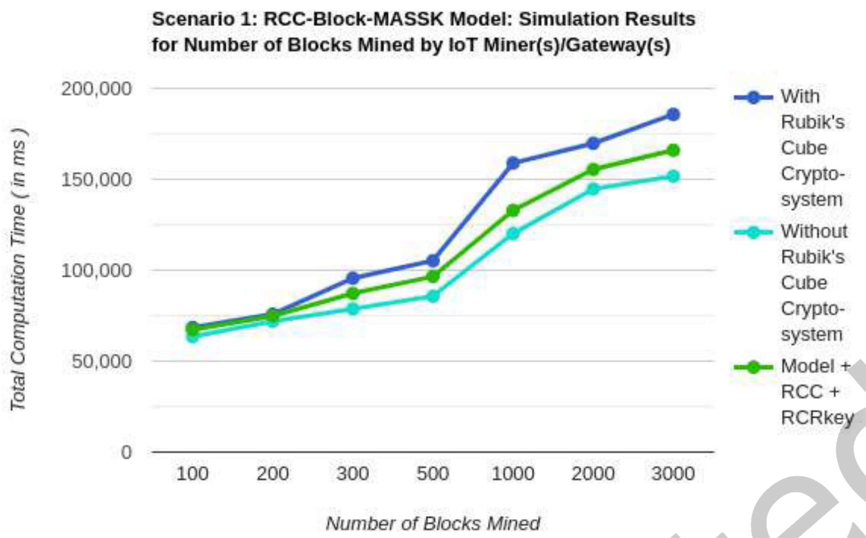


Figure 11. Scenario 1: Model performance after increasing number of blocks mined per chain for Case-1, Case-2 and Case-3.

Scenario 2 : This simulation result is obtained by considering the number of ledger based service exchange transactions among IoT network mining peers(GW_{NODE_N}). Figure 12 shows that total computation time for Case 1, Case 2 and Case 3 are almost identical with slight changes observed when the number of transactions per block reaches up to 180. At “160” it is observed that the complete RCC-Block-MASSK model in case 2 consumes a lesser amount of time as compared to the case 1, whereas takes almost equal amount of time as taken by case 3.

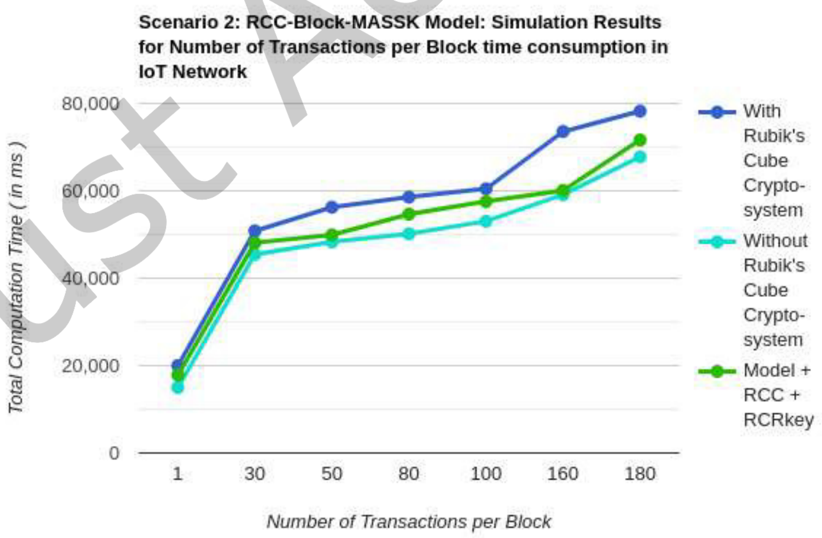


Figure 12. RCC-Block-MASSK model Simulation results for Case 1, Case 2 and Case 3 based on Number of transaction per Block time consumption in a Blockchain-IoT network

Scenario 3: This simulation result is obtained by considering the scalability of the proposed model in terms of number of nodes. As the number of nodes in a Blockchain driven IoT network increases, it is observed that All the three cases of the BIoT network, Case 1, Case 2 and Case 3 show similar trend and the time complexity of the cases are, *Time Complexity:* *Case 1> Case 2> Case 3,*

whereas, the IoT security trend from the network attacks such as, MITM, Packet Snooping, Identity theft , Sensor capture attack etc is found to be,

Attack-Security: *Case 2> Case 1> Case 3,*

For the three cases Figure 13 shows the total computation time for *Case 1, Case 2* and *Case 3*, which is almost similar whereas security for case 2 is the highest.

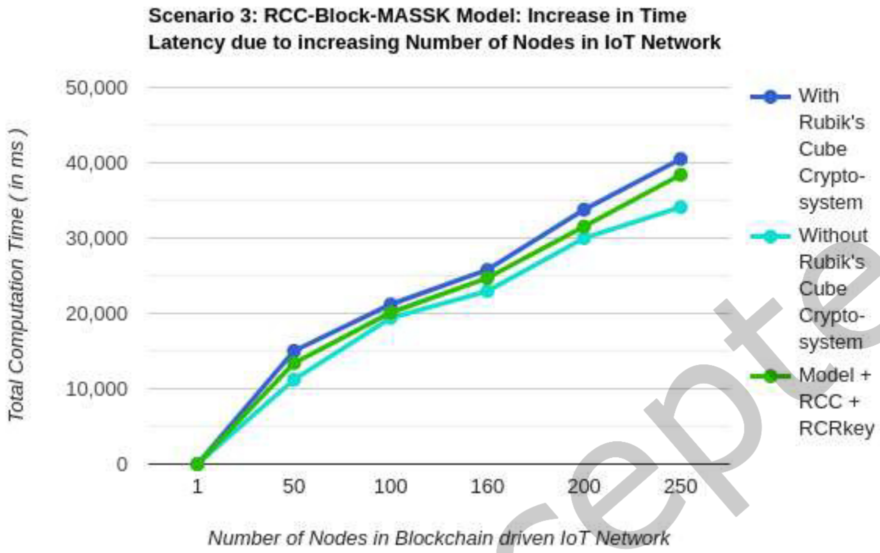


Figure 13. RCC-Block-MASSK Model performance in terms of time taken to process requests with the increase in number of nodes in the BIoT network for Case1, Case 2, and Case 3

Time taken for each individual operation in the algorithm of the model is estimated using *NS-3* and *MIRACLE* tools. Table 2 Depicts the maximum and minimum time taken for a single atomic operation for the respective category in BIoT environment simulation. Multiple iterations of experiments were exhibited to evaluate the final Average time consumption by the individual parameter in the *RCC-Block-MASSK* model.

Table 2. NS-3 and Miracle tool based Time Complexity Costs estimated for RCC-Block-MASSK Model and their Average Values

Time Complexity Parameter	Maximum consumed Time Quantum in (EC_Q) [in milliseconds]	Minimum consumed Time Quantum in (EC_Q) [in milliseconds]	Average Time (in ms)
C_F	1.4988	0.8870	1.1929
$T_{ioT\text{-}EDGE}$	4.4163	0.9454	2.6808
$TS_{EDyn\text{-}DGE}$ (IoT-EDGE)	0.0627	0.0579	0.0603
TS_{GW} (Gateway)	0.0463	0.0442	0.0453
T_{h0} (Hashing)	0.2130	0.0313	0.1221
$T_{Block\text{-}Update}$ (One Block in Blockchain)	240.00	80.000	160.00
T_{ADD} (Addition)	0.0015	0.0010	0.0012
$T_{MULTIPLY}$	0.0069	0.0010	0.0039
T_{XOR} (Digital Signature Verification)	1.7412	0.8455	1.2933

Clearly, the most time consuming operation is where a block is being added to the blockchain taking on an average of 160 ms per block. Addition operation is atomic and it proves to be the least time consuming, taking 0.0012 milli-seconds followed by $T_{MULTIPLY}$ with 0.0039 ms. It is to be noted that $T_{ioT\text{-}EDGE}$ is the most time consuming operation during authentication and session key generation phase with 2.6808 ms. It is so because it takes time to mutually authenticate the gateway(s) and the intermediate nodes before the actual edge device can receive the commands from CT_{ADMIN} / S_{SERVER}. *RCC-Block-MASSK* model has efficient digital signature generation and verification mechanisms as it takes about 1.2933ms to verify a digital signature. Tools used for experimentation and calculation of both simulation and real time values of time were carried out using, *NS-3* (for simulation), *MIRACLE* and *Raspberry Pi 4* setup with the hardware specified in section 6.1. We proposed a *RCC*-Rubik's cube based hybrid cryptosystem that utilizes the probabilistic complexity and randomness of a rubik's cube puzzle alongwith the image encryption/decryption techniques. To analyze the *3-D* to *2-D* conversion and pixel diffusion success rate based on variety of parameters such as histogram analysis, chi-square tests, information entropy, *UACI, NPCR* etc, we experimented the *RCC*-cryptosystem on standard as well as local images for encryption and decryption analysis. Figure 14 shows the histogram analysis of plain image and cipher image using both standard and local experimental images. *RCC* cryptosystem of the *RCC-Block-MASSK* model behaves differently even when there are slight changes in the image. Figure 14 (a) shows the histogram of plain "*Monkey*" standard image, (b) depicts the histogram for Monkey image flipped vertically followed by rotation of each pixel by 90° degrees. Figure 14 (c) shows the histogram after the scrambling process of the data and then its encryption using RCC-cryptosystem. It shows a lot of uniformity is obtained after encryption proving its pixel diffusion reliability and security.

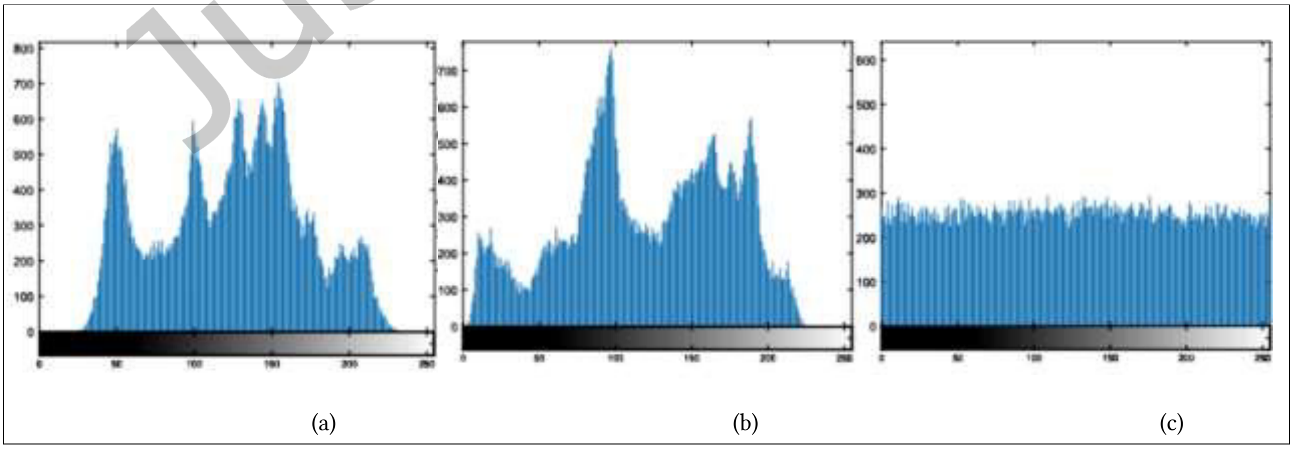
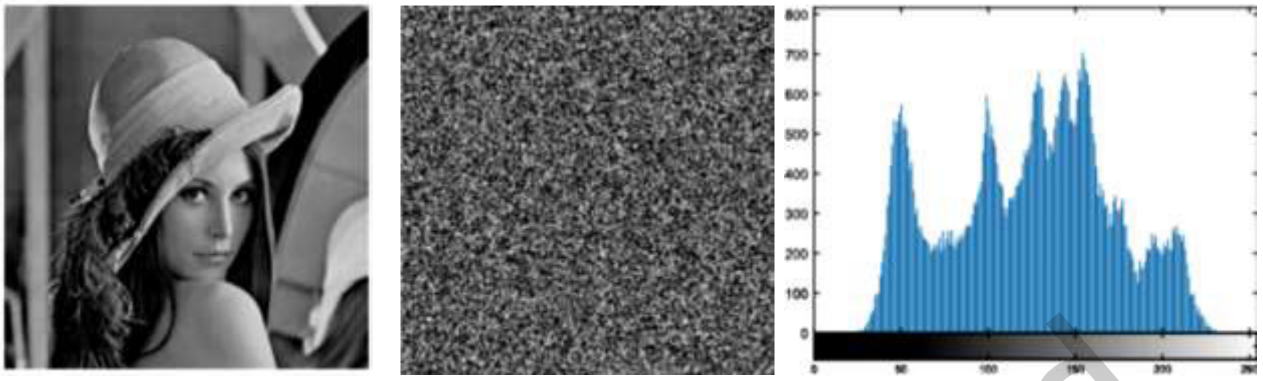


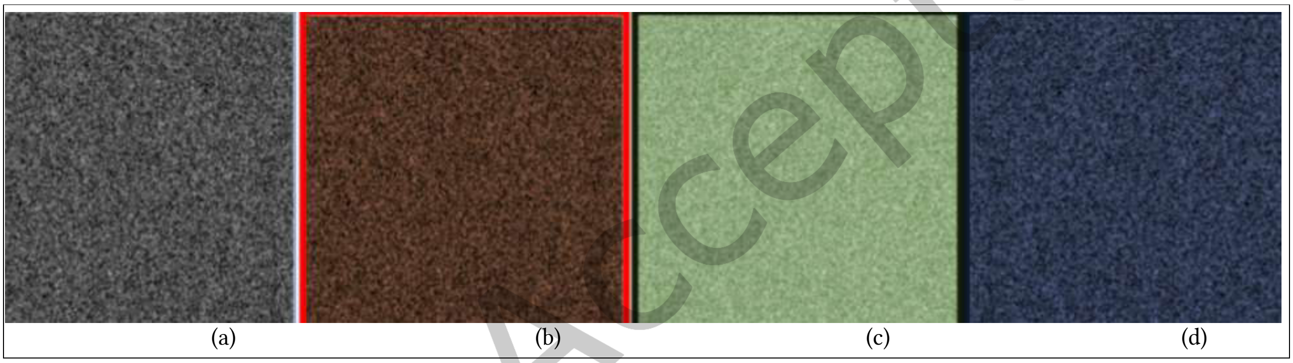
Figure 14. The Histogram analysis of plain image and cipher images: (a) Histogram of plain "Monkey" image, (b) Histogram for Monkey image flipped vertically followed by rotation of each pixel by 90° degrees. Obtained (c) after the scrambling process of the data and then its encryption using RCC-cryptosystem

Figure 15(a) is a plain Lena image, (b) is the histogram of image obtained after pixel scrambling using N-Order Rubik's cube. Finally, Figure 15 (c) shows the uniform and stable histogram for the same image obtained after a successful image encryption process.



(a) (b) (c)

Figure 15. The Histogram analysis of plain image and cipher images:(a) plain Lena image, (b) Histogram obtained after pixel scrambling using N-Order Rubik's cube. Figure 14 (c) Histogram for Lena obtained after image encryption process.



(a) (b) (c) (d)

Figure 16. Cipher Results for (a) RCC-Block-MASSK Cipher, (b) Red (c) Green (d) Blue

Figure 16. depicts the ciphers generated for standard image "*Lena*" as per our proposed cipher-image generation mechanism. Figure 16 (a) shows simple *RCC-Block-MASSK* cipher image, (b) red channel cipher image, (c) green channel cipher image and (d) shows blue channel cipher images generated by the proposed encryption mechanism. The evaluation of RCC cryptosystem based on chi-square test with 5% significance, information entropy.UACI and NPCR tests is given in Table 3. The percentage performance of the proposed cryptosystem evaluates to be satisfactory in terms of being light-weight as well as supportive for faster communications among IoT network components while providing the desired security. RCC crypto-system is strong enough to avoid MITM, generic passive, Brute force attacks, key-copy attacks, Botnet attack, *Node/IoT$_{EDGE}$* impersonation and gateway impersonation attacks.

Table 3. RCC-Block-MASSK Model: Rubik's Cube Crypto-system security analysis on the basis of Chi-square test, Entropy, UACI and NPCR performance evaluation

File-Name	P-value (Chi-Square test)(5% of Significance Testing)	Information Entropy	UACI Results (%)	NPCR Results (%)
Lena	0.7022	8.2033	29.8714	97.2489
Monkey	0.1549	8.1115	31.6548	96.9943
Local image 1	0.9976	8.6590	31.5712	97.2054
Local Image 2	0.6854	8.7713	30.9868	97.8013
Local Image 3	0.6354	7.9845	31.2275	95.0032
Local Image 4	0.7885	8.4136	33.8974	96.8869
Local Image 5	0.8096	8.1862	29.3296	97.0988

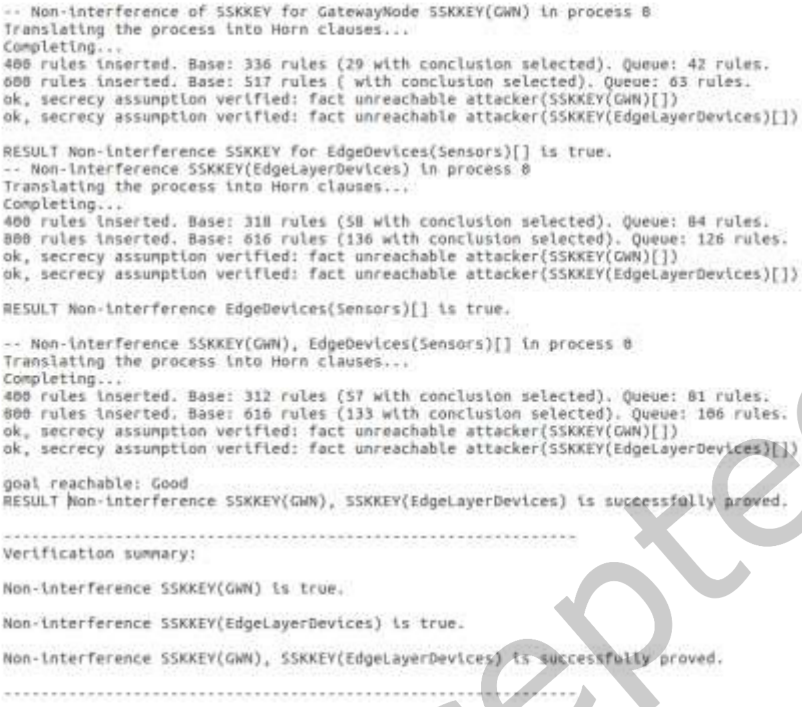
The security of the *RCC-Block-MASSK* model resides in the degree of parameter computation randomness, which is by default in-built with the use of cube mechanism, and the incorporation of significant sized prime numbers discussed in section 3. The amount of time taken to compute cryptographic parameters plays a key role in determining the toll taken by the IoT resource constrained devices on executing the security mechanism. Table 3 provides the experimental data obtained after numerous iterations of experiments to analyze the proposed *RCC-Block-MASSK* modelon a real timeIoT network integrated with Blockchain environment for both static and dynamic edge devices. Table 4 depicts the minimum(T_{MIN}) , maximum(T_{MAX}) and average ($T_{AVERAGE}$) computation time for the respective cryptographic security parameter for static, dynamic modes and server end time delay computation due to RCR_{KEY} randomizer.

Table 4. RCC-Block-MASSK Model: Cryptographic parameters and Intermediate parameters computation time in terms of minimum, maximum and average values for Static, Dynamic devices and Server. (in milliseconds) [Real-Time System setup with 12 Raspberry Pi 4 and POLWSCT consensus Blockchain]

Type of Device	Cryptographic Parameter	Minimum Time in milliseconds (T_{MIN})	Maximum Time in milliseconds (T_{MAX})	Average Time in milliseconds ($T_{AVERAGE}$)
Dynamic	$A_{EDGE(i)}$	2.023	3.047	2.535
	$Sig_{EDGE(i)}$	2.289	3.561	2.925
	$Dyn\text{-}P_{KEY}$	2.055	4.098	3.077
	$Pub\text{-}Dyn\text{-}IoT_{EDGE}$	2.087	5.164	3.633
	SSK_{KEY}	5.335	6.786	6.065
	A_{GW}	1.360	2.367	1.868
	$new\text{-}TID_{EDGE}$	2.459	5.001	3.734
Static	$A_{EDGE(i)}$	1.008	2.691	1.849
	$Sig_{EDGE(i)}$	1.444	3.713	2.579
	SP_{KEY}	2.028	3.554	2.791
	$Pub\text{-}IoT_{EDGE}$	1.002	4.997	3.000
	SSK_{KEY}	4.368	5.225	4.777
	A_{GW}	0.981	2.159	1.540
	$new\text{-}TID_{EDGE}$	2.153	4.872	3.519
S_{SERVER}	C_2 *(scrambled secure session key cube)*	6.756	8.301	7.578
	RCR_{KEY}*(Deleting* C_1 *Cube)*	0.051	1.297	0.6759

6.3 Formal Cryptosystem/Protocol Security verfiication: Proverif Protocol Verifier

In order to verify the security strength and parameter reachability of session keys being generated in the proposed model, we used the proverifPostHeadPara protocol verifier. It is used to prove the rachability of the SSK_{KEY}(Gateway Nodes) and SSK_{KEY}(for edge layered devices -both static and dynamic nodes). The results are promising and it indicates and verifies the secrecy assumption for the desired hypothesis to be "*True*".



```
-- Non-interference of SSKKEY for GatewayNode SSKKEY(GWN) in process 0
Translating the process into Horn clauses...
Completing...
400 rules inserted. Base: 336 rules (29 with conclusion selected). Queue: 42 rules.
600 rules inserted. Base: 517 rules ( with conclusion selected). Queue: 63 rules.
ok, secrecy assumption verified: fact unreachable attacker(SSKKEY(GWN)[])
ok, secrecy assumption verified: fact unreachable attacker(SSKKEY(EdgeLayerDevices)[])

RESULT Non-interference SSKKEY for EdgeDevices(Sensors)[] is true.
-- Non-interference SSKKEY(EdgeLayerDevices) in process 0
Translating the process into Horn clauses...
Completing...
400 rules inserted. Base: 318 rules (58 with conclusion selected). Queue: 84 rules.
800 rules inserted. Base: 616 rules (136 with conclusion selected). Queue: 126 rules.
ok, secrecy assumption verified: fact unreachable attacker(SSKKEY(GWN)[])
ok, secrecy assumption verified: fact unreachable attacker(SSKKEY(EdgeLayerDevices)[])

RESULT Non-interference EdgeDevices(Sensors)[] is true.

-- Non-interference SSKKEY(GWN), EdgeDevices(Sensors)[] in process 0
Translating the process into Horn clauses...
Completing...
400 rules inserted. Base: 312 rules (57 with conclusion selected). Queue: 81 rules.
800 rules inserted. Base: 616 rules (133 with conclusion selected). Queue: 106 rules.
ok, secrecy assumption verified: fact unreachable attacker(SSKKEY(GWN)[])
ok, secrecy assumption verified: fact unreachable attacker(SSKKEY(EdgeLayerDevices)[])

goal reachable: Good
RESULT Non-interference SSKKEY(GWN), SSKKEY(EdgeLayerDevices) is successfully proved.

-------------------------------------------------------------
Verification summary:

Non-interference SSKKEY(GWN) is true.

Non-interference SSKKEY(EdgeLayerDevices) is true.

Non-interference SSKKEY(GWN), SSKKEY(EdgeLayerDevices) is successfully proved.

-------------------------------------------------------------
```

Figure 17. Formal security verification results using Proverif protocol verifier: Security breach attacks were successfully prevented and the reachability of the SSK_{KEY} (Shared Secret Session Key) was proved to be Strong and it withstood all the attacks that were targetted at compromising SSK_{KEY}

This means the secure session key(s) under the current management mechanism are secure and the secrecy of the parameters being shared among IoT network devices are unreachable to the attacker. The unsecure parameters being passed on from one node to another maybe visible to the attacker but the end result, which is our SSK_{KEY} (secure shared session key) being generated is found to be completely unreachable and highly secure. The attacker was not able to penetrate the security boundaries and during experimental analysis, no security breaches were found during various attack scenarios (Brute force, Man-in-the-middle, sniffing, phishing and DDOS attacks). Following Figure 17. shows the formal verification results obtained by using the proverif protocol verifier. Security breach attacks were successfully prevented and the reachability of the SSK_{KEY} (Shared Secret Session Key) was proven to be strong and reliable. It withstood all the attacks that were targetted at compromising SSK_{KEY}.

7 CONCLUSION AND FUTURE POSSIBILITIES

Blockchain is the technology which is trending as well revolutionizing the present security solutions while maintaining records in the ledger. It provides security from data record tampering by the attacker and provides various means to authenticate the networking devices. Seeking the growing demands and dependencies of human life on the IoT enabled devices, we proposed a light-weight blockchain consensus mechanism that drives the trust among IoT networks devices and provides higher level of security solutions. The model is integrated with the newly proposed Rubik's cube based crypto-system that provides a tougher encryption/decryption mechanism for the attacker to breach. IoT edge devices are low levelled battery constrained devices that are confined to limited battery support. These devices can not run high-end security protocols as they consume high amounts of battery life, making the device run out of power before these device(s) even contribute any significant data to the IoT network. To cope with multiple IoT attacks possible due to lack of security options and numerous vulnerabilities, we provided a secure session key generation mechanism that not only helps generate an intermediate trust among devices, but also allows to create a secure session for data transmission using hashing techniques. In section 2 we discussed in detail the two prime components of our proposed *RCC-Block-MASSK* model, the first one is the POLWSCT consensus mechanism, which is a light-weight approach for establishing consensus. Second one is the transaction updation methodoly where the algorithm keeps the track of service exchange using the “zvma” service token. This approach of Blockchain integration with the IoT network helps to quickly authenticate the communicating peers in the network while maintaining a healthy record of service exchange of transactions. In section 3 we introduced a newly designed bit scrambling Rubik’s cube mechanism that helps randomizing the binary converted data into a randomly achieved scrambled form. Due to this scrambling, the attacker becomes incapable of guessing the order of shuffling/scrambling in order to decrypt the cipher. Our cryptosystem performs at the same level of performance speed, as of any s-box permutator, provided, the data is first converted into binary format. Section 2 and 3 constituted the Phase 1 and Phase 2 of the proposed *RCC-Block-MASSK* model. In section 4 we discussed the phase 3 of our model that described the proper mutual authentication mechanism. This mechanism provides the parametric and technical details of how the common secure session key SSK_{KEY} is generated after successful authentication among communicating network nodes/edge devices/gatewaynodes/blockchain. Our model provides flexibility in applying multiple mechanisms within the heterogeneous variety of the devices in the network, based on their resource count and computation capabilities.We compared the *POLWSCT* mechanism with the other consensus algorithm giving a briefing of how well our model performs against the pre-existing models. Digital signatures utilized in our model diversify the applicability of this model over a wide range of research areas, such as Biometric authentication, Individuality recognition, Smart-device interaction, Healthcare, Smart-cities, smart-vehicles etc. We observed that our model could have been even better if the hashing algorithms were more efficient. There is scope of improvement in the hashing functions and their utilization in cryptographic parameter passing among peers. We can further try to enhance the performance of this model by finding a way to opt for tuple hash or parallel hash functions. Our model is flexible and the cryptosystem(s) used in our model can be replaced with an even lighter crypto-graphic technique in the forthcoming future (current version is the best version of the proposed model). Our model provides flexibility to incorporate futuristic crypto-systems and hash based crypto-verifiers in order to authenticate devices in the BIoT network before actual data transmission. We conclude this paper hoping for better emerging technologies to come in the near future so that we can improve this model even further.

REFERENCES

[1] Lan, L., Shi, R., Wang, B., & Zhang, L. (2019). An iot unified access platform for heterogeneity sensing devices based on edge computing.IEEE access,7, 44199-44211.

[2] Vogt, H. (2002, August). Efficient object identification with passive RFID tags. In the International Conference on Pervasive Computing(pp. 98-113). Springer, Berlin, Heidelberg.

[3] Dwivedi, A. D., Srivastava, G., Dhar, S., & Singh, R. (2019). A decentralized privacy-preserving healthcare blockchain for IoT.Sensors,19(2), 326.

[4] Dorri, A., Kanhere, S. S., Jurdak, R., & Gauravaram, P. (2017, March). Blockchain for IoT security and privacy: The case study of a smart home. In 2017 IEEE international conference on pervasive computing and communications workshops (PerCom workshops)(pp. 618-623). IEEE.

[5] Celik, Z. B., Babun, L., Sikder, A. K., Aksu, H., Tan, G., McDaniel, P., & Uluagac, A. S. (2018). Sensitive information tracking in commodity IoT. In 27th {USENIX} Security Symposium ({USENIX} Security 18) (pp. 1687-1704).

[6] Alam, M. E., Kader, M. A., Parvin, R., Sultana, S., Sultana, Z., & Muhammad, S. D. (2021, January). IoT based biometric seat reservation and transport management system for university bus. In 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)(pp. 649-653). IEEE.

[7] El-Haii, M., Chamoun, M., Fadlallah, A., & Serhrouchni, A. (2018, October). Analysis of cryptographic algorithms on iot hardware platforms. In 2018 2nd Cyber Security in Networking Conference (CSNet)(pp. 1-5). IEEE.

[8] Soe, Y. N., Feng, Y., Santosa, P. I., Hartanto, R., & Sakurai, K. (2020). Towards a lightweight detection system for cyber attacks in the IoT environment using corresponding features.Electronics, 9(1), 144.

[9] Melki, R., Noura, H. N., & Chehab, A. (2020). Lightweight multi-factor mutual authentication protocol for IoT devices.International Journal of Information Security,19(6), 679-694.

[10] Yan, Z., Zhang, P., & Vasilakos, A. V. (2014). A survey on trust management for the Internet of Things. Journal of network and computer applications, 42, 120-134.

[11] Collina, M., Corazza, G. E., & Vanelli-Coralli, A. (2012, September). Introducing the QEST broker: Scaling the IoT by bridging MQTT and REST. In 2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications-(PIMRC)(pp. 36-41). IEEE.

[12] Mahendran, R. K., Prabhu, V., Parthasarathy, V., Thirunavukkarasu, U., & Jagadeesan, S. (2021). An energy-efficient centralized dynamic time scheduling for internet of healthcare things.Measurement, 186, 110230.

[13] Tournier, J., Lesueur, F., Le Mouël, F., Guyon, L., & Ben-Hassine, H. (2021). A survey of IoT protocols and their security issues through the lens of a generic IoT stack. Internet of Things, 16, 100264.

[14] Ferrag, M. A., Shu, L., Djallel, H., & Choo, K. K. R. (2021). Deep Learning-Based Intrusion Detection for Distributed Denial of Service Attack in Agriculture 4.0. Electronics, 10(11), 1257.

[15] Agiollo, A., Conti, M., Kaliyar, P., Lin, T., & Pajola, L. (2021). DETONAR: Detection of routing attacks in RPL-based IoT. IEEE Transactions on Network and Service Management.

[16] Liu, X., Zeng, Q., Du, X., Valluru, S. L., Fu, C., Fu, X., & Luo, B. (2021, October). SniffMislead: Non-Intrusive Privacy Protection against Wireless Packet Sniffers in Smart Homes. In 24th International Symposium on Research in Attacks, Intrusions and Defenses(pp. 33-47).

[17] Kim, W. B., & Lee, I. Y. (2021). Survey on Data Deduplication in Cloud Storage Environments. Journal of Information Processing Systems, 17(3), 658-673.

[18] Da Xu, L., Lu, Y., & Li, L. (2021). Embedding blockchain technology into IoT for security: a survey. IEEE Internet of Things Journal.

[19] Xu, C., Qu, Y., Luan, T. H., Eklund, P. W., Xiang, Y., & Gao, L. (2021). A Light-weight and Attack-Proof Bidirectional Blockchain Paradigm for Internet of Things. IEEE Internet of Things Journal.

[20] Alhejazi, M. M., & Mohammad, R. M. A. (2021). Enhancing the blockchain voting process in IoT using a novel blockchain Weighted Majority Consensus Algorithm (WMCA). Information Security Journal: A Global Perspective, 1-19.

[21] Mu, R., Gong, B., Ning, Z., Zhang, J., Cao, Y., Li, Z., ... & Wang, X. (2022). An identity privacy scheme for blockchain-based on edge computing. Concurrency and Computation: Practice and Experience, 34(1), e6545.

[22] Iftikhar, Z., Javed, Y., Zaidi, S. Y. A., Shah, M. A., Iqbal Khan, Z., Mussadiq, S., & Abbasi, K. (2021). Privacy preservation in resource-constrained IoT devices using blockchain—A survey. Electronics, 10(14), 1732.

[23] Mai, T., Yao, H., Zhang, N., Xu, L., Guizani, M., & Guo, S. (2021). Cloud mining pool aided blockchain-enabled internet of things: An evolutionary game approach. IEEE Transactions on Cloud Computing.

[24] Siddiqui, F., Beley, J., Zeadally, S., & Braught, G. (2021). Secure and lightweight communication in heterogeneous IoT environments. Internet of Things, 14, 100093.

[25] Korf, R. E. (1997, July). Finding optimal solutions to Rubik's Cube using pattern databases. In AAAI/IAAI (pp. 700-705).

[26] Singh, S., Hosen, A. S., & Yoon, B. (2021). Blockchain security attacks, challenges, and solutions for the future distributed iot network. IEEE Access, 9, 13938-13959.

[27] Henderson, T. R., Lacage, M., Riley, G. F., Dowell, C., & Kopena, J. (2008). Network simulations with the ns-3 simulator. SIGCOMM demonstration, 14(14), 527.

[28] Österlind, F., Eriksson, J., & Dunkels, A. (2010, November). Cooja TimeLine: a power visualizer for sensor network simulation. In Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems (pp. 385-386).

[29] Zamani, E., He, Y., & Phillips, M. (2020). On the security risks of the blockchain. Journal of Computer Information Systems, 60(6), 495-506.

[30] Stergiou, C. L., Psannis, K. E., & Gupta, B. B. (2020). IoT-based big data secure management in the fog over a 6G wireless network. IEEE Internet of Things Journal, 8(7), 5164-5171.

[31] Stergiou, C. (2021). Efficient and secure algorithms for big data handling, processing, and delivery in cloud computing for Internet of Things networks (Doctoral dissertation, Πανεπιστήμιο Μακεδονίας. Σχολή Επιστημών Πληροφορίας. Τμήμα Εφαρμοσμένης Πληροφορικής).

[32] Zhu, X., & Jiang, C. (2021). Integrated Satellite-Terrestrial Networks Toward 6G: Architectures, Applications, and Challenges. IEEE Internet of Things Journal, 9(1), 437-461.

[33] Al-Joboury, I. M., & Al-Hemiary, E. H. (2020). Virtualized Fog Network with Load Balancing for IoT based Fog-to-Cloud. JOIV: International Journal on Informatics Visualization, 4(3), 123-126.

[34] Das, D., Banerjee, S., & Biswas, U. (2021). A secure vehicle theft detection framework using Blockchain and smart contract. *Peer-to-Peer Networking and Applications*, *14*(2), 672-686.

[35] Qu, Q., Xu, R., Chen, Y., Blasch, E., & Aved, A. (2021). Enable Fair Proof-of-Work (PoW) Consensus for Blockchains in IoT by Miner Twins (MinT). *Future Internet*, *13*(11), 291.

[36] Sun, Y., Yan, B., Yao, Y., & Yu, J. (2021). DT-DPoS: A Delegated Proof of Stake Consensus Algorithm with Dynamic Trust. *Procedia Computer Science*, *187*, 371-376.

[37] Maqbool, A., Sattar, S., Naheed, A., Khalid, S., Rana, T., Afzal, F., & Cancan, M. (2021). A comparative analysis of consensus protocols for dealing power theft issues in Pakistan. *Journal of Information and Optimization Sciences*, *42*(7), 1523-1540.

[38] George, J. T. (2022). Consensus Algorithms for Blockchains. In *Introducing Blockchain Applications* (pp. 149-161). Apress, Berkeley, CA.

[39] Benomar, Z., Campobello, G., Segreto, A., Battaglia, F., Longo, F., Merlino, G., & Puliafito, A. (2021). A Fog-based Architecture for Latency-sensitive Monitoring Applications in the Industrial Internet of Things. *IEEE Internet of Things Journal.*

[40] Unal, D., Al-Ali, A., Catak, F. O., & Hammoudeh, M. (2021). A secure and efficient Internet of Things cloud encryption scheme with forensics investigation compatibility based on identity-based encryption. Future Generation Computer Systems, 125, 433-445.

[41] Wazid, M., Das, A. K., & Park, Y. (2021). Blockchain-enabled secure communication mechanism for IoT-driven personal health records. Transactions on Emerging Telecommunications Technologies, e4421.

[42] Fan, Q., Chen, J., Shojafar, M., Kumari, S., & He, D. (2022). SAKE*: A Symmetric Authenticated Key Exchange Protocol with Perfect Forward Secrecy for Industrial Internet of Things. IEEE Transactions on Industrial Informatics.

[43] Kasyoka, P., Kimwele, M., & Mbandu Angolo, S. (2020). Certificateless pairing-free authentication scheme for wireless body area network in healthcare management system. Journal of medical engineering & technology, 44(1), 12-19.

[44] Tan, X., Zhang, J., Zhang, Y., Qin, Z., Ding, Y., & Wang, X. (2020). A PUF-based and cloud-assisted lightweight authentication for multi-hop body area networks. Tsinghua Science and Technology, 26(1), 36-47.

[45] Li, C., Zhang, J., Yang, X., & Youlong, L. (2021). Lightweight blockchain consensus mechanism and storage optimization for resource-constrained IoT devices. Information Processing & Management, 58(4), 102602.

[46] Bouras, M. A., Lu, Q., Dhelim, S., & Ning, H. (2021). A Lightweight Blockchain-Based IoT Identity Management Approach. Future Internet, 13(2), 24.

[47] Abed, S. E., Jaffal, R., Mohd, B. J., & Al-Shayeji, M. (2021). An analysis and evaluation of lightweight hash functions for blockchain-based IoT devices. Cluster Computing, 24(4), 3065-3084.

[48] Ferdush, J., Begum, M., & Uddin, M. S. (2021). Chaotic Lightweight Cryptosystem for Image Encryption. *Advances in Multimedia*, *2021*.

[49] Rao, V., & Prema, K. V. (2021). A review on lightweight cryptography for Internet-of-Things based applications. *Journal of Ambient Intelligence and Humanized Computing*, *12*(9), 8835-8857.

[50] Feng, Q., He, D., Wang, H., Zhou, L., & Choo, K. K. R. (2019). Lightweight collaborative authentication with key protection for smart electronic health record systems. *IEEE Sensors Journal*, *20*(4), 2181-2196.

[51] Afianti, F., & Suryani, T. (2019). Lightweight and DoS resistant multi user authentication in wireless sensor networks for smart grid environments. *IEEE Access*, *7*, 67107-67122.

[52] Dar, M. A., Khan, U. I., & Bukhari, S. N. (2019). Lightweight Session Key Establishment for Android Platform Using ECC. In *Advances in Computer, Communication and Control* (pp. 347-359). Springer, Singapore.

[53] Tamilarasi, K., & Jawahar, A. (2020). Medical Data Security for Healthcare Applications Using Hybrid Lightweight Encryption and Swarm Optimization Algorithm. *Wireless Personal Communications*, *114*(3).

[54] Shen, C., Zhang, K., & Tang, J. (2021). A COVID-19 Detection Algorithm Using Deep Features and Discrete Social Learning Particle Swarm Optimization for Edge Computing Devices. *ACM Transactions on Internet Technology (TOIT)*, *22*(3), 1-17.

[55] Yang, Z., Jin, Y., & Hao, K. (2018). A bio-inspired self-learning coevolutionary dynamic multiobjective optimization algorithm for internet of things services. *IEEE transactions on evolutionary computation*, *23*(4), 675-688.

[56] Bouteghrine, B., Tanougast, C., & Sadoudi, S. (2021). Novel image encryption algorithm based on new 3-d chaos map. *Multimedia Tools and Applications*, 1-23.

[57] Shrivastava, M., Roy, S., Kumar, K., Pandey, C. V., & Grover, J. (2021). LICCA: a lightweight image cipher using 3-D cellular automata. *Nonlinear Dynamics*, *106*(3), 2679-2702.

[58] Khaitan, S., Sagar, S., & Agarwal, R. (2021). Chaos based image encryption using 3-Dimension logistic map. *Materials Today: Proceedings*.

[59] Fataf, N. A. A., Rahim, M. A., He, S., & Banerjee, S. (2021). A Communication Scheme based on Fractional Order Chaotic Laser for the Internet of Things. *Internet of Things*, 100425.

[60] Zhang, J., Guo, M., Li, B., & Lu, R. (2021). A transport monitoring system for cultural relics protection based on blockchain and internet of things. *Journal of Cultural Heritage*, *50*, 106-114.

[61] Ge, C., Susilo, W., Liu, Z., Xia, J., Szalachowski, P., & Fang, L. (2020). Secure keyword search and data sharing mechanism for cloud computing. IEEE Transactions on Dependable and Secure Computing, 18(6), 2787-2800.

[62] Ge, C., Liu, Z., Xia, J., & Fang, L. (2019). Revocable identity-based broadcast proxy re-encryption for data sharing in clouds. IEEE Transactions on Dependable and Secure Computing, 18(3), 1214-1226.

[63] Ge, C., Susilo, W., Baek, J., Liu, Z., Xia, J., & Fang, L. (2021). Revocable attribute-based encryption with data integrity in clouds. IEEE Transactions on Dependable and Secure Computing.

[64] Ge, C., Susilo, W., Baek, J., Liu, Z., Xia, J., & Fang, L. (2021). A verifiable and fair attribute-based proxy re-encryption scheme for data sharing in clouds. IEEE Transactions on Dependable and Secure Computing, 19(5), 2907-2919.

[65] Wells, D., Beck, N., Kleusberg, A., Krakiwsky, E. J., Lachapelle, G., Langley, R. B., ... & Delikaraoglou, D. (1987). Guide to GPS positioning. In Canadian GPS Assoc.

[66] Pekonen, O. (2021). Cubed: The Puzzle of Us All by Ernő Rubik.

[67] OpenAI, I. A., Andrychowicz, M., Chociej, M., Litwin, M., McGrew, B., Petron, A., ... & Zhang, L. (2019). Solving rubik's cube with a robot hand.

[68] Agostinelli, F., McAleer, S., Shmakov, A., & Baldi, P. (2019). Solving the Rubik's cube with deep reinforcement learning and search. Nature Machine Intelligence, 1(8), 356-363.

[69] Knudsen, L. R., & Robshaw, M. J. (2011). Brute force attacks. In The Block Cipher Companion (pp. 95-108). Springer, Berlin, Heidelberg.

[70] Tournier, J., Lesueur, F., Le Mouël, F., Guyon, L., & Ben-Hassine, H. (2021). A survey of IoT protocols and their security issues through the lens of a generic IoT stack. Internet of Things, 16, 100264.

[71] Su, Y., & Ranasinghe, D. C. (2022). Leaving Your Things Unattended is No Joke! Memory Bus Snooping and Open Debug Interface Exploits. arXiv preprint arXiv:2201.07462.

[72] Sheikh, Z. A., & Singh, Y. (2021). Lightweight De-authentication DoS Attack Detection Methodology for 802.11 Networks Using Sniffer. In Proceedings of Second International Conference on Computing, Communications, and Cyber-Security (pp. 67-80). Springer, Singapore.

[73] Sain, M., Normurodov, O., Hong, C., & Hui, K. L. (2021, February). A Survey on the Security in Cyber Physical Systems with Multi-Factor Authentication. In 2021 23rd International Conference on Advanced Communication Technology (ICACT) (pp. 1-8). IEEE.

[74] Doshi, N. (2022). Cryptanalysis of authentication protocol for WSN in IoT based electric vehicle environment. Materials Today: Proceedings.

[75] Dai, C., & Xu, Z. (2022). A secure three-factor authentication scheme for multi-gateway wireless sensor networks based on elliptic curve cryptography. Ad Hoc Networks, 102768.

[76] Devis, Y., Irawan, Y., Zoromi, F., & Amartha, M. R. (2021, March). Monitoring System of Heart Rate, Temperature and Infusion in Patients Based on Microcontroller (Arduino Uno). In Journal of Physics: Conference Series (Vol. 1845, No. 1, p. 012069). IOP Publishing.

[77] Zhang, X., Song, M., Xu, Y., Dai, Z., & Zhang, W. (2021, May). Intelligent Door Lock System Based on Raspberry Pi. In 2021 2nd International Conference on Artificial Intelligence and Information Systems (pp. 1-7).

[78] Kortli, Y., Gabsi, S., Voon, L. F. L. Y., Jridi, M., Merzougui, M., & Atri, M. (2022). Deep embedded hybrid CNN-LSTM network for lane detection on NVIDIA Jetson Xavier NX. Knowledge-Based Systems, 107941.

[79] Almalki, F. A., & Soufiene, B. O. (2021). EPPDA: an efficient and privacy-preserving data aggregation scheme with authentication and authorization for IoT-based healthcare applications. Wireless Communications and Mobile Computing, 2021.

[80] Zuech, R., Hancock, J., & Khoshgoftaar, T. M. (2021, July). Detecting SQL Injection Web Attacks Using Ensemble Learners and Data Sampling. In 2021 IEEE International Conference on Cyber Security and Resilience (CSR) (pp. 27-34). IEEE.

[81] Gowtham, M., & Pramod, H. B. (2021). Semantic Query-Featured Ensemble Learning Model for SQL-Injection Attack Detection in IoT-Ecosystems. IEEE Transactions on Reliability.

[82] Shi, H., Chen, Y., & Hu, J. Y. (2021). Deep learning on information retrieval using agent flow email reply system for IoT enterprise customer service. Journal of Ambient Intelligence and Humanized Computing, 1-14.

[83] Mothukuri, V., Khare, P., Parizi, R. M., Pouriyeh, S., Dehghantanha, A., & Srivastava, G. (2021). Federated Learning-based Anomaly Detection for IoT Security Attacks. IEEE Internet of Things Journal.

[84] Medjek, F., Tandjaoui, D., Djedjig, N., & Romdhani, I. (2021). Multicast DIS attack mitigation in RPL-based IoT-LLNs. Journal of Information Security and Applications, 61, 102939.

[85] Yang, N., Chen, K., & Wang, M. (2021). SmartDetour: Defending Blackhole and Content Poisoning Attacks in IoT NDN Networks. IEEE Internet of Things Journal.

[86] Kumar, P., Kumar, R., Gupta, G. P., & Tripathi, R. (2021). A Distributed framework for detecting DDoS attacks in smart contract-based Blockchain-IoT Systems by leveraging Fog computing. Transactions on Emerging Telecommunications Technologies, 32(6), e4112.

[87] Tushir, B., Sehgal, H., Nair, R., Dezfouli, B., & Liu, Y. (2021). The Impact of DoS Attacks onResource-constrained IoT Devices: A Study on the Mirai Attack. arXiv preprint arXiv:2104.09041.

Plasmonics
https://doi.org/10.1007/s11468-023-01814-z

A Sustainable Approach to Develop Gold Nanoparticles with *Kalanchoe fedtschenkoi* and Their Interaction with Protein and Dye: Sensing and Catalytic Probe

Neha Bhatt[1] · **Mohan Singh Mehata**[1]

Received: 4 February 2023 / Accepted: 28 February 2023

Abstract

In this study, highly stable gold nanoparticles (AuNPs) of different sizes ranging from 15 to 55 nm were synthesized via an eco-friendly, sustainable, and cost-efficient approach using a specific plant, *Kalanchoe fedtschenkoi*. The AuNPs demonstrated an absorption maximum at around 525 nm, hence exhibiting a strong surface plasmon resonance (SPR) band that is created when the free electrons of the AuNPs oscillate in harmony with the frequency of incident light. The impact of physiochemical environments, pH, and temperature was examined. The crystal structure and stability of the produced AuNPs were validated with an X-ray diffractogram, zeta potential analysis, and absorption. The morphology, structure, and bonds were examined using HRTEM and FTIR, respectively. The interaction of AuNPs (concentrations range of 0–181 μM) with plasma protein bovine serum albumin was explored using absorption and fluorescence studies. Furthermore, AuNPs were utilized as an active catalyst for the degradation of dye methylene blue (MB) in the presence of $NaBH_4$. MB was degraded by 94%, and the solution became colorless within 16 min with a rate constant of 0.175 min^{-1}.

Keywords Gold nanoparticles · Surface plasmon resonance · Fluorescence · Protein · Catalyst · Sensing

Introduction

Nanotechnology, a science that encompasses several disciplines and involves chemistry, physics, biology, environment, medicine, and agriculture, has the potential to solve various problems such as drug delivery, solar energy conversions [1], wastewater treatment [2], and cancer treatment [3]. In recent years, the astounding advancements in nanotechnology have attracted researchers to engage themselves in developing reliable and efficient methods to produce nanomaterials ranging from 1 to 100 nm [4]. Because of the variation from their bulk counterparts in terms of optical, electronic, physiochemical, and magnetic properties, the interest in nanomaterials has intensified [5]. Major categories of nanostructures of biological significance include metallic, polymeric, carbon-based, semiconductor quantum dots, and magnetic nanoparticles. Quantum dots' size-dependent emission characteristics make them effective for biological identification and detection. For cell sorting, magnetic nanoparticles have been used [6].

Numerous chemical and physical techniques have been employed for the large-scale synthesis of various nanomaterials [7]. Chemical methods, including electrochemical process [8], precipitation [9], sonochemical route [10], sol–gel, hydrothermal [11], chemical bath deposition [12], chemical reduction [13], chemical vapor deposition [14], microemulsion technique, and microwave-assisted [15] synthesis, are the main techniques through the chemical approach using harsh reducing agents, organic compounds, and hazardous substances as well as producing hazardous by-products that are extremely damaging to the environment [16]. The physical methods of synthesis, such as plasma, pulsed laser, gamma radiation, vacuum vapor deposition [17], and mechanical milling, require high energy and are quite time-consuming. Given the limitations of chemical and physical processes, designing an efficient and ecologically friendly approach to producing nanomaterials is essential [18].

Gold nanostructures and nanoparticles (AuNPs) are one of the most commonly used noble metal nanoparticles

✉ Mohan Singh Mehata
msmehata@gmail.com

[1] Laser-Spectroscopy Laboratory, Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi 110042, India

Published online: 09 March 2023

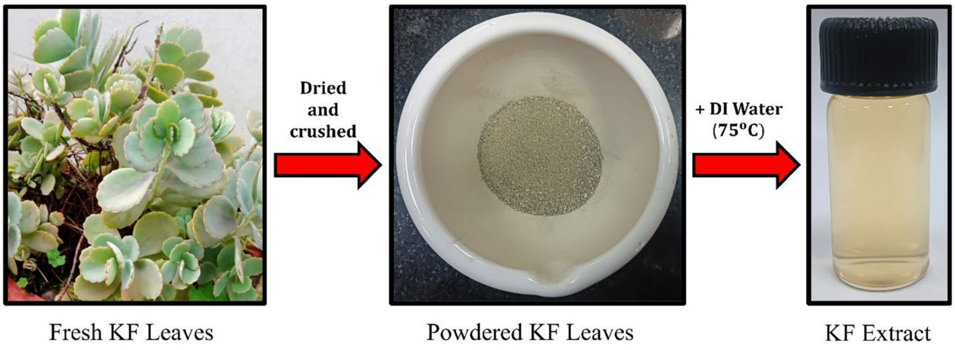
(NPs) and are applied in a variety of fields [19]. AuNPs have proven to be an efficient choice for purposes such as leukemia therapy [20], biomolecular immobilization [21], biosensor production [22], cancer therapy [23], antibacterial treatments [24], antimicrobial treatments [25], and labeling for contrast enhancement in cryoelectron microscopy [26]. Apart from biological applications, AuNPs have been utilized in various other applications, including catalysis, detection [27], and optoelectronic devices [28]. The surface plasmon resonance (SPR) observed in AuNPs, which depends on particle morphology and suspension medium, is responsible for a wide range of applications of AuNPs [29].

AuNPs have been prepared using several techniques, such as chemical, physical, and biological methods. The most effective and environmentally benign approaches are biological ones, which draw on natural resources like plant components, bacteria, yeasts, molds, enzymes, agricultural wastes, fungi, and algae [15]. The need for environmentally friendly nanoparticle production developed because physical and chemical procedures are expensive and environmentally harmful. Green synthesis of nanoparticles utilizes environment-friendly, non-toxic, and secure natural agents [30]. Since they are produced using a one-step process, nanoparticles created utilizing green technologies have a variety of natures, remarkable stability, and optimum sizes [31]. The issue of toxic surface compounds is not present in nanoparticle synthesis using biological techniques [16]. AuNPs were synthesized from various sources, for example, using *Annona squamosa L.* peel [15], *Zingiber officinale* extract [32], *honey* extract [33], *Murraya koenigii* leaf [34], *Rosa hybrida* petal [35], *Macrotyloma uniflorum* [36], *Adiantum philippense L. Frond* [37], *Punica granatum* [38], *Salvia officinalis*, *Lippia citriodora*, *Pelargonium graveolens* [39], *Dendropanax morbifera* leaf [40], *Couroupita guianensis* flower [3], *Trachyspermum ammi* seeds [5], *Allium ampeloprasum* leaf extract [41], *Nyctanthes arbortristis* flower [42], *Morinda citrifolia* leaf [43], and *Trigonella foenum-graecum* [44].

The *Kalanchoe* plant, which belongs to the *Crassulaceae* and is mostly found in Madagascar and Southeast Africa, is distributed worldwide in warm regions [45]. In these tropics, plants of the genus *Kalanchoe* are utilized as traditional remedies and have a variety of other ethnobotanical purposes. This species is employed as an analgesic in Brazil. The antibacterial properties of the plant were demonstrated by the growth inhibition exhibited by *K. fedtschenkoi* (KF) extracts displayed opposing gram-negative bacteria species such as *P. aeruginosa* and *A. baumannii*, as well as gram-positive bacteria *S. aureus* [46].

Bovine serum albumin (BSA) is crucial for maintaining the blood pH and osmotic pressure as well as for transporting, binding, and delivering numerous substances to their intended organs [47]. Since the structure and characteristics of BSA are well understood, it is utilized as a model for research of conformational changes following interaction with AuNPs [48]. The structure of BSA involves 583 amino acid residues forming a single polypeptide chain, and 17 disulfide links with a single thiol (SH) group. The BSA molecule is quite compact due to the presence of these disulfide bonds, which also help stabilize the helical structure of BSA. Fatty acids, which are insoluble in plasma, are transported mainly by BSA. The adsorption of serum albumins to metal oxides and the interaction of BSA with metal hydroxide suspensions have been thoroughly investigated [48]. However, it is known that the chemistry of the particle's surface and the protein's conformational state both significantly impact how thionine interact with AuNPs [49]. This makes it challenging to study the behavioral conformity of proteins for a nanoparticle-protein system, as protein adsorption can lead to the denaturation of the protein's tertiary and secondary structures [50]. Absorption and fluorescence spectroscopy are prominent techniques for examining the interactions between metals and proteins due to their high sensitivity and straightforwardness [51]. Tryptophan residues Trp 134 and Trp 213 have the greatest impact on BSA's fluorescence (FL) [52]. Tyrosine and phenylalanine (Phe) residues make up only a tiny percentage of the yield due to their low FL quantum yield. As reported previously, the alterations in the area around the microenvironment of residues may account for the variations of protein conformation on adding 2-azido acrylates [53]. The absorption or optical density maxima of BSA is located at 278 nm, with the fluorescence maxima of BSA appearing at 351 nm, quenched by adding AuNPs [54]. A putative conjugation mechanism is also proposed after looking at the conformational changes in BSA when interacting with AuNPs, based on the evidence gained using these approaches.

One of the main issues with environmental degradation is the water contamination brought on by industrial development. Water contamination is influenced by a number of variables, one of which is the presence of synthetic dye in wastewater [55]. The release of water waste containing many organic dyes might obstruct plant photosynthesis and sunlight absorption. In addition, a lot of synthetic dyes pose a serious threat to human health [56]. These challenges have been overcome using various techniques, including chemical oxidation, adsorption, fabric filtration, and catalytic degradation [57]. Owing to their novel physical, chemical, and electrical characteristics, which are different from their bulk counterparts, catalytic degradation using metal nanoparticles offers a convenient degrading approach for hazardous dyes among these techniques [58]. Using biocompatible, environmentally safe nanocatalyst to degrade toxic dyes is the simplest approach that does not require using organic solvents [59]. The aromatic dye methylene blue (MB) has a heterocyclic structure. The color of MB in its crystallized



Scheme 1 The schematic diagram for the preparation of plant extract

state is greenish-brown. MB solutions in water are blue. MB is a harmful industrial dye and is employed as a staining agent in the field of medicine [60].

In the present work, gold nanoparticles were synthesized utilizing a new plant named *Kalanchoe Fedtschenkoi* with a green, more efficient, and eco-friendly approach than other methods. Highly stable AuNPs were obtained using *Kalanchoe fedtschenkoi* plant extract, acting as a reducing and stabilizing agent. Furthermore, their interaction with BSA, the most abundant plasma protein, was considered. Also, these biosynthesized AuNPs were employed as an active catalyst for degrading MB dye.

Experimental Section

Chemicals

$HAuCl_4.H_2O$ (tetrachloroauric (III) acid), sodium hydroxide (NaOH), and sodium borohydride ($NaBH_4$) were procured from Sigma-Aldrich Chemicals Co. Bovine albumin fraction V (BSA) and methylene blue dye were procured from CDH Chemicals Ltd. The chemicals were all utilized in their original form without any modifications. Deionized water (DI), having a specific resistance of 18.2 MΩ cm, was employed as a solvent for all experiments.

Protein Solution

1.67 mg of BSA was added to 50 mL DI water to prepare a 0.5 µM solution. This solution was stirred for about 15 min to mix well and reached an equilibrium state. The solution was used for investigating the interaction between BSA and AuNPs through absorption and fluorescence analysis.

Dye Samples

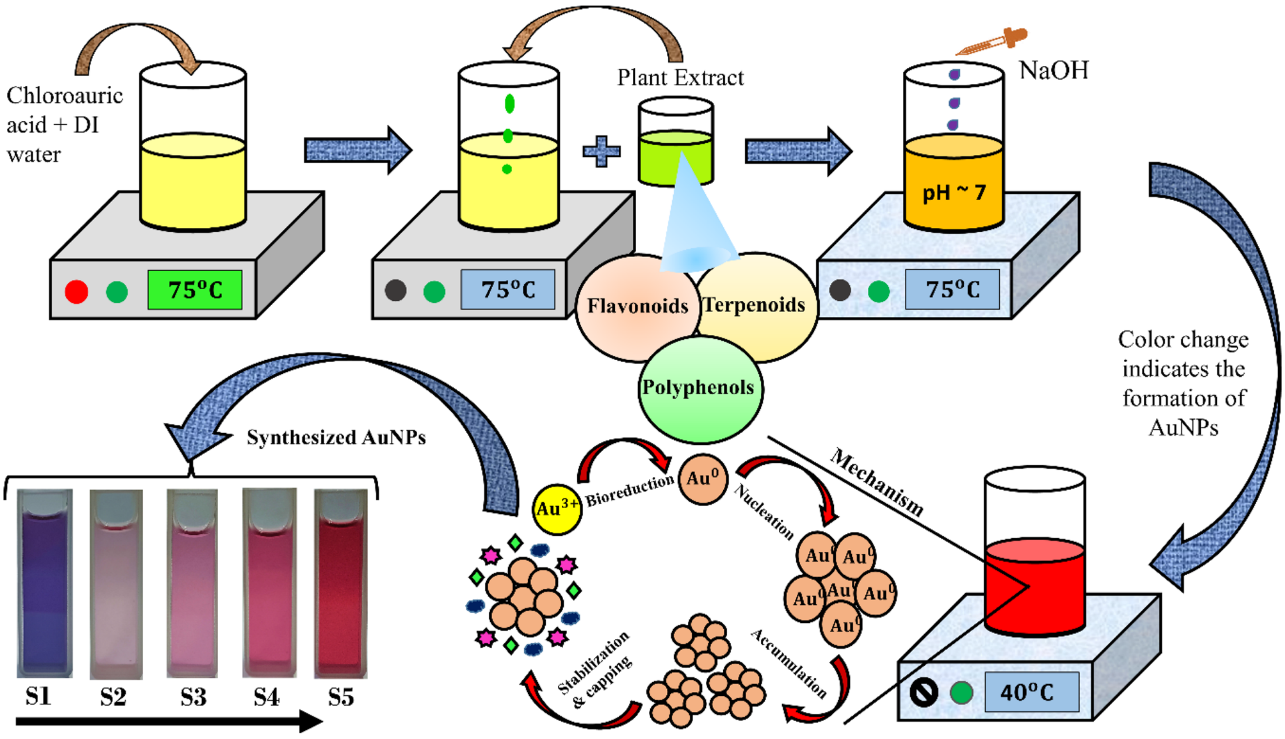
Initially, a stock solution of 50 µM of MB dye (formula: $C_{16}H_{18}N_3SCl$; M.M.: 319.85 g/mol) was prepared by mixing 200 mL of DI water and 3.2 mg of the dye. This was further diluted to 10 µM by adding 80 mL of DI water to 20 mL of the stock solution. This solution was divided into three vials with 20 mL each. The first solution was degraded by $NaBH_4$, the second by AuNPs, and the third by $NaBH_4$ + AuNPs. The amount of AuNPs added, 2 mL, was kept the same for both the second and third vials.

Preparation of Plant Extract

Fresh *Kalanchoe fedtschenkoi* (KF) leaves were picked from Dehradun, Uttarakhand, India, for use in this study. The leaves were further cleansed two to three times to remove surface impurities. The leaves were dried in the oven at 75 °C for about a day until all the surface moisture was obliterated. The leaves were further crushed to form a fine powder. Two grams of powder was boiled with 50 mL of DI water for 30 min at 75 °C, which was further filtered using the Whatman filter paper and kept at 4 °C in the refrigerator for further use. A schematic representation of the preparation of plant extract is illustrated in Scheme 1.

Gold Nanoparticle Synthesis

33.98 mg of $HAuCl_4$ was added to 100 mL of DI water to prepare 1 mM of tetrachloroauric acid solution. Forty milligrams of NaOH was stirred in 10 mL DI water to prepare a 0.1 M solution of NaOH. Ten milliliters of 1 mM chloroauric acid solution was placed in a conical flask and subjected to heating at 75 °C at 400 rpm for a duration of 15 min. Furthermore, 2 mL of *Kalanchoe fedtschenkoi* extract was added to the solution. The heating was turned off. A few drops of the basic solution of NaOH were introduced into the mixture to adjust the pH to ~ 7 and decrease the reaction time. The mixture changed from light yellow to various shades from colorless, purple, light pink, pink, and finally red as the reaction progressed over time. Various samples were formed, namely S1, S2, S3, S4, and S5, corresponding to their reaction time of 1, 2, 4, 6, and 8 h, respectively, to gain a deeper understanding of the formation mechanism and to control the size of the AuNPs. The longer and



Scheme 2 Synthesis of gold nanoparticles using the eco-friendly and cost-effective route

continuous reduction of gold nanoparticles leads to the formation of more uniform and symmetrical nanoparticles [61]. These samples were stored in the refrigerator for further examination. The illustration of the process of synthesizing gold nanoparticles is depicted in Scheme 2, along with the color change.

Characterization Techniques

UV–visible absorption and fluorescence spectroscopy are the most frequently employed methods to identify active species due to their robust functionality and high sensitivity, even to small samples. PerkinElmer, Lambda 750 UV/VIS/NIR dual beam spectrometer was utilized for the UV–vis spectroscopic studies. A Horiba Jobin Yvon Fluorolog-3 spectrofluorometer, equipped with a xenon lamp of 450 W and a photomultiplier tube, was used for steady-state FL and FL-excitation measurements. The sample container was a quartz cuvette with an optical path of 10 mm. Drop-casting was used to coat the colloidal AuNPs on a glass substrate in order to prepare a thin film of AuNPs to measure the X-ray diffractogram. BRUKER-D8 advanced was used to record the XRD pattern of a thin film of AuNPs. TALOS thermo-scientific instrument (Acc. Vol. 200 kV) was used to record the high-resolution transmission electron microscopic (HRTEM) images. The zeta potential of colloidal AuNPs and their size distribution were recorded using a Zetasizer nano series ZS (Malvern Panalytical). Fourier transform infrared (FTIR) studies in 400 to 4000 cm^{-1} were carried out using PerkinElmer two-spectrum FTIR spectrometer. Furthermore, BSA was used in experiments with increasing concentrations of AuNPs ranging from 0.91 to 181 μM.

Results and Discussion

X-ray Diffraction Analysis

Figure 1 represents the XRD pattern for biosynthesized AuNPs of sample S5 (a thin film of AuNPs overlay on a glass substrate) along with the JCPDS pattern of AuNPs. The XRD peaks occur at 2θ = 77.80°, 64.88°, 44.64°, and 38.40°, and were indexed as (311), (220), (200), and (111) planes, respectively, based on the FCC structure of AuNPs (JCPDS file no. 04–0784) [62]. The acquired XRD pattern showed that the synthesized AuNPs were crystallite in nature, which was confirmed by comparing it to the standard pattern for AuNPs. The intense diffraction peak at 38.40° indicates the favored direction of orientation in (111) direction [63]. This describes that molecular-sized structures have an identical spacing between each atom or molecule in a repeating 3D pattern [64]. The average crystallite size

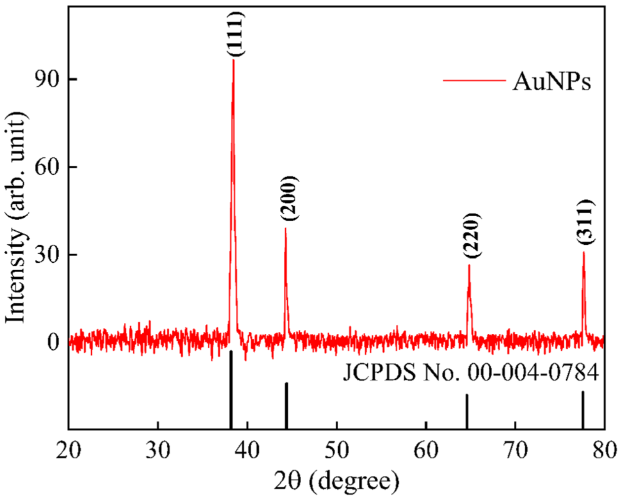


Fig. 1 XRD pattern of synthesized AuNP film

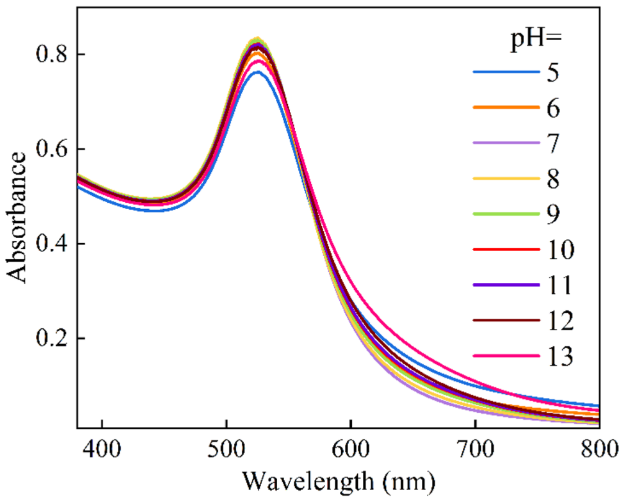


Fig. 3 Absorption spectra of colloidal AuNPs at different pH

estimated using Debye-Scherer's equation $D = 0.9\ \lambda/\beta\cos\theta$ was 18 nm.

UV–vis Absorption Spectra

The size, shape, refractive index, and interaction of gold colloids with their medium affect the SPR band of AuNPs appearing in the UV–vis spectrum [44]. It is observed that the maximum plasmon resonance peak of gold nanoparticles varies from 561 to 525 nm with varying average particle sizes from S1 to S5. Figure 2a shows the absorption spectra of KF extract, $HAuCl_4$ solution, and gold nanoparticles (S5). Figure 2b shows the normalized absorption spectra of five different-sized AuNPs. The SPR band of colloid S1 occurs at 561 nm. This long wavelength absorption is caused by the SPR occurring within the plane, which indicates a notable difference in the shape of the AuNPs [65]. The size of the nanoparticles may be correlated linearly with the absorption wavelength [51]. From the spectra, it can be observed that as the reaction time increased from 1 to 8 h, the SPR band was seen to shift towards the shorter wavelength, indicating a decrease in particle size. Therefore, it can be mentioned that the reaction time plays an important role in influencing the shape and size distribution of AuNPs.

Effect of pH

A pH of a solution has a great impact on controlling the shape and size of AuNPs. The absorption spectra of colloidal AuNPs produced at different pH levels between 5 and 13 are displayed in Fig. 3. Increasing the pH from acidic towards neutral (~5–8) increases the absorption intensity and reaches a maximum at pH 8. However, a further increase in pH beyond 8, from 8 to 13, results in a drop in absorption intensity. The electrical charges on biomolecules are altered by changes in pH, which also affects the peculiarities of the capping and stabilizing agents [61]. Although, the change in pH of the AuNPs colloidal solution did not bring any significant change in the position of the peak of the absorption band. Some aggregations may form with further increasing of pH [56]. This indicates that pH 8 is optimal for synthesizing AuNPs with *Kalanchoe fedtschenkoi* extract.

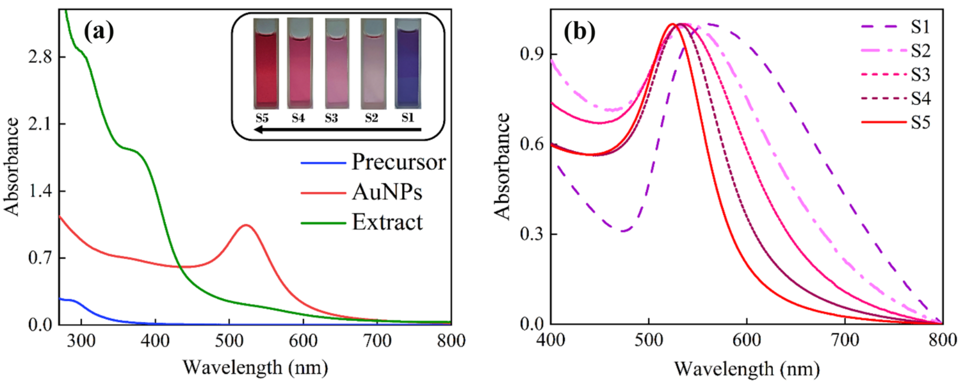


Fig. 2 Absorption spectra of **a** precursor, plant extract, and AuNPs (S5) along with color change (insets) and **b** normalized absorption spectra of different-sized AuNPs (S1, 52 nm to 19 nm, S5)

Effect of Temperature

The temperature can significantly impact the morphology of the synthesized AuNPs [66]. Figure 4 shows the absorption spectra of the synthesized colloidal AuNPs (S5) at temperatures ranging from 0 to 100 °C at an interval of 10 °C. The absorption maximum and intensity of AuNPs show no significant variation for numerous temperatures. However, at 90 and 100 °C, there is a slight decrease in the absorbance intensity, which might be due to the effect of agglomeration of nanoparticles at high temperatures. The observed results indicate that the synthesized AuNPs are stable at various temperatures and agree with the previous report [67].

Stability of AuNPs

One of the critical parameters determining the stability of AuNPs is the period at which they hold without significant change in their properties. The more stable AuNPs can be utilized successfully for various biomedical applications. The absorption spectra of synthesized AuNPs (S5) were measured at an interval of 10 days for about 4 months. Figure 5 shows the absorption spectra for AuNPs recorded for 110 days at an interval of 10 days. A slight decrease in the absorption intensity was noticed even after 110 days without any shift in the maximum absorption wavelength. Therefore, the AuNPs synthesized using *Kalanchoe fedtschenkoi* were far more stable than the stability observed in previous works [68], where the change in absorbance was quite significant in 10 days only.

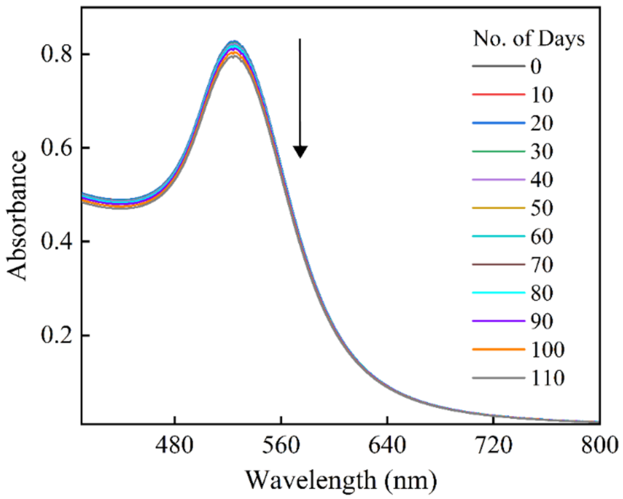


Fig. 5 Absorption spectra of colloidal AuNPs on different days show excellent stability

Zeta Potential Analysis

Zeta potential is a critical factor influencing the stability and morphology of colloidal suspensions [69]. It indicates the nature and magnitude of the charge associated with the particle. Zeta potential in colloidal suspensions denotes electrostatic repulsion between nearby, similar-charged particles [70]. In general, stable suspensions of colloidal nanoparticles are formed when the zeta potential values are more positive or negative than ± 30 mV form. This is due to the inter-particle electrostatic repulsion. The plot in Fig. 6 displays the distribution of particle sizes for sample S5 with the largest size intensity at

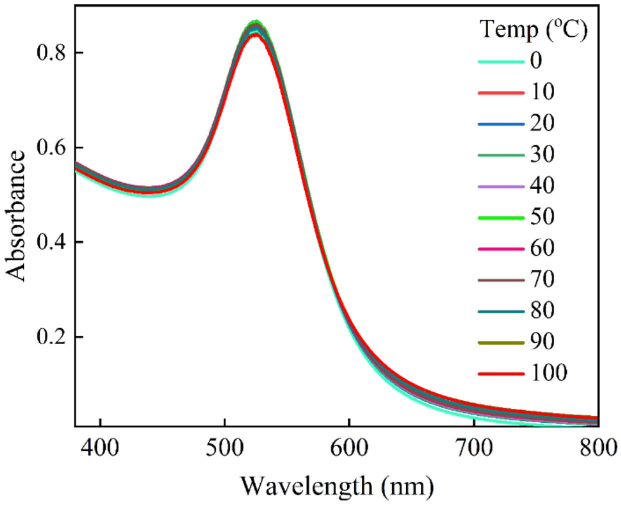


Fig. 4 Absorption spectra of colloidal AuNPs at different temperatures

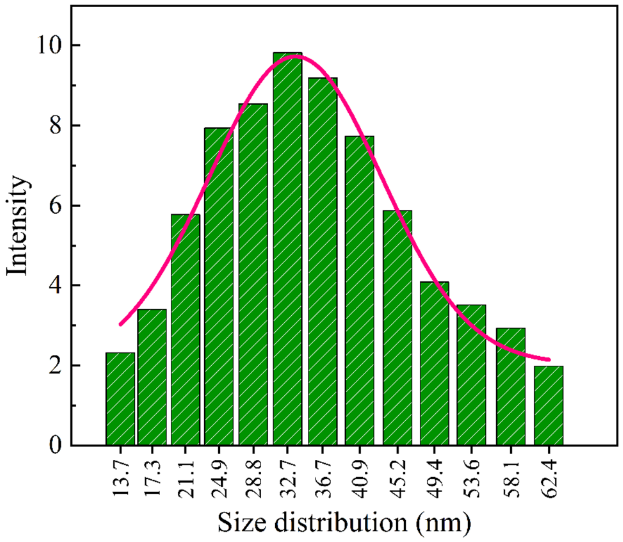


Fig. 6 Size distribution of colloidal AuNPs obtained from DLS analysis

32.7 nm. Thus, the average size of synthesized AuNPs is around 32.7 nm and the value of zeta potential recorded is – 29.6 mV, showing very high stability of the synthesized AuNPs as compared to previously reported zeta potential values [71]. Dynamic light scattering (DLS) uses the scattered light intensity as a parameter for the measurement of the hydrodynamic diameter of a sample, which possibly explains the increase in average size compared to crystallite size [72].

HRTEM Analysis

Figure 6 shows the HRTEM images at various magnifications for samples S1, S3, and S5. Figure 7a and b shows the images of sample S1 of biosynthesized AuNPs at magnifications of 20 and 50 nm, respectively, along with the magnification of 5 nm (the inset of a). These images indicate that sample S1 contains particles having different shapes, such as oval and spherical. Figure 7c illustrates the particle size distribution obtained using HRTEM images with the average size of sample S1 is 52 nm. Figure 7d and e shows the images for sample S3, indicating that upon increasing synthesis time, the particles tend to acquire a more symmetrical and uniform shape. Figure 7f represents the size distribution of S3, which indicates the average particle size for S3 to be 21 nm. With a further increase in synthesis time, the nanoparticles will gain a more uniform and symmetrical shape,

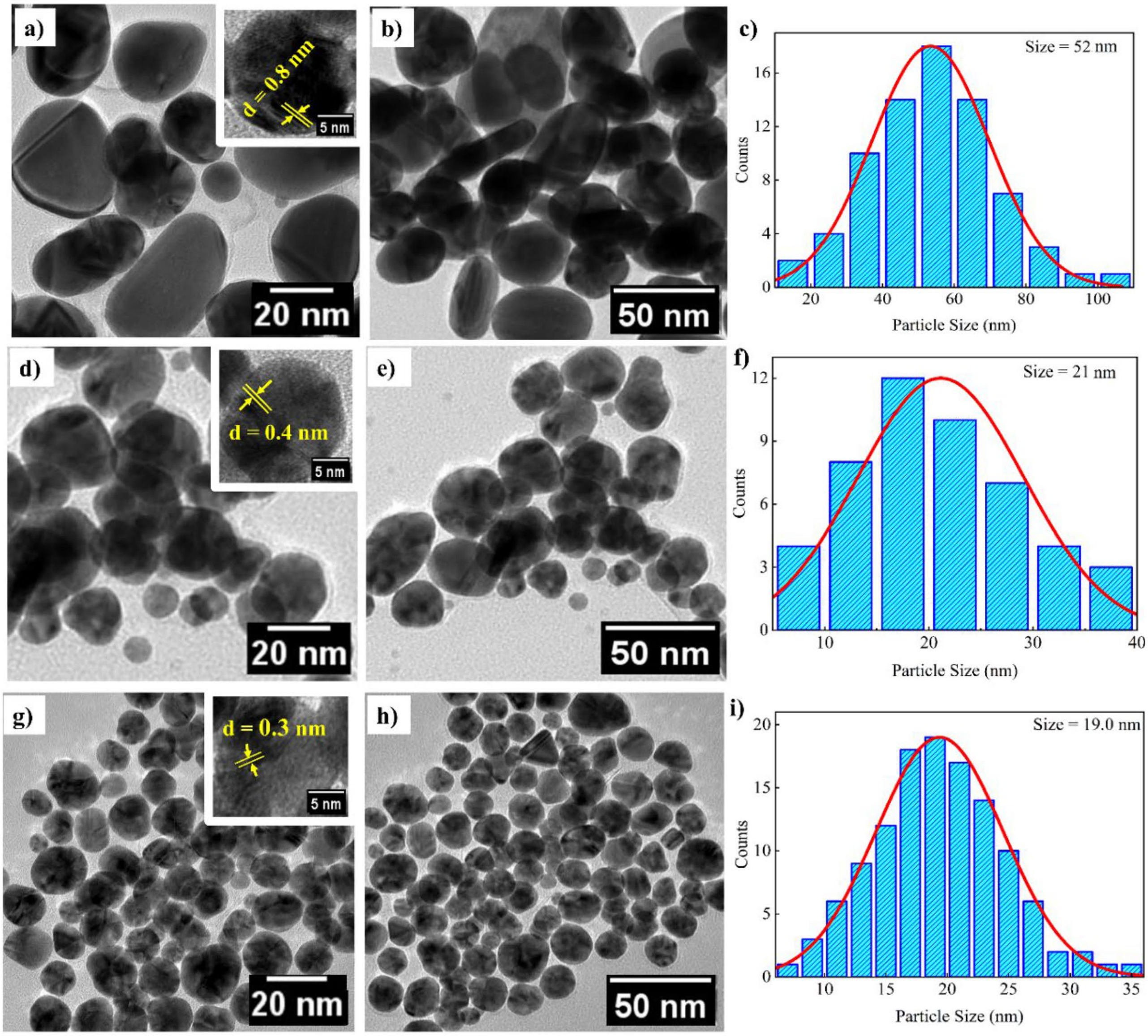


Fig. 7 HRTEM images at a magnification of 20 and 50 nm along with particle size distribution of S1 (**a**, **b**, **c**), S3 (**d**, **e**, **f**), and S5 (**g**, **h**, **i**), respectively. The inset of (**a**, **d**, **g**) represents the image at a 5 nm magnification

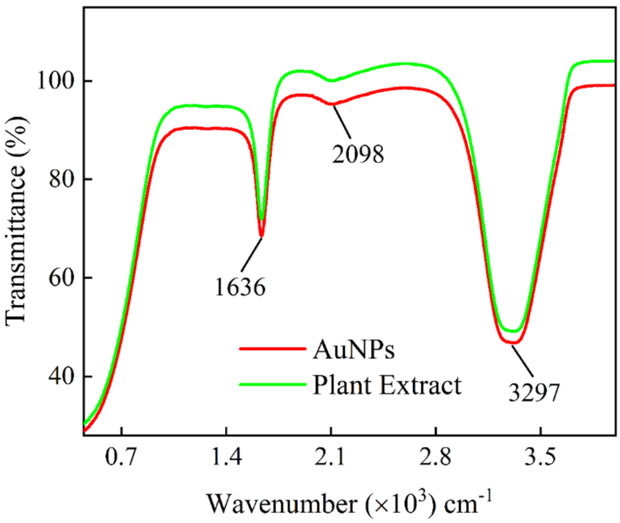


Fig. 8 FTIR spectra of biosynthesized colloidal AuNPs and plant extract

as shown in Fig. 7g and h, showing spherical gold nanoparticles for sample S5, which is likely due to the increased opportunity for nucleation and growth of the nanoparticles [73]. Figure 7i illustrates the particle size distribution of S5, indicating an average particle size of AuNPs to be 19 nm, close to crystallite size.

FTIR Analysis

Several phytochemicals and biomolecules have been reported to be present in *Kalanchoe fedtschenkoi* plant, including organic acids such as malic acid and citric acid, flavonoids such as quercetin and kaempferol, alkaloids such as bufadienolides and glycosides, and polysaccharides [74]. FTIR analysis of gold nanoparticles synthesized using *Kalanchoe fedtschenkoi* extract was measured to detect different functional groups involved in the formation of AuNPs. The FTIR spectra of sample S5 and the plant extract are shown in Fig. 8. The broader peak recorded at 3297 cm^{-1} can be attributed to the vibrations of the hydroxyl (O–H) bond, which indicates the presence of alcoholic and phenolic compounds [75] and is also observed in terpene and fatty acids. The band at 1636 cm^{-1} can be due to the stretching vibrations of C=C bonds [76]. The presence of an aromatic component is evident by the weak band observed at 2098 cm^{-1} [77]. The FTIR analysis showed the presence of hydroxyl, carbonyl, and carboxyl groups, which are commonly found in flavonoids, organic acids, and polysaccharides, which are possibly responsible for the reduction and stabilization of AuNPs.

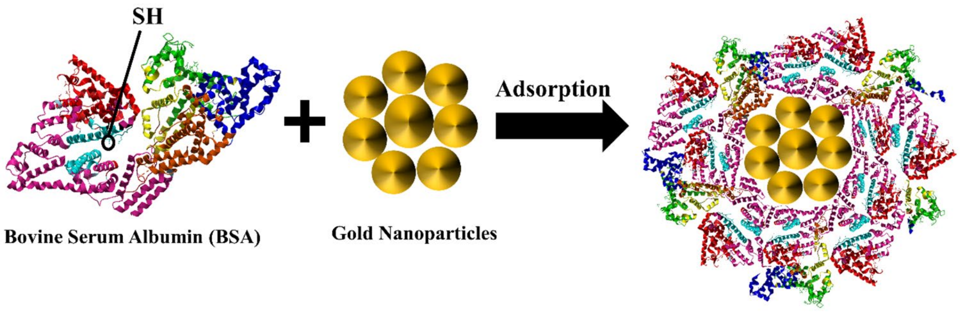
Interaction of AuNPs with BSA

The structure of BSA protein is defined by 583 amino acid residues forming a single polypeptide chain and 17 disulfide links and a single thiol (SH) group. The possible mechanism underlying the interaction of the protein with AuNPs may be passive adsorption [78], in which particular charge functional protein groups are joined to the surface of gold nanoparticles forming covalent or non-covalent interactions. Figure 9 shows the pictorial representation of BSA adsorption on AuNPs. In BSA, the SH group in the albumin cysteine residues interacts with the Au atoms on the surface of AuNPs, initiating the creation of Au–S covalent bonds. Because BSA contains binding sites, direct adsorption could be accomplished by simply incubating gold nanoparticles with BSA. Figure 10 shows the surface diagram of BSA illustrating the interaction of the thiol group of BSA with AuNPs, possibly in the form of adsorption, where the thiol group of BSA binds with the Au atoms present on the surface of the AuNPs, leading to the formation of a stable complex between BSA and AuNPs [79]. The methodological ease and economy of this adsorption strategy, which avoids the employment of extra reagents and extreme conditions, make it exceptional and sustainable.

Absorption of BSA

The interaction between BSA and AuNPs was examined by measuring the absorption spectra of BSA, along with the

Fig. 9 Plausible route of adsorption of BSA on AuNPs



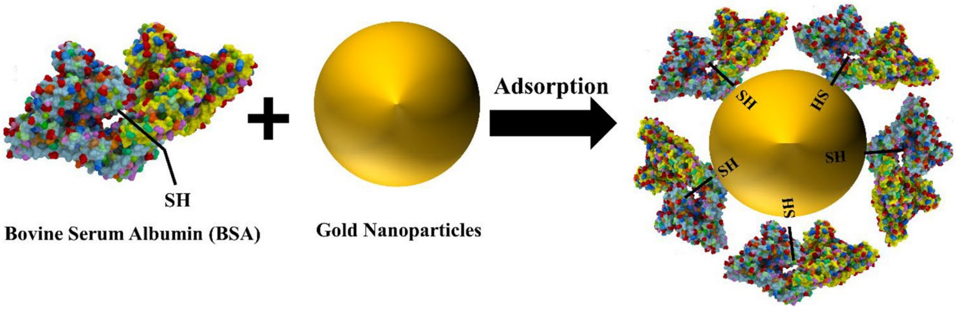


Fig. 10 Possible interaction of SH group of BSA with AuNPs

increasing concentration of AuNPs from 0.9 to 181 μM. Figure 11 shows that BSA has a strong absorption band at 278 nm. The absorption band intensity gradually increases along with the rise in the concentration of AuNPs with no significant shift in absorption maxima. The stable complex in the ground state formed due to interaction between BSA and AuNPs, as the thiol group of BSA binds with the gold atoms present on the surface of AuNPs, may be the plausible cause of the increase in intensity [80]. As observed, the concentration of AuNPs used has no discernible optical density in the region of BSA absorption spectra; the enhanced BSA absorption is most likely the result of forming a ground-state stable complex due to intermolecular interactions [81].

Figure 12 illustrates the Benesi-Hildebrand (B-H) absorption plot for increasing the concentration of AuNPs. The binding constant K_b was determined using the method reported by [82] using Eq. (1).

$$\frac{1}{A - A_0} = \frac{1}{A_{co} - A_0} + \frac{1}{K_b\left(A_{co} - A_0\right)[Q]} \tag{1}$$

where A represents the absorbance of BSA with different concentrations of AuNPs at 278 nm, A_0 and A_{co} indicate the absorbance of BSA at initial concentration and in the presence of AuNPs at 278 nm, respectively, and $[Q]$ is the AuNP concentration in M. The plot of $1/(A - A_0)$ vs. $1/[Q]$ is linear with a slope that equals to $1/K_b(A_{co} - A_0)$ and intercept, which equals to $1/(A_{co} - A_0)$. The plot showed a linear relation with $R^2 = 0.99$ with a value of K_b as $4 \times 10^4 \mathrm{M}^{-1}$, hence showing a strong binding.

Fluorescence of BSA

The interaction of BSA with AuNPs was monitored by measuring the change in fluorescence (FL) intensity, which was quenched by increasing concentrations of AuNPs from 0 to 181 μM. The strong FL band of BSA at 351 nm is shown in Fig. 13. In the presence of AuNPs, BSA's FL intensity reduces, indicating that the former interacts with one of the protein's two tryptophan residues (Trp-134 or Trp-213) [83]. It is significant to notice that in the experimental conditions, the BSA's excitation wavelength (280 nm) does not coincide with the SPR peak (525 nm) of AuNPs, demonstrating that the quenching process is carried out by nanoparticles [51]. The thiol group of BSA molecules gets adsorbed on the surface of AuNPs. When the binding site is close to AuNPs, FL from the tryptophan moiety

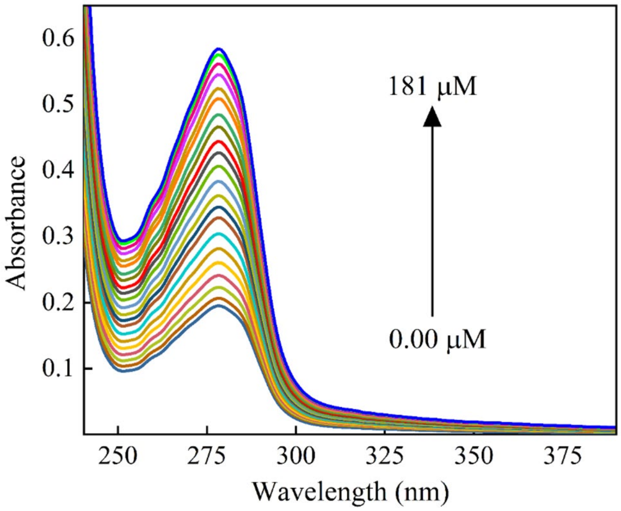


Fig. 11 Absorption spectra of BSA (0.5 μM) with increasing concentration of colloidal AuNPs

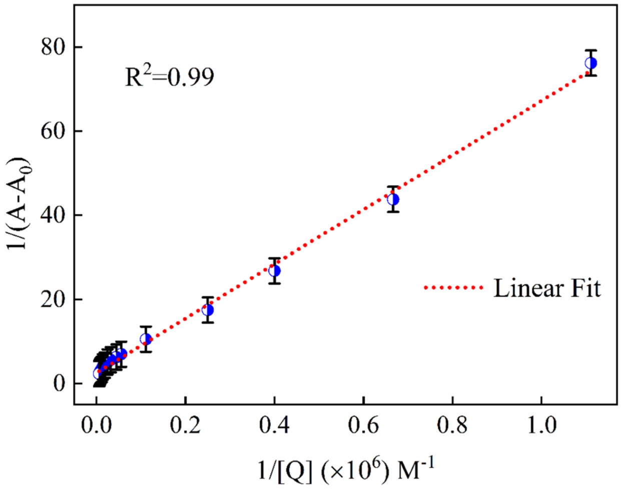


Fig. 12 Plot (B-H) of $(\frac{1}{A-A_0})$ *vs.* $1/[Q]$

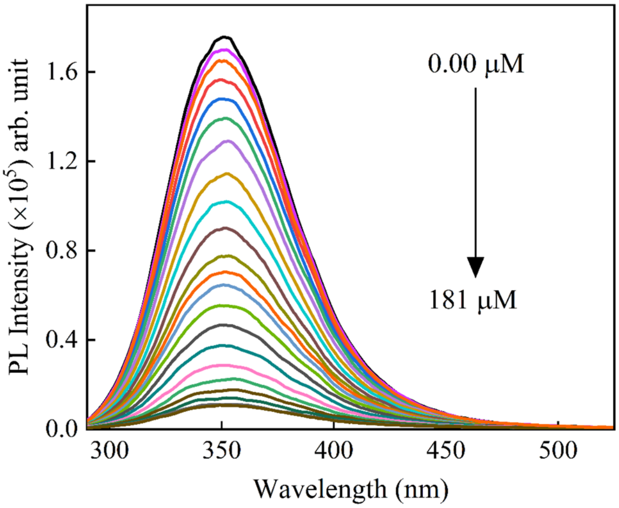


Fig. 13 Fluorescence spectra of BSA (0.5 μM) with increasing concentration of AuNPs

of BSA is quenched, and the free BSA in the solution emits the remaining fluorescence [84]. As an outcome, the un-adsorbed probe molecule of the BSA is responsible for the signal contribution to the FL spectra [52]. The linear Stern–Volmer indicates that there is only one sort of quenching in the system. Considering that internal energy transfer requires a good overlap between the FL and absorption spectra of the donor and acceptor [51], due to the significant Stokes shift, the resulting overlap between the FL and absorption spectra is insufficient for enabling the energy transfer process.

To investigate the mechanism of quenching, FL intensity was recorded with varying AuNP concentrations, and the Stern–Volmer (S-V) plot was obtained (Fig. 14) with Eq. 2 [85].

$$\frac{F_0}{F} = 1 + K_{SV}[Q] \tag{2}$$

where F_0 and F represent FL intensities in the absence and presence of AuNPs, respectively. The S-V plot revealed a linear relationship between the concentration of AuNPs and FL intensity with $R^2 = 0.99$. K_{SV}, referred to as the S-V constant or the quenching constant, is estimated to be $7.2 \times 10^4 \mathrm{M}^{-1}$ using Eq. (2). The corresponding limit of detection (LoD) for AuNPs was calculated using $3\sigma/\mathrm{K}$ [86], where σ indicates the standard deviation and K is the slope of the plot, to be 6 μM.

Catalytic Performance of AuNPs

Additionally, the well-crystalline AuNPs were used for the catalytic reaction and to degrade the textile dye methylene blue (MB) in the presence of $\mathrm{NaBH_4}$, serving as a reference and reducing agent. As observed, MB dye deteriorated

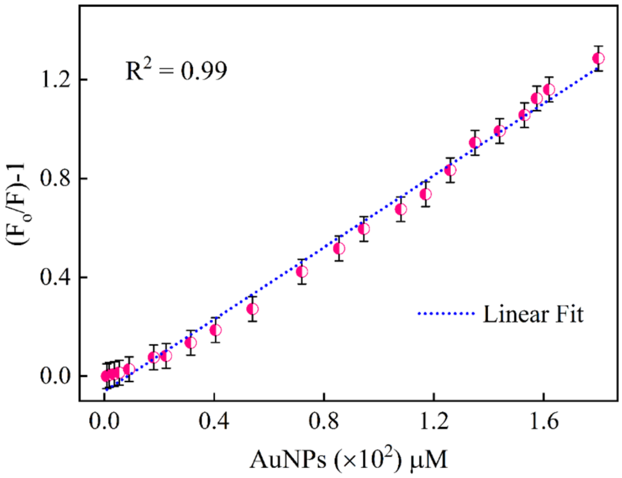


Fig. 14 The plot of (F_0/F)-1 *vs.* concentration of AuNPs

from a vivid blue to almost colorless after 16 min (insets of Fig. 15c). Figure 15 represents the absorption spectra of methylene blue dye in DI water with $\mathrm{NaBH_4}$ in the presence and absence of AuNPs. At around 664 nm, MB dye exhibits its distinctive lower energy absorption band, which corresponds to the $n - \pi^*$ transitions [56]. MB is carcinogenic and mutagenic to living things and is toxic in natural water, and its concentration in the body can be hazardous [55]. The absorption intensity was somewhat reduced in the addition of $\mathrm{NaBH_4}$ and AuNPs alone (Fig. 15a, b), demonstrating no discernible color change of MB solutions (insets of Fig. 15a, b) and the MB dye degraded was 29 and 18%, respectively. However, the addition of $\mathrm{NaBH_4}$ reduced the amount of MB and stabilized it, which led to a minor decrease in absorption intensity. After adding a modest amount of AuNPs, the absorption intensity reduces steadily and almost entirely after 16 min. The absorption intensity is decreased by about 94%, a point at which the color is lost (Fig. 15c). AuNPs act as a charge carrier and initialize a transfer of electrons from nucleophilic $\mathrm{BH_4^-}$ ions to electrophilic dye molecules [87], which might be a plausible reason for the reduction of MB dye to leucomethylene blue [55]. AuNPs perform as active catalysts in the reduction of MB dye using $\mathrm{NaBH_4}$ by providing a surface for $\mathrm{BH_4^-}$ donor ions to get adsorbed [88]. The absorption peak of MB dye arises around 300 nm (not shown), increases initially, and decreases subsequently, which is plausibly due to the overlap of the absorption of plant extract and due to the degradation of dye [87]. The presence of AuNPs increases the concentration of active sites, allowing the reaction to occur faster.

However, when used alone, AuNPs and $\mathrm{NaBH_4}$ provide electrons for dye reduction reactions, but their combined use is far more efficient. The degradation percentage was calculated using $\left(\frac{A_0 - A_t}{A_0}\right) \times 100\%$, where A_0 and A_t indicate the absorbance inten-

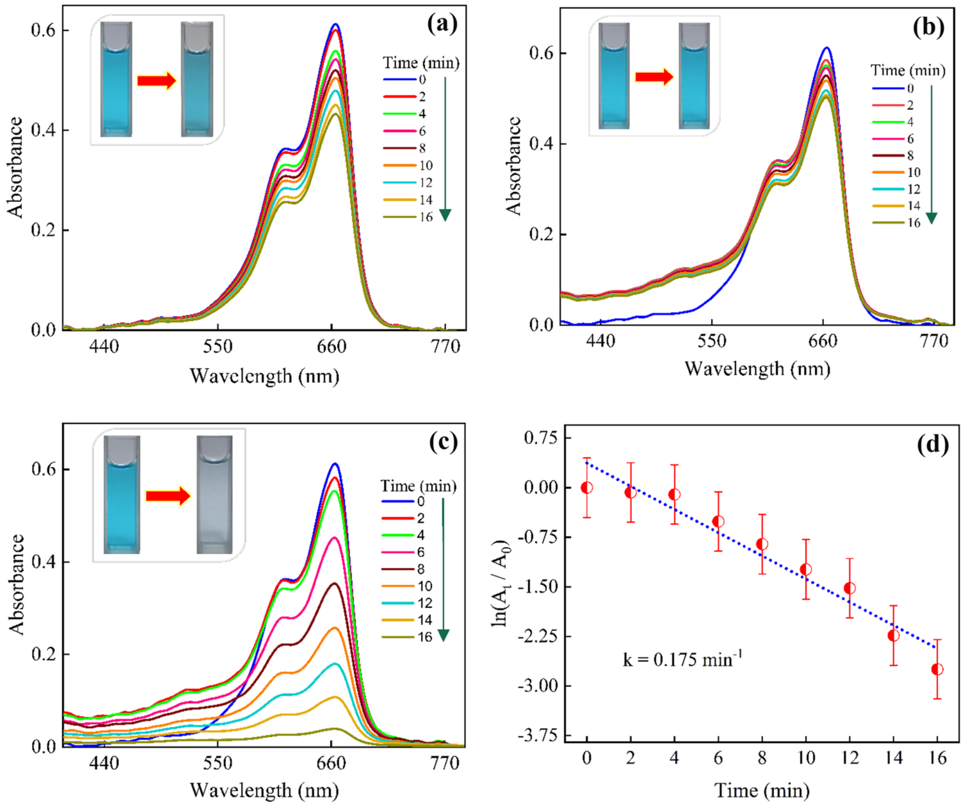


Fig. 15 Absorption spectra of methylene blue with reaction time in the presence of $NaBH_4$ (**a**), AuNPs (**b**), and $NaBH_4$ + AuNPs (**c**). The plot of ln ($\frac{A_t}{A_0}$) of MB with $NaBH_4$ + AuNPs as a function of time (**d**)

sities of dye at 664 nm initially at time $t = 0$ (pure dye) and at time t, respectively. The rate constant for the degradation of MB dye is calculated using the relation $\ln\left(\frac{A_t}{A_0}\right) = -kt$, where k is the rate constant of reaction and t represents the reaction time [56]. The plot of $\ln\left(\frac{A_t}{A_0}\right)$ *vs*. t is given in Fig. 15d, which shows a linear relation, showing the pseudo-first-order reaction kinetics of degradation reaction of MB dye [59]. The rate constant (k) for the degradation of MB dye using AuNPs with sodium borohydride was determined to be 0.175 min^{-1}.

AuNPs with $NaBH_4$ degraded the dye by 94% in 16 min, which is quite faster than the degradation time reported in previous reports [89, 90], hence demonstrating a more efficient catalyst while synthesized using *Kalanchoe fedtschenkoi* plant. However, a faster rate has also been reported in some reports [55, 91]. This variation and slowness in the reaction rate in the present system can be attributed to the differences in the synthesis approaches, various physiochemical factors, and the concentration and amount of different materials used for the experimental process. Further optimization of these parameters could potentially lead to even faster degradation rates.

Conclusion

Highly efficient gold nanoparticles were successfully synthesized using an environmentally friendly and cost-effective approach using the leaves of the plant *Kalanchoe fedtschenkoi*. The morphology and size of produced AuNPs reveal a consistent spherical shape. The AuNPs exhibited excellent crystalline structure having an average particle size of 19 nm. The AuNPs show a strong SPR band at 525 nm. The AuNPs are highly stable and examined with a zeta potential of − 29.6 mV and absorption spectra of 4 months since the absorption spectra did not significantly change over time.

Furthermore, interactions of AuNPs with BSA, forms a very important component of plasma which functions as a drug carrier and helps digest fatty acids, were examined by recording the change in absorption and FL intensities of BSA with increasing concentration of AuNPs. The fluorescence intensity of BSA was quenched following the linear S-V relation with LoD of 6 µM. The adsorption of BSA on the surface of AuNPs indicated improved drug transfer properties of BSA and extra stability of AuNPs.

Employment of AuNPs degraded textile dye MB remarkably in 16 min. Ninety-four percent of the dye was degraded in the presence of $NaBH_4$ + AuNPs, while degraded only by 29 and 18% in the presence of $NaBH_4$ and AuNPs, respectively. In addition, AuNPs have potential applications in areas such as antibacterial, antifungal, antioxidant, antimicrobial, and anticancer properties.

Acknowledgements The author Ms. Neha Bhatt expresses gratitude to Ms. Aneesha and other lab scholars for their constant support and valuable motivation throughout this study.

Author Contribution Neha Bhatt: data curation; formal analysis; investigation; methodology; writing—original draft. Mohan Singh Mehata: supervision; conceptualization; funding acquisition; resources; validation; writing—review and editing.

Data Availability All data generated or analyzed during this study are included in this published article.

Declarations

Consent for Publication All authors have approved the paper and agree with its publication.

Competing Interests The authors declare no competing interests.

References

1. Jain S, Mehata MS (2017) Medicinal plant leaf extract and pure flavonoid mediated green synthesis of silver nanoparticles and their enhanced antibacterial property. Sci Rep 7:15867. https://doi.org/10.1038/s41598-017-15724-8
2. Dasgupta N, Ranjan S, Ramalingam C (2017) Applications of nanotechnology in agriculture and water quality management. Environ Chem Lett 15:591–605. https://doi.org/10.1007/s10311-017-0648-9
3. Geetha R, Ashokkumar T, Tamilselvan S et al (2013) Green synthesis of gold nanoparticles and their anticancer activity. Cancer Nanotechnol 4:91–98. https://doi.org/10.1007/s12645-013-0040-9
4. Das M, Shim KH, An SSA, Yi DK (2011) Review on gold nanoparticles and their applications. Toxicol Environ Health Sci 3:193–205. https://doi.org/10.1007/s13530-011-0109-y
5. Perveen K, Husain FM, Qais FA et al (2021) Microwave-assisted rapid green synthesis of gold nanoparticles using seed extract of trachyspermum ammi: Ros mediated biofilm inhibition and anticancer activity. Biomolecules 11:197. https://doi.org/10.3390/biom11020197
6. Kolhatkar AG, Jamison AC, Litvinov D et al (2013) Tuning the magnetic properties of nanoparticles. Int J Mol Sci 14:15977–16009. https://doi.org/10.3390/ijms140815977
7. Umer A, Naveed S, Ramzan N, Rafique MS (2012) Selection of a suitable method for the synthesis of copper nanoparticles. NANO 7:1230005. https://doi.org/10.1142/S1793292012300058
8. Anand V, Srivastava VC (2015) Zinc oxide nanoparticles synthesis by electrochemical method: optimization of parameters for maximization of productivity and characterization. J Alloys Compd 636:288–292. https://doi.org/10.1016/j.jallcom.2015.02.189
9. Lin CC, Wei YL (2020) Enhanced reactivity of copper nanoparticles mass-produced by reductive precipitation in a rotating packed bed with blade packings. J Mater Res Technol 9:12328–12334. https://doi.org/10.1016/j.jmrt.2020.08.080
10. Silva N, Ramírez S, Díaz I et al (2019) Easy, quick, and reproducible sonochemical synthesis of CuO nanoparticles. Materials 12:804. https://doi.org/10.3390/MA12050804
11. Aneesh PM, Vanaja KA, Jayaraj MK (2007) Synthesis of ZnO nanoparticles by hydrothermal method. Nanophotonic Mater IV 6639:66390J. https://doi.org/10.1117/12.730364
12. Téllez VC, Portillo MC, Santiesteban HJ et al (2021) Green synthesis of palladium mixed with PdO nanoparticles by chemical bath deposition. Opt Mater (Amst) 112:110747. https://doi.org/10.1016/j.optmat.2020.110747
13. Daruich De Souza C, Ribeiro Nogueira B, Rostelato MECM (2019) Review of the methodologies used in the synthesis gold nanoparticles by chemical reduction. J Alloys Compd 798:714–740. https://doi.org/10.1016/j.jallcom.2019.05.153
14. Lassègue P, Noé L, Monthioux M, Caussat B (2017) Fluidized bed chemical vapor deposition of copper nanoparticles on multi-walled carbon nanotubes. Surf Coatings Technol 331:129–136. https://doi.org/10.1016/j.surfcoat.2017.10.046
15. Gangapuram BR, Bandi R, Alle M et al (2018) Microwave assisted rapid green synthesis of gold nanoparticles using Annona squamosa L peel extract for the efficient catalytic reduction of organic pollutants. J Mol Struct 1167:305–315. https://doi.org/10.1016/j.molstruc.2018.05.004
16. Virkutyte J, Varma RS (2013) Green synthesis of nanomaterials: environmental aspects. ACS Symp Ser 1124:11–39. https://doi.org/10.1021/bk-2013-1124.ch002
17. Sheth P, Sandhu H, Singhal D et al (2012) Nanoparticles in the pharmaceutical industry and the use of supercritical fluid technologies for nanoparticle production. Curr Drug Deliv 9:269–284. https://doi.org/10.2174/156720112800389052
18. Nasrollahzadeh M, Sajjadi M, Iravani S, Varma RS (2021) Green-synthesized nanocatalysts and nanomaterials for water treatment: current challenges and future perspectives. J Hazard Mater 401:123401. https://doi.org/10.1016/j.jhazmat.2020.123401
19. Shakibaie M, Forootanfar H, Mollazadeh-Moghaddam K et al (2010) Green synthesis of gold nanoparticles by the marine microalga Tetraselmis suecica. Biotechnol Appl Biochem 57:71–75. https://doi.org/10.1042/ba20100196
20. Mukherjee P, Bhattacharya R, Bone N et al (2007) Potential therapeutic application of gold nanoparticles in B-chronic lymphocytic leukemia (BCLL): enhancing apoptosis. J Nanobiotechnology 5:4. https://doi.org/10.1186/1477-3155-5-4
21. Petkova GA, Záruba K, Žvátora P, Král V (2012) Gold and silver nanoparticles for biomolecule immobilization and enzymatic catalysis. Nanoscale Res Lett 7:287. https://doi.org/10.1186/1556-276X-7-287
22. Zeng S, Yong KT, Roy I et al (2011) A review on functionalized gold nanoparticles for biosensing applications. Plasmonics 6:491–506. https://doi.org/10.1007/s11468-011-9228-1
23. Cai W, Gao T, Hao Hong JS (2008) Applications of gold nanoparticles in cancer nanotechnology. Nanotechnol Sci Appl 1:17–32. https://doi.org/10.4018/978-1-5225-3158-6.ch035
24. Sathiyaraj S, Suriyakala G, Dhanesh Gandhi A et al (2021) Biosynthesis, characterization, and antibacterial activity of gold nanoparticles. J Infect Public Health 14:1842–1847. https://doi.org/10.1016/j.jiph.2021.10.007
25. Zhang Y, Shareena Dasari TP, Deng H, Yu H (2015) Antimicrobial activity of gold nanoparticles and ionic gold. J Environ Sci Heal - Part C Environ Carcinog Ecotoxicol Rev 33:286–327. https://doi.org/10.1080/10590501.2015.1055161
26. Beales CT, Medalia O (2022) Gold nanomaterials and their potential use as cryo-electron tomography labels. J Struct Biol 214:107880. https://doi.org/10.1016/j.jsb.2022.107880
27. Jans H, Huo Q (2012) Gold nanoparticle-enabled biological and chemical detection and analysis. Chem Soc Rev 41:2849–2866. https://doi.org/10.1039/c1cs15280g
28. Shkir M, Yahia IS, Ganesh V et al (2018) A facile synthesis of Au-nanoparticles decorated PbI_2 single crystalline nanosheets for optoelectronic device applications. Sci Rep 8:13806. https://doi.org/10.1038/s41598-018-32038-5
29. El-Brolossy TA, Abdallah T, Mohamed MB et al (2008) Shape and size dependence of the surface plasmon resonance of gold nanoparticles studied by photoacoustic technique. Eur Phys J Spec Top 153:361–364. https://doi.org/10.1140/epjst/e2008-00462-0
30. García-Quintero A, Palencia M (2021) A critical analysis of environmental sustainability metrics applied to green synthesis of nanomaterials and the assessment of environmental risks associated with the nanotechnology. Sci Total Environ 793:148524. https://doi.org/10.1016/j.scitotenv.2021.148524

31. Khalaj M, Kamali M, Costa MEV, Capela I (2020) Green synthesis of nanomaterials - A scientometric assessment. J Clean Prod 267:122036. https://doi.org/10.1016/j.jclepro.2020.122036
32. Kumar KP, Paul W, Sharma CP (2011) Green synthesis of gold nanoparticles with Zingiber officinale extract: characterization and blood compatibility. Process Biochem 46:2007–2013. https://doi.org/10.1016/j.procbio.2011.07.011
33. Philip D (2009) Honey mediated green synthesis of gold nanoparticles. Spectrochim Acta - Part A Mol Biomol Spectrosc 73:650–653. https://doi.org/10.1016/j.saa.2009.03.007
34. Philip D, Unni C, Aromal SA, Vidhu VK (2011) Murraya Koenigii leaf-assisted rapid green synthesis of silver and gold nanoparticles. Spectrochim Acta - Part A Mol Biomol Spectrosc 78:899–904. https://doi.org/10.1016/j.saa.2010.12.060
35. Noruzi M, Zare D, Khoshnevisan K, Davoodi D (2011) Rapid green synthesis of gold nanoparticles using Rosa hybrida petal extract at room temperature. Spectrochim Acta - Part A Mol Biomol Spectrosc 79:1461–1465. https://doi.org/10.1016/j.saa.2011.05.001
36. Aromal SA, Vidhu VK, Philip D (2012) Green synthesis of well-dispersed gold nanoparticles using Macrotyloma uniflorum. Spectrochim Acta - Part A Mol Biomol Spectrosc 85:99–104. https://doi.org/10.1016/j.saa.2011.09.035
37. Sant DG, Gujarathi TR, Harne SR et al (2013) Adiantum philippense L. Frond assisted rapid green synthesis of gold and silver nanoparticles. J Nanoparticles 2013:1–9. https://doi.org/10.1155/2013/182320
38. Ahmad N, Sharma S, Rai R (2012) Rapid green synthesis of silver and gold nanoparticles using peels of Punica granatum. Adv Mater Lett 3:376–380. https://doi.org/10.5185/amlett.2012.6357
39. Elia P, Zach R, Hazan S, Kolusheva S, Porat ZE, Zeiri Y (2014) Green synthesis of gold nanoparticles using plant extracts as reducing agents. Int J Nanomedicine 9:4007–4021. https://doi.org/10.2147/IJN.S57343
40. Wang C, Mathiyalagan R, Kim YJ et al (2016) Rapid green synthesis of silver and gold nanoparticles using Dendropanax morbifera leaf extract and their anticancer activities. Int J Nanomedicine 11:3691–3701. https://doi.org/10.2147/IJN.S97181
41. Hatipoğlu A (2021) Rapid green synthesis of gold nanoparticles: synthesis, characterization, and antimicrobial activities. Prog Nutr 23:e2021242. https://doi.org/10.23751/pn.v23i3.11988
42. Gogoi N, Bora U (2011) Green synthesis of gold nanoparticles using Nyctanthes arbortristis flower extract. Bioprocess Biosyst Eng 34:615–619. https://doi.org/10.1007/s00449-010-0510-y
43. Suman TY, Radhika Rajasree SR, Ramkumar R et al (2014) The green synthesis of gold nanoparticles using an aqueous root extract of Morinda citrifolia L. Spectrochim Acta - Part A Mol Biomol Spectrosc 118:11–16. https://doi.org/10.1016/j.saa.2013.08.066
44. Aswathy Aromal S, Philip D (2012) Green synthesis of gold nanoparticles using Trigonella foenum-graecum and its size-dependent catalytic activity. Spectrochim Acta - Part A Mol Biomol Spectrosc 97:1–5. https://doi.org/10.1016/j.saa.2012.05.083
45. Smith GF, Figueiredo E (2017) Kalanchoe fedtschenkoi Raym.-Hamet & H.Perrier (Crassulaceae) is spreading in South Africa's Klein Karoo. Bradleya 35:80–86. https://doi.org/10.25223/brad.n35.2017.a7
46. Richwagen N, Lyles JT, Dale BLF, Quave CL (2019) Antibacterial activity of Kalanchoe mortagei and K. fedtschenkoi against ESKAPE pathogens. Front Pharmacol 10:67. https://doi.org/10.3389/fphar.2019.00067
47. Iosin M, Canpean V, Astilean S (2011) Spectroscopic studies on pH- and thermally induced conformational changes of bovine serum albumin adsorbed onto gold nanoparticles. J Photochem Photobiol A Chem 217:395–401. https://doi.org/10.1016/j.jphotochem.2010.11.012
48. Wangoo N, Suri CR, Shekhawat G (2008) Interaction of gold nanoparticles with protein: a spectroscopic study to monitor protein conformational changes. Appl Phys Lett 92:133104. https://doi.org/10.1063/1.2902302
49. Ding Y, Chen Z, Xie J, Guo R (2008) Comparative studies on adsorption behavior of thionine on gold nanoparticles with different sizes. J Colloid Interface Sci 327:243–250. https://doi.org/10.1016/j.jcis.2008.07.057
50. Roach P, Farrar D, Perry CC (2005) Interpretation of protein adsorption: surface-induced conformational changes. J Am Chem Soc 127:8168–8173. https://doi.org/10.1021/ja042898o
51. Pramanik S, Banerjee P, Sarkar A, Bhattacharya SC (2008) Size-dependent interaction of gold nanoparticles with transport protein: a spectroscopic study. J Lumin 128:1969–1974. https://doi.org/10.1016/j.jlumin.2008.06.008
52. Bisht B, Dey P, Singh AK et al (2022) Spectroscopic investigation on the interaction of direct yellow-27 with protein (BSA). Methods Appl Fluoresc 10:044009. https://doi.org/10.1088/2050-6120/ac8a8b
53. Ariyasu S, Hayashi H, Xing B, Chiba S (2017) Site-specific dual functionalization of cysteine residue in peptides and proteins with 2-azidoacrylates. Bioconjug Chem 28:897–902. https://doi.org/10.1021/acs.bioconjchem.7b00024
54. Chaves OA, Teixeira FSM, Guimarães HA et al (2017) Studies of the interaction between BSA and a plumeran indole alkaloid isolated from the stem bark of aspidosperma cylindrocarpon (Apocynaceae). J Braz Chem Soc 28:1229–1236. https://doi.org/10.21577/0103-5053.20160285
55. Kim B, Song WC, Park SY, Park G (2021) Green synthesis of silver and gold nanoparticles via Sargassum serratifolium extract for catalytic reduction of organic dyes. Catalysts 11:347. https://doi.org/10.3390/catal11030347
56. Mehata MS (2021) Green route synthesis of silver nanoparticles using plants/ginger extracts with enhanced surface plasmon resonance and degradation of textile dye. Mater Sci Eng B Solid-State Mater Adv Technol 273:115418. https://doi.org/10.1016/j.mseb.2021.115418
57. Chai HY, Lam SM, Sin JC (2019) Green synthesis of magnetic Fe-doped ZnO nanoparticles via Hibiscus rosa-sinensis leaf extracts for boosted photocatalytic, antibacterial and antifungal activities. Mater Lett 242:103–106. https://doi.org/10.1016/j.matlet.2019.01.116
58. Muraro PCL, Mortari SR, Vizzotto BS et al (2020) Iron oxide nanocatalyst with titanium and silver nanoparticles: synthesis, characterization and photocatalytic activity on the degradation of Rhodamine B dye. Sci Rep 10:3055. https://doi.org/10.1038/s41598-020-59987-0
59. Nadaf NY, Kanase SS (2019) Biosynthesis of gold nanoparticles by Bacillus marisflavi and its potential in catalytic dye degradation. Arab J Chem 12:4806–4814. https://doi.org/10.1016/j.arabjc.2016.09.020
60. Ruby A, Mehata MS (2022) Surface plasmon resonance allied applications of silver nanoflowers synthesized from Breynia vitis-idaea leaf extract. Dalt Trans 51:2726–2736. https://doi.org/10.1039/D1DT03592D
61. Thanh NTK, Maclean N, Mahiddine S (2014) Mechanisms of nucleation and growth of nanoparticles in solution. Chem Rev 114:7610–7630. https://doi.org/10.1021/cr400544s
62. Rajeshkumar S (2016) Anticancer activity of eco-friendly gold nanoparticles against lung and liver cancer cells. J Genet Eng Biotechnol 14:195–202. https://doi.org/10.1016/j.jgeb.2016.05.007
63. Krishnamurthy S, Esterle A, Sharma NC, Sahi SV (2014) Yucca-derived synthesis of gold nanomaterial and their catalytic potential. Nanoscale Res Lett 9:627. https://doi.org/10.1186/1556-276X-9-627

64. Khalil MMH, Ismail EH, El-Magdoub F (2012) Biosynthesis of Au nanoparticles using olive leaf extract. 1st Nano Updates. Arab J Chem 5:431–437. https://doi.org/10.1016/j.arabjc.2010.11.011
65. Philip D (2010) Green synthesis of gold and silver nanoparticles using Hibiscus rosa sinensis. Phys E Low-Dimens Syst Nanostructures 42:1417–1424. https://doi.org/10.1016/j.physe.2009.11.081
66. Holm VRA, Greve MM, Holst B (2016) Temperature induced color change in gold nanoparticle arrays : investigating the annealing effect on the localized surface plasmon resonance. J Vac Sci Technol B 501:06K501. https://doi.org/10.1116/1.4963153
67. Dutta A, Chattopadhyay A (2016) RSC Advances The effect of temperature on the aggregation kinetics of partially bare gold nanoparticles. RSC Adv 6:82138–82149. https://doi.org/10.1039/c6ra17561a
68. Yi S, Xia L, Lenaghan SC et al (2013) Bio-synthesis of gold nanoparticles using English ivy (Hedera helix). J Nanosci Nanotechnol 13:1649–1659. https://doi.org/10.1166/jnn.2013.7183
69. Sankhla A, Sharma R, Singh R, Kashyap D (2016) Biosynthesis and characterization of cadmium sulfi de nanoparticles e An emphasis of zeta potential behavior due to capping. Mater Chem Phys 170:44–51. https://doi.org/10.1016/j.matchemphys.2015.12.017
70. Aryan R, Mehata MS (2021) Green synthesis of silver nanoparticles using Kalanchoe pinnata leaves (life plant) and their antibacterial and photocatalytic activities. Chem Phys Lett 778:138760. https://doi.org/10.1016/j.cplett.2021.138760
71. Shabestarian H, Homayouni-Tabrizi M, Soltani M et al (2017) Green synthesis of gold nanoparticles using sumac aqueous extract and their antioxidant activity. Mater Res 20:264–270. https://doi.org/10.1590/1980-5373-MR-2015-0694
72. Akabri B, Tavandashti MP, Zandrahimi M (2011) Particle size characterization of nanoparticles- A practical approach. Iranian J Mater Sci Eng 8:48–56. http://ijmse.iust.ac.ir/article-1-341-en.html. Accessed 20 Jan 2023
73. Sperling RA, Gil PR, Zhang F et al (2008) Biological applications of gold nanoparticles. Chem Soc Rev 37:1896–1908. https://doi.org/10.1039/b712170a
74. Huang HC, Lin MK, Yang HL et al (2013) Cardenolides and bufadienolide glycosides from Kalanchoe tubiflora and evaluation of cytotoxicity. Planta Med 79:1362–1369. https://doi.org/10.1055/s-0033-1350646
75. Ahmad T, Irfan M, Bhattacharjee S (2016) Parametric study on gold nanoparticle synthesis using aqueous Elaise guineensis (oil palm) leaf extract : effect of precursor concentration. Procedia Eng 148:1396–1401. https://doi.org/10.1016/j.proeng.2016.06.558
76. Khademi-Azandehi P, Moghaddam J (2014) Green synthesis, characterization and physiological stability of gold nanoparticles from Stachys lavandulifolia Vahl extract. Particuology 19:22–26. https://doi.org/10.1016/j.partic.2014.04.007
77. Folorunso A, Akintelu S, Oyebamiji AK et al (2019) Biosynthesis, characterization and antimicrobial activity of gold nanoparticles from leaf extracts of Annona muricata. J Nanostructure Chem 9:111–117. https://doi.org/10.1007/s40097-019-0301-1
78. Bolaños K, Kogan MJ, Araya E (2019) Capping gold nanoparticles with albumin to improve their biomedical properties. Int J Nanomedicine 14:6387–6406. https://doi.org/10.2147/IJN.S210992
79. Awotunde O, Okyem S, Chikoti R, Driskell JD (2020) Role of free thiol on protein adsorption to gold nanoparticles. Langmuir 36:9241–9249. https://doi.org/10.1021/acs.langmuir.0c01550
80. Alsamamra H, Hawwarin I, Sharkh SA, Abuteir M (2018) Study the interaction between gold nanoparticles and bovine serum albumin: spectroscopic approach. J Bioanal Biomed 10:43–49. https://doi.org/10.4172/1948-593x.1000203
81. Roy S, Das TK (2016) Interaction of biosynthesized gold nanoparticles with BSA and CTDNA: a multi-spectroscopic approach. Polyhedron 115:111–118. https://doi.org/10.1016/j.poly.2016.05.002
82. Singh G, Priyanka SA et al (2021) Schiff base-functionalized silatrane-based receptor as a potential chemo-sensor for the detection of Al^{3+} ions. New J Chem 45:7850–7859. https://doi.org/10.1039/d1nj00943e
83. Samari F, Hemmateenejad B, Shamsipur M et al (2012) Affinity of two novel five-coordinated anticancer Pt(II) complexes to human and bovine serum albumins: a spectroscopic approach. Inorg Chem 51:3454–3464. https://doi.org/10.1021/ic202141g
84. Brewer SH, Glomm WR, Johnson MC et al (2005) Probing BSA binding to citrate-coated gold nanoparticles and surfaces. Langmuir 21:9303–9307. https://doi.org/10.1021/la050588t
85. Aneesha ON, Mehata MS (2023) In situ synthesis of WS_2 QDs for sensing of H_2O_2: quenching and recovery of absorption and photoluminescence. Mater Today Commun 34:105013. https://doi.org/10.1016/j.mtcomm.2022.105013
86. Sun T, Su Y, Sun M, Lv Y (2021) Homologous chemiluminescence resonance energy transfer on the interface of WS_2 quantum dots for monitoring photocatalytic H_2O_2 evaluation. Microchem J 168:106344. https://doi.org/10.1016/j.microc.2021.106344
87. Shukla S, Masih A, Aryan MMS (2022) Catalytic activity of silver nanoparticles synthesized using Crinum asiaticum (Sudarshan) leaf extract. Mater Today Proc 56:3714–3720. https://doi.org/10.1016/j.matpr.2021.12.468
88. Saikia P, Miah AT, Das PP (2017) Highly efficient catalytic reductive degradation of various organic dyes by Au/CeO_2-TiO_2 nano-hybrid. J Chem Sci 129:81–93. https://doi.org/10.1007/s12039-016-1203-0
89. Rabeea MA, Owaid MN, Aziz AA et al (2020) Mycosynthesis of gold nanoparticles using the extract of Flammulina velutipes, Physalacriaceae, and their efficacy for decolorization of methylene blue. J Environ Chem Eng 8:103841. https://doi.org/10.1016/j.jece.2020.103841
90. Das J, Velusamy P (2014) Catalytic reduction of methylene blue using biogenic gold nanoparticles from Sesbania grandiflora L. J Taiwan Inst Chem Eng 45:2280–2285. https://doi.org/10.1016/j.jtice.2014.04.005
91. Suvith VS, Philip D (2014) Catalytic degradation of methylene blue using biosynthesized gold and silver nanoparticles. Spectrochim Acta - Part A Mol Biomol Spectrosc 118:526–532. https://doi.org/10.1016/j.saa.2013.09.016

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

A Truncated SVD Framework for Online Hate Speech Detection on the ETHOS Dataset

Anusha Chhabra
Department of Information Technology
Biometric Research Laboratory
Delhi Technological University
Delhi-110042, India
anusha.chhabra@gmail.com

Dinesh Kumar Vishwakarma
Department of Information Technology
Biometric Research Laboratory
Delhi Technological University
Delhi-110042, India
dinesh@dtu.ac.in

2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT) | 978-1-6654-9414-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICITIIT57246.2023.10068574

***Abstract*—Hate content on social media is currently one of the most significant risks, where the victim is either a single individual or a group of people. In the current scenario, online web platforms are one of the most prominent ways to contribute to an individual's opinions and thoughts. Free sharing of ideas on an event or situation also bulks on the web. Information sharing is sometimes a bane for society if primarily used platforms are utilized with some lousy intention to spread hatred for intentionally creating chaos/ confusion among the public. Users take this as an opportunity to spread hate to get some monetary benefits, the detection of which is of paramount importance. This article utilizes the concept of truncated singular value decomposition (SVD) for detecting hate content on the ETHOS (Binary-Label) dataset. Compared with the baseline results, our framework has performed better in various machine learning algorithms like SVM, Logistic Regression, XGBoost, and Random Forest.**

Keywords—Hate Speech, Machine Learning, SVD, Binary-label Classification, TF-IDF

I. Introduction

There has been substantial usage of social media platforms by more people and exponential growth in the data. People share their thoughts and views on almost everything without considering the impact on society. According to statistics, Twitter is the most usable platform having nearly 340 million active users [1] and about 200 million tweets per year. The mentioned statistics and many users are also flooding hate content. Therefore, identifying hate content is a very prominent research area. Hate content can be defined as controversial, attacking group characteristics based on religion, gender, ethnicity, etc. Fig *1* shows that a leader is porting a divisive statement targeting those who raise their voices against CAA, NRC, and NPR [2]. Perhaps, Major social media platforms are curbing hate content at an initial stage. Still, hate content is sowing its roots almost in every form of content characteristics.

To improve the binary classification of social media texts, researchers and practitioners are paying more attention to the upcoming techniques of machine learning and deep learning. Considerable efforts have been spent on creating new and practical features that better classify hate speech on social media [3], [4], [5]. In addition, the challenges related to specific hate content detection lie in the need for more guidelines, benchmarks [6]**,** and the non-availability of multimodal datasets. This paper presents a framework for identifying hate content on the ETHOS dataset.

The Major contributions of this manuscript are:

- Training the models on one dataset and cross-validation is done on another dataset which is approximately 24 times greater.
- To show the vulnerability of a small dataset with another related large dataset.

The rest of the paper is organized as follows: Section II provides an overview of the recent works on hate content detection using unsupervised machine learning approaches.

Section III illustrates the framework to detect hate content, followed by the discussion of the experimental results in Section IV. The conclusion and further scope are discussed in Section V.

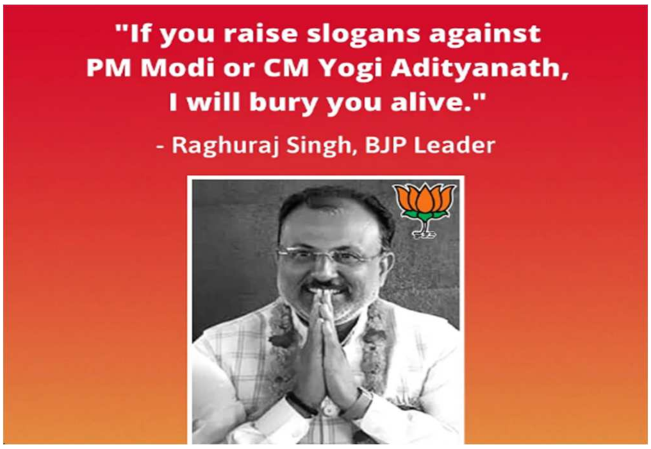

"If you raise slogans against
PM Modi or CM Yogi Adityanath,
I will bury you alive."
- Raghuraj Singh, BJP Leader

Fig 1. Example of Hate Speech

II. Related Work

Identifying hate content is crucial for millions of users to have freedom of expression. Authors and Academicians are focusing on multimodal, multilingual, and multiclass hate speech detection using supervised, unsupervised and semi-supervised machine learning techniques.

Machine learning has played its role very well in the last two decades. Specifically for hate speech and offensive language detection, Naïve Bayes, Support Vector Machine, Logistic Regression, Random Forest Decision Tree, and ensemble techniques are used as machine learning classifiers[7] for hate speech detection. The work in this area is found to be done in various languages. The same probabilistic and predictive analysis techniques are used by [8] for hate speech detection in Indonesian languages. [9] applies a supervised SVM technique for racist text classification. It is also observed that by ignoring the word-order sequence, BoW showed better accuracy in text classification. To overcome the limitation of BoW, Researchers perform N-gram approaches[10]. Manual labeling of large data is a time-consuming task that leads to the requirement of an unsupervised method. It takes advantage of detecting hate speech in a huge stream of data. Authors in [11] used Kohonen maps for the detection of cyberbullying, claiming an accuracy of 72%. PCA is also another class to

978-1-6654-9414-4/23/$31.00 ©2023 IEEE

detect violent or non-violent tweets. The authors examined the classification rate using various machine-learning algorithms after analyzing the sentence and language features[13].

The benefit of manual labeling is that it is effective for domain-specific tasks, but its drawback is execution time. In order to train and test the classifiers, the authors employed unsupervised approach for dimensionality reduction [12], applying k-means clustering and assigning each cluster as one human-annotated data from the Twitter dataset using supervised machine learning text classifiers. Twitter data is categorized into hateful and antagonistic labels using a Bayesian logistic regression model, according to [7]. The effectiveness of several supervised approaches for hate speech identification is evaluated and contrasted by authors [6] with an emphasis on South Asian languages. labeled and unlabeled data can be used by semi-supervised learning systems. Labeling data pertaining to unlabeled data can effectively increase productivity. [14] analyzed that supervised learning can sufficiently catch small-scale events whereas unsupervised learning has limited ability to deal with limited-scale events; yet, the necessity to manually label the data lowers down the models scalability.

III. Dataset Description

ETHOS dataset contains ~1K comments from YouTube and Reddit validated through a Figure-Eight Platform. This dataset repository has two csv files: Binary and Multi- Label. Out of 998 comments in binary file, 565 are manually labeled as non-hate while rest of them are labeled as hate. For Multi-label file, 8 labels are created as violence (if it incites (1) or not (0) violence), directed vs general (if it is directed to a person (1) or a group (0)), and 6 labels about the category of hate speech like gender, race, national origin, disability, religion and sexual orientation.

IV. Proposed Methodology

Hate speech is now a threat to society, affecting the dignity of an individual, unity, and nation. Many hate words are used alternatively. Fig 2 shows the word cloud for hate speech explicitly generated from [3]. Therefore, Eliminating and classifying hate content over social platforms is crucial and requires an hour.

Data preprocessing is a component of data preparation. Several techniques are used to normalize the data before it is fed into any machine learning or AI development pipeline. Fig 3 represents the broader approach used in the classification. Tokenization is the initial step in any NLP pipeline, used to break unstructured data into chunks of discrete values. Then, stemming is used as a normalized technique in which tokenized words are converted into short words to remove redundancies. Finally, the cleaned data is used to create a dictionary for key: value pairs. TF-IDF is used for mapping words to vectors of real numbers then a vector matrix is given as an input to classify as hate or non-hate.

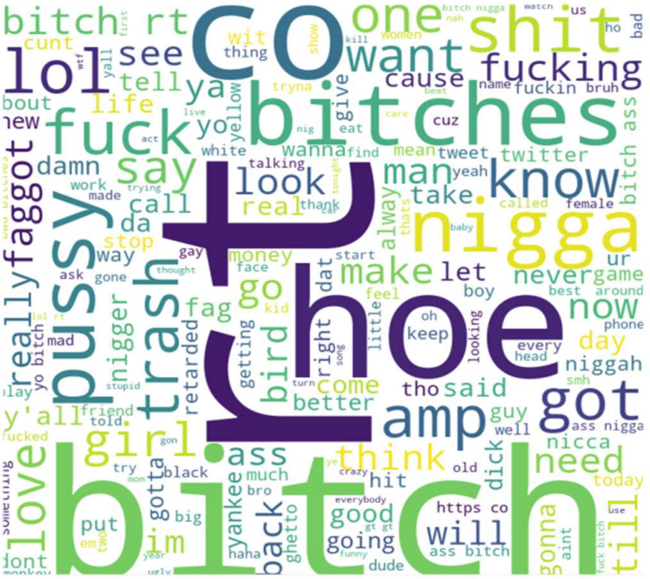


Fig 2. Word Cloud

Error! Reference source not found. represents the flowchart adopted for implementation. In the process flow, tweets are preprocessed in the first stage. Second stage implements TF-IDF, to quantify the words. Truncated SVD is used for dimensionality reduction for simplifying the calculations. Hyper parameter tuning such as L1 regularization is also done for logistic regression, XGBoost and SVM. Finally, the Prediction is done using various machine learning algorithms like Logistic Regression, Random Forest, Support Vector Machines, and XGBoost.

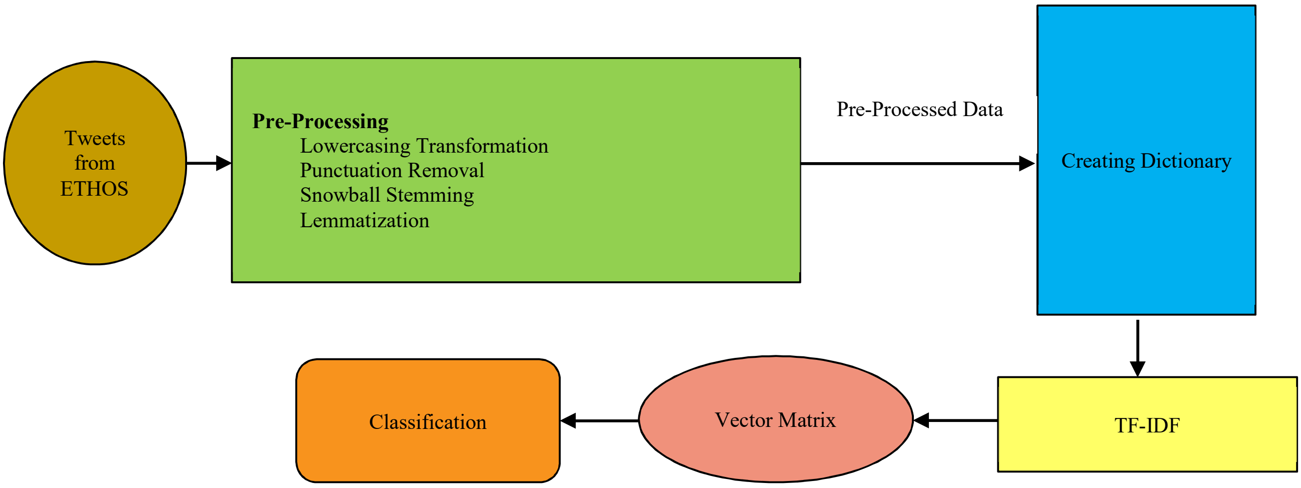


Fig 3. Data Pre Processing Steps

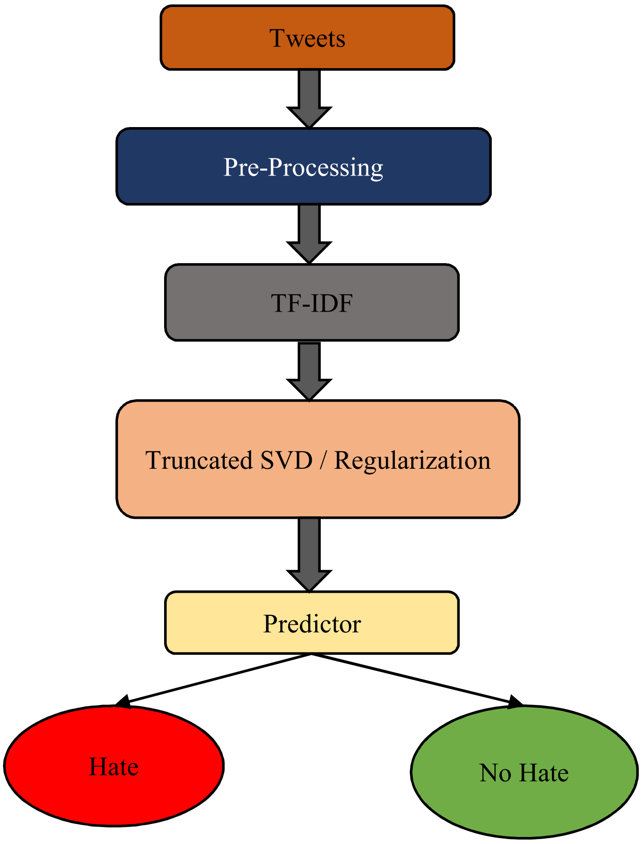


Fig 4. Process Flow

V. Experimental Setup

Although the dataset size is very small. To prove that a dataset of higher quality is more useful than the larger datasets, we have considered a dataset D1[3] which is approximately 24 times greater than ETHOS. In this experiment, we train various machine learning models with default parameters on the ETHOS dataset and compare the results with D1 dataset. The results are compared in terms of F1 score and balanced accuracy.

F1 score (Eq 1) is defined as the combination of precision and recall of a classifier into a single metric by considering their harmonic mean.

$$F1 = \frac{2(Precision*Recall)}{Precision+Recall} \qquad (1)$$

Table I F1 Scores of ETHOS and D1 from SVM, LR, RF, XGBoost

Models	**ETHOS**			**D1**		
	F1 Score	**F1 Score (Hate)**	**F1 Score (No Hate)**	**F1 Score**	**F1 Score (Hate)**	**F1 Score (No Hate)**
SVM	67.71	59.60	73.63	75.47	12.86	79.30
LR	69.13	60.84	75.27	78.76	14.89	82.67
RF	67.01	58.85	73.03	67.21	12.73	70.55
XGBoost	65.30	54.50	73.44	75.39	10.62	79.35

Table I shows the results in the form of overall F1 scores, F1 Score (Hate) and F1 score (No Hate) of four machine learning models implemented on ETHOS and D1 datasets. The results are obtained when the models are trained on ETHOS and cross validation is done on D1 dataset.

Balanced accuracy (Eq 2) is defined as the arithmetic mean of sensitivity and specificity. It is also considered as the further development in standard accuracy metric.

$$Balanced\ Accuracy = \frac{Specificity+Sensitivity}{2} \qquad (2)$$

Balanced Accuracies are shown in the Table II representing that our proposed approach using truncated SVD and hyper parameter tuning gives better results than baseline results. The graphical representation of balanced accuracies are shown in Fig 4.

Table II Comparison Table_Balanced Accuracy

Balanced Accuracy		
Models	**ETHOS_Our Approach**	**ETHOS_Baseline**
SVM	***66.70***	66.43
LR	***67.07***	66.94
RF	***68.17***	65.04
XGBoost	***64.42***	64.33

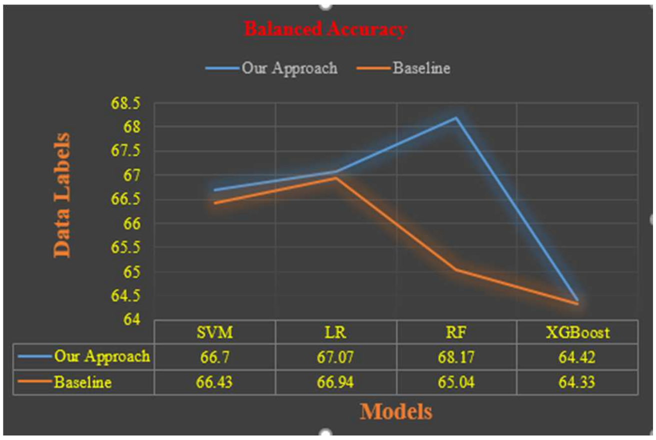


Fig 4. Comparison Graph of Balanced Accuracy: ETHOS_BINARY (Our Approach vs Baseline)

VI. Conclusion & Future Scope

From the empirical evaluation done in the paper, it is seen that reducing features using Truncated SVD along with hyper parameter tuning helped in increasing balanced accuracy and F1 score for algorithms like Logistic Regression, SVM and XGBoost when compared to the baseline results. For Random Forest, only change in hyper parameter is giving good results. The paper covers the basic ML algorithms for detecting hate speech. So, more SOTA algorithms and ensemble techniques can be implemented as a future task. Moreover, ETHOS dataset can be combined with other similar datasets for more evaluations.

References

[1] J clement, "https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/," *Number of monthly active Twitter users worldwide*, 2021. .

[2] M. Bose, "https://www.thequint.com/news/politics/senior-bjp-leaders-giving-india-a-free-tutorial-in-hate-speech#read-more," *The Quint*, 2020. .

[3] T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated hate speech detection and the problem of offensive language,"

Proc. 11th Int. Conf. Web Soc. Media, ICWSM 2017, pp. 512–515, 2017.

[4] Z. Waseem and D. Hovy, "Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter," pp. 88–93, 2016, doi: 10.18653/v1/n16-2013.

[5] P. Burnap and M. L. Williams, "Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making," *Policy and Internet*, vol. 7, no. 2, pp. 223–242, 2015, doi: 10.1002/poi3.85.

[6] M. M. Khan, K. Shahzad, and M. K. Malik, "Hate Speech Detection in Roman Urdu," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 20, no. 1, pp. 1–19, 2021, doi: 10.1145/3414524.

[7] P. Burnap and M. Williams, "Hate Speech, Machine Classification and Statistical Modelling of Information Flows on Twitter: Interpretation and Communication for Policy Decision Making," in *Internet, Policy & Politics*, 2014, pp. 1–18, [Online]. Available: http://orca.cf.ac.uk/id/eprint/65227%0A.

[8] I. Alfina, R. Mulia, M. I. Fanany, and Y. Ekanata, "Hate speech detection in the Indonesian language: A dataset and preliminary study," in *2017 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2017*, 2018, vol. 2018-Janua, pp. 233–237, doi: 10.1109/ICACSIS.2017.8355039.

[9] E. Greevy and A. F. Smeaton, "Classifying racist texts using a support vector machine," *Proc. Sheff. SIGIR - Twenty-Seventh Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, no. January 2004, pp. 468–469, 2004, doi: 10.1145/1008992.1009074.

[10] W. B. Cavnar, J. M. Trenkle, and A. A. Mi, "N-Gram-Based Text Categorization," *Proc. SDAIR-94, 3rd Annu. Symp. Doc. Anal. Inf. Retr.*, pp. 161–175, 1994, [Online]. Available: http://www.let.rug.nl/~vannoord/TextCat/textcat.pdf.

[11] M. Di Capua, E. Di Nardo, and A. Petrosino, "Unsupervised cyber bullying detection in social networks," *Proc. - Int. Conf. Pattern Recognit.*, vol. 0, pp. 432–437, 2016, doi: 10.1109/ICPR.2016.7899672.

[12] K. E Abdelfatah, G. Terejanu, and A. A Alhelbawy, "Unsupervised Detection of Violent Content in Arabic Social Media," pp. 01–07, 2017, doi: 10.5121/csit.2017.70401.

[13] Y. Chen, Y. Zhou, S. Zhu, and H. Xu, "Detecting offensive language in social media to protect adolescent online safety," 2012.

[14] V. Tech, C. Lu, H. J. H. I. I. I. College, and F. Chen, "STED : Semi-Supervised Targeted-Interest Event Detection."

International Journal of Advanced Technology and Engineering Exploration, Vol 10(98)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2021.874950

Research Article

Analysis of hardness for dissimilar stainless-steel joint by mathematical modelling

Deeksha Narwariya[1*] and Aditya Kumar Rathi[2]
Research Scholar, Department of Mechanical Engineering, MPAE Division, Netaji Subhas University of Technology, India[1]
Associate Professor, Department of Mechanical Engineering, MPAE Division, Netaji Subhas University of Technology[2]

Received: 15-July-2021; Revised: 14-January-2023; Accepted: 17-January-2023

Abstract

Gas tungsten arc welding (GTAW) is very popular globally as it is capable of doing similar and dissimilar material welding. Most commonly steel and aluminium of different grades are joined using this process. In this research work, two unalike grades of stainless steel i.e., 304 and 316 were welded together with different combination of parameters. An attempt was made to find out the effect on the surface hardness of the joint. The parameters under consideration were current, welding speed and torch angle, having two limits, maximum limit (+1) and lower limit (-1). A mathematical model was developed between input parameters and the responses such as hardness was analyzed by using the factorial approach. The result of the analysis shows that increase in current and welding speed decreases the hardness whereas an increase in torch angle increases the surface hardness. Hardness is maximum at the weld zone of two dissimilar metals and minimum at the heat affected zone.

Keywords

Stainless steels, GTAW, Surface hardness, Input parameters, ANOVA.

1.Introduction

There are certain concerns when we need that a joint must be of different metal compositions. This also applies to higher temperature applications or of wear situations. In such situations and other related problems, one has to think of dissimilar metal joints. Two different grades of stainless steel with different combination of the input parameter can be joined together. The combination of the different parameter leads to obtain desired results. These input parameters can be current, table speed and torch angle, variation in these parameters can cause alteration in mechanical properties. These changes are also responsible for the change in stability of arc, melting and deposition rate [1]. For welding dissimilar metals, the following four things we have to keep in mind, melting point of metals, coefficients of their thermal expansion, electrochemical difference and solubility of each metal [2]. If fusion welding is used, melting point plays an important role. Similarly, the coefficient of thermal expansion can create excessive strain in the weld zone if there is more difference in it. Electrochemical difference can relate to corrosion in the inter metallic zone [3].

Metals that fit closer to electrochemical scale can provide simple welding process than those that are far apart. Solubility of each metal is another problem which is to be taken care [4]. If these are not interpolable with each other than third metal is used as a filler metal which is soluble with these metals [5]. The process available for joining dissimilar metals are fusion and nonfusion welding along with low dilution welding. The last two methods are used for high output and special welding application [6]. Most commonly used process for dissimilar metal welding (DMW) in power and process industries are fusion welding [4].

The purpose of the paper was to study the surface hardness of welded joints. The combination of these materials will make a joint very useful for certain applications because these can sustain the extreme

*Author for correspondence

environmental condition for corrosion. These are easily weldable at the given thickness so the problem of solubility and melting point can be dealt easily with such type of dissimilar fusion welding. Some of the important applications are heat exchangers, pipelines, pressure vessels, flanges and fittings. Effect of welding parameters on the hardness an be easily seen in this work.

This paper is organized and explored in the following sections. Literature discussion in section 2. Methods have been explored in section 3. Results investigation and the discussion has been presented in section 4. The concluding remark has been elaborated in section 5.

2.Literature review

The metallurgical and mechanical characterization of dissimilar joints between low carbon steel and stainless steel was made by laser autogenous welding. The results demonstrate that the affirmative difference in yield between the weld metal and the base materials assist the joint from being plastically deformed [7]. Another researcher presented in his paper that deeper qualitative analysis of the welding procedure's influences the fracture toughness of the High strength low alloy (HSLA) steel welded by manual metal arc (MMA) and metal active gas (MAG) process [8]. High temperature during welding has an affirmative outcome on the micro-structure and it was observed that material failure occurred in the base material near the heat affected zone (HAZ) while manufacturing American iron and steel institute (AISI) 316L stainless steel tubes with Selective laser melting (SLM) technology [9]. Hydrogen analysis and scanning electron microscopy (SEM) were conducted to find the effect of charging time on the hydrogen concentration and surface Morphology of hydrogen exposed 316L stainless steel welded joints at ambient and cryogenic temperature. The result shows that vulnerability to hydrogen decreased the absorbed energy and ductility of 316L stainless steel at all tested temperatures but not much difference was found among the pre-charging times [10]. Gas tungsten arc welding (GTAW) can be more widely used in aerospace components and food processing units by welding sheets and pipes [11]. Stainless steel 316 is mostly used in Thermal power plants, its study for temperature time precipitation (TTP), when it is used in high temperature applications [12]. Underwater wet welding of 317L stainless steel with nickel based tubular wire shows that fully austenitic weld metal without cracks and pores with nickel-based filler wire can be achieved as observed in the micro-structure analysis [13]. Dissimilar welding was done on small thick plates with different fluxes on Tungsten inert gas (TIG) to micro-structure examination with the best outcome in case of TiO_2 ingredients [14]. Similarly, a dissimilar welding was performed on unlike grades of steel where related parameters were optimized by Taguchi analysis on a tensile test specimen [15]. A dissimilar weldment between P91 and American iron and steel institute (AISI) 316L austenitic steel fabricated by activated TIG welding was performed on 8 mm thick plates using a single pass with pre-coated mixture of metallic oxides the result shows that elimination of hot cracking in the joint compared with conventional methods is very helpful [16]. The mechanical and micro-structural properties can be enhanced by using filler material such as 309L when joining similar grades of stainless steel by TIG welding, the best result is obtained by 120 A current [17]. Influence of welding parameters such as welding current and welding speed was carried out to find the strength of joint welded by TIG welding in between low carbon steel and aluminium alloy AA1050 the result shows that at an optimized value of strength with the help of Taguchi analysis can be achieved [18]. SS 304 and carbon steel welded together with TIG and A-TIG process and it was found that A-TIG welded joint has better joint efficiency and mechanical properties as compared to TIG welded joints [19]. Optimization of process parameters was done on naval steel to get the maximum depth of penetration by using GTAW process. Response surface methodology (RSM) D-optimal method and Taguchi optimization techniques were compared, and it showed that RSM result is better than Taguchi analysis for better penetration [20]. Artificial neural network (ANN) was used for simulation, and genetic algorithm (GA) was used to optimize the process parameters such as welding current, welding speed, nozzle deflection distance, travel angle and wire feed frequency for GTAW process on AISI 1020 steel blank material with two tubular wires [21]. A TIG welding process was used to study the effect of welding speed and welding current on welding of mild steel plates it was concluded that current can be increased up to 110 A for 6 mm thick plates to get the best hardness result. A further increase in current will reduce the hardness [22]. Shielded metal arc welding (SMAW) was performed on A 36 carbon steel welded joint to study the effect of welding current on the micro structure and hardness of the joint. Results indicate that the higher current reduces strength and hardness [23]. A dissimilar joint of AA 6063 with AA 7075 was made with the help of friction stir welding (FSW) process to study the effect of spindle speed and welding speed, the result shows that increasing the spindle speed and decreasing the welding speed gave the best result of tensile strength [24]. A study of metal inert gas (MIG) welding process carried out on carbon steel plates says that the hardness of the joint increases with the decrease in welding speed and increase in welding current [25]. The effect of gas flow rate on welded stainless-steel alloy SS-202 joint was studied and the result shows that at a low gas flow rate high tensile strength and low ductility were achieved whereas at high welding speeds high tendency of welding defects and improper penetration takes place [26].

As per the latest research, software simulation is also useful to study the dissimilar metal welded plates of 20/0Cr18Ni9 by finite element analysis the result shows that residual stress of each model is significantly reduced. The stress reduction of 20 steel side is significantly larger than on the 0Cr18Ni9 side and deformation was smaller in each case after the heat treatment [27]. Dissimilar metal joining investigation of AISI 316L-Alloy800 by activated TIG. (A-TIG) shows that joint has a decent combination of tensile strength and impact toughness [28]. The propeller shaft made by joining steel ST41 and 316L stainless steel with a variety of electrode in SMAW process the result of the analysis shows that the tensile strength of the base material increases and use of E6013 electrode has lower corrosion resistance [29]. Micro-structure and corrosion behavior of copper and 316L stainless steel welded joints on an electrolytic copper cathode plate was investigated. Clear phase separation was noted in the weld metal with the γ- and ε-Cu phase dominating the Fe-rich and Cu-rich zones, respectively [30]. A dissimilar joint of aluminium alloy and commercially used copper alloy by friction stir spot welding (FSSW) process shows that the optimized result of welding parameters can be obtained by the weighted aggregated sum product assessment coupled with grey wolf optimization method and pin length is the most significant factor [31]. Mechanical properties and crack propagation behavior was investigated of pressurized water reactor for cladding layer material 304L and SA508 the result indicated that the strength value at the fusion boundary is largest and yield strength reaches at 689 Mpa [32]. As per the literature review, one can say that a lot of work already done on the dissimilar weld joint but to the best knowledge, dissimilar TIG joint of SS 304 and SS 316 of 2 mm is still remains. So, an effort was made on dissimilar grades of SS 304 and SS 316 steel on the TIG welding by considering the combination of different level of input parameters.

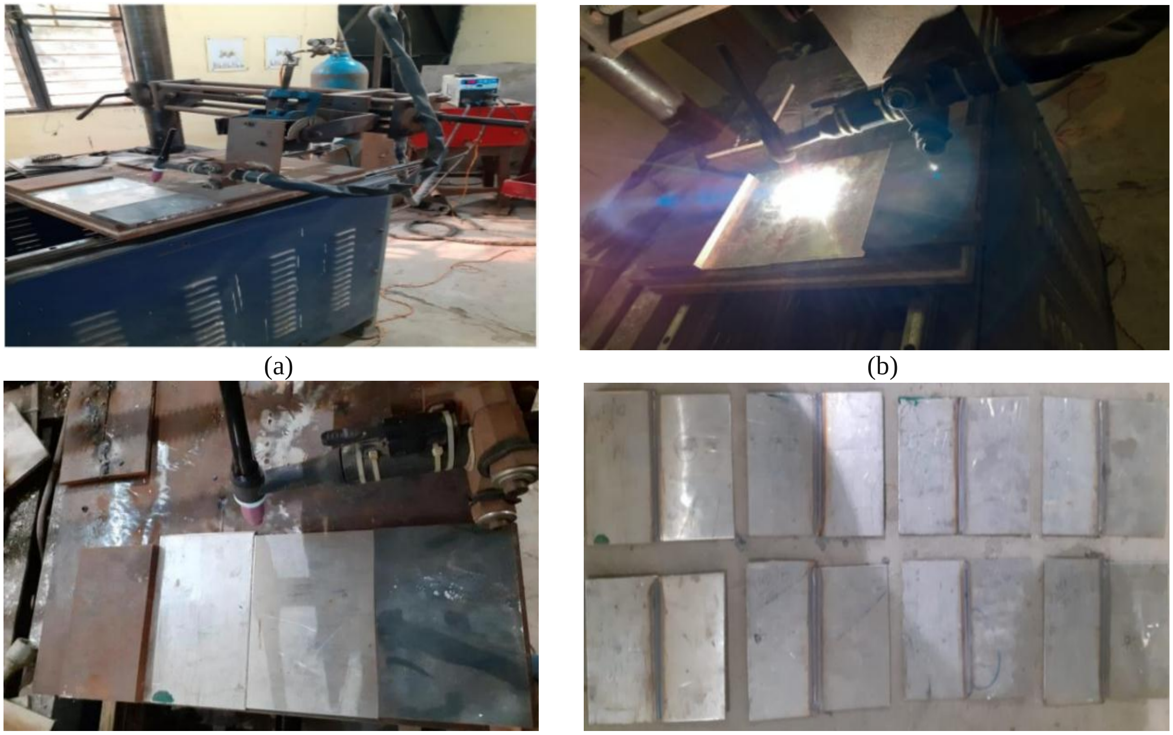
3.Methods

The following path was followed while carrying out the fact-finding work.

1. Selection of material
2. Problem with welding of stainless steel
3. Recognizing the input parameters
4. Establishment of a design matrix
5. Experimentation for the responses
6. Generation of the mathematical model
7. Analysing the graphs and plots

3.1Experimental setup

In the present study, TIG welding machine of flat V-I (voltage and current) characteristics was used. For sample preparation shielded gas as argon was used to weld two different grades of steel plates with non consumable electrode of 2.5 mm diameter. The sample dimensions were kept as 150 mm square and 2 mm thick. The Rockwell hardness test is the most widely used it has 3 type of scales A, B, C. B scale was used in this case according to the data book as maximum hardness of SS 304 is 70 Rockwell hardness at B scale (HRB) and SS 316 is 78 HRB. Experimental setup is shown in *Figure 1(a)*, consist of a welding torch and the platform to weld the specimen. Welding of dissimilar steel 2 mm plates in process is shown *Figure 1(b)*. Fixing of plates to make a butt joint is shown in *Figure 1(c)* before the welding. Plates are butted against each other and a clamping device will hold the plates together. The joint after the welding operation is shown in *Figure 1(d)* before making the specimen for other testing processes.



(a) (b)

(c) (d)

Figure 1 (a) TIG welding setup (b) welding of specimen (c) SS304 and SS316 specimen (d) welded specimen

3.2Selection of material

Stainless steels are a ferrous alloy that contains chromium which protects the ferrous content from rusting and also enhance heat resistant property. Stainless steel being resistant to corrosion, can be used in many applications. The important point is that it can be easily shaped into plates, sheets, bars and tubes. Important applications are heavy industries, construction material and surgical equipments. It also does not require any surface treatment mostly used in kitchen appliances and food processing units. So, AISI 304 and AISI 316 were selected for the study purpose. The chemical composition of these two have been discussed.

AISI 304 which is Chromium-Nickel based steel, mostly used as 300 series stainless steel, it has superior forming and welding attributes. The cautiously controlled investigation of 304 made it to be deep drawn more seriously than types 301 and 302 without in-between heat demulcent. AISI 304 also has prominent welding attributes. In a gently corrosive environment, no post-weld annealing required to regenerate the superior performance of this grade. Chemical composition of SS304 is given in *Table 1*.

Table 1 Composition of SS304

S. No.	Elements	Weight (%)
1	Carbon	0.060
2	Manganese	0.86
3	Silicon	0.031
4	Chromium	18.35
5	Nickel	8.20
6	Phosphorus	0.031
7	Sulphur	0.01

By adding 2.5% molybdenum in 316/316L an austenitic steel increased ace corrosion resistance of this type 304 grade. 316L has reinforced indentation corrosion opposition and has superior mechanical phenomenon to sulphates, phosphates and others. 316/316L has better resistance than standard 18/8 types of salt water, acids, chlorides, bromides and iodides. Its chemical mixture is given in *Table 2.*

Table 2 Components of SS316

S. No.	Elements	Weight (%)
1	Carbon	0.04
2	Manganese	1.6
3	Silicon	0.05
4	Chromium	19
5	Nickel	12.5
6	Phosphorus	0.03
7	Sulphur	0.03

3.3Problem in TIG welding of stainless steel

Steel is mainly used because of its anti-corrosion property. Chromium and carbon adversely effected in the form of corrosion called weld decay or inter-crystalline corrosion. During welding nearby carbon dissolves with chromium and then on cooling precipitate as chromium carbide on the grain boundary hence, it is going to lose its corrosion resistance property. There are so many techniques. The most obvious are either reduce carbon content or add some alloying element which can react and avoid the formation of chromium carbide. Higher expansion in the heat affected zone results in increased thermal stresses hence more distortion. For welding of thin sheet where more dimensional tolerance is required, which can be achieved by use of accelerated cooling process like copper chills or freezing gas. That makes processing more tedious and costly. TIG welding of stainless steel is carried out within the inert gas atmosphere otherwise chromium present in the ferrous alloy reacts and form a compound. Generally, argon gas is used and sometimes nitrogen can also be used, but there is the risk of hot cracking in the case of nitrogen. Stress corrosion cracking in fabrication industries, is the other problem. It may occur due to halides or strong alkali solution. Cracking occurs at high stress region. Stresses in the region near by weld reaches to yield point of metal and cracking propagate rapidly. Stress relieving process at the temperature 700 to 900˚C has to be used to avoid this failure.

3.4Important process factor

On the ground of literature survey, test runs were performed, it showed that weld properties are related to current, welding speed and torch angle very much. Where as no significant effect was seen due to gas flow rate. Keeping in mind it was kept constant 15 l/min throughout the experiment. Once the important input variable was found, their working extent was ascertained by performing trial runs. Such as, increasing one parameter over its range and keeping others as constant by checking the quality of weld in reachable range as far as weld bead and any surface fault is concerned. All the process parameters are given in *Table 3.*

3.5Formation of a design matrix

Using MINITAB software, the design formula was prepared. Two levels of welding current were chosen as 120 and 140 Amp. Welding speed levels was chosen as 25 and 35 cm/min. Torch angel was varied from 0-45˚. Further welding was performed according to the design matrix, whereas weld run was performed in different combination of input parameters. The design matrix in coded form is shown in *Table 4.* +1 indicates higher level and -1 indicates lower level. Process flow diagram is shown in *Figure 2*.

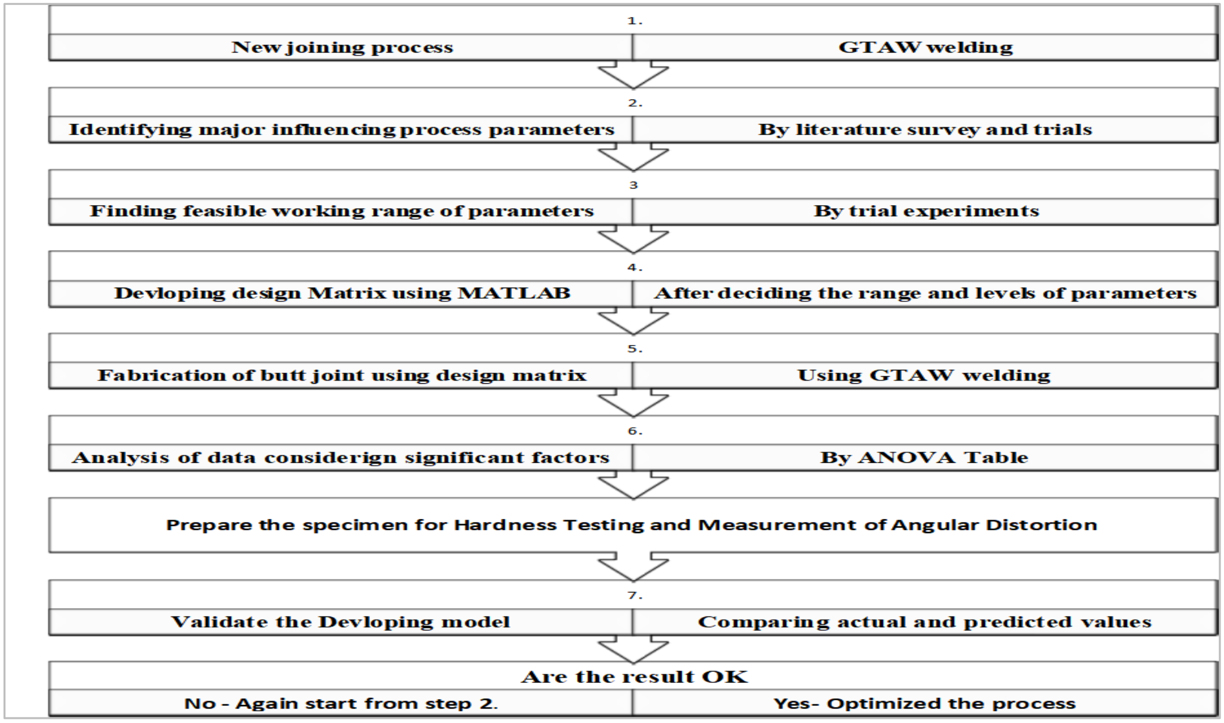

1.
New joining process
GTAW welding
2.
Identifying major influencing process parameters
By literature survey and trials
3
Finding feasible working range of parameters
By trial experiments
4.
Devloping design Matrix using MATLAB
After deciding the range and levels of parameters
5.
Fabrication of butt joint using design matrix
Using GTAW welding
6.
Analysis of data considerign significant factors
By ANOVA Table
Prepare the specimen for Hardness Testing and Measurement of Angular Distortion
7.
Validate the Devloping model
Comparing actual and predicted values
Are the result OK
No - Again start from step 2.
Yes- Optimized the process

Figure 2 Process flow diagram

Table 3 Important factors

S. No.	Parameters	Parameter limits
1	Welding current	120 - 140 Amp
2	Welding speed	25 - 35 cm/min
3	Torch angle	0 - 45 ˚
4	Voltage	15 V
5	Tungsten electrode diameter	2.5 mm
6	Inert gas	Argon
7	Nozzle to tip distance	2mm

Table 4 Design matrix

S. No.	Current	Welding speed	Torch angle
1	-1	+1	-1
2	+1	-1	+1
3	-1	-1	+1
4	-1	-1	-1
5	+1	+1	+1
6	+1	-1	-1
7	-1	+1	+1
8	+1	+1	-1

4.Results

4.1Experimentation for the response: (Hardness)

The surface hardness was measured using a Rockwell hardness testing machine with diamond indenter. Initially, a load of 10 kgf was applied so that there is good connect between work piece and diamond shaped indenter. After that a primary load of 150 kgf was applied as per the workpiece material. Waiting time of 60 sec was set so that any fluctuation in dial gauge can be minimized. Now release the load and note down the scale reading. It was planned to take 3 readings, one in the weld area and two on the both sides of the weld bead. The machine used to take the Brinell hardness number (BHN) of the specimens at three different points is shown in *Figure 3*. Measured hardness of different specimens is shown in *Table 5*.

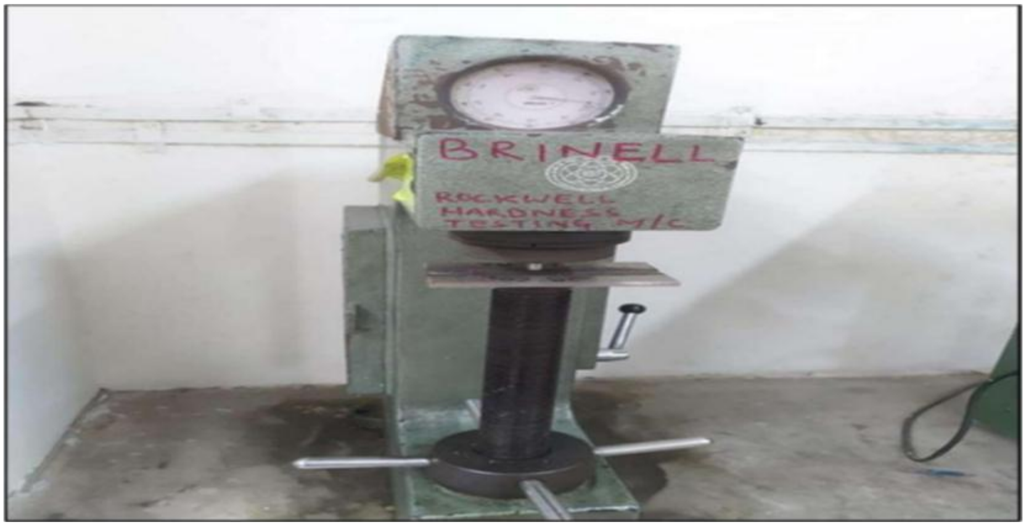


Figure 3 Measurement of the hardness using the Rockwell hardness testing machine

Table 5 Measurement of surface hardness

S. No.	Current (Amp)	Welding speed (cm/min)	Torch angle (degree)	H1	H2	H3
1	140	35	0	47	65	53
2	120	25	45	80	98	90
3	140	35	45	54	70	62
4	140	25	0	69	88	81
5	120	25	0	72	89	83
6	120	35	0	63	86	75
7	120	35	45	76	95	88
8	140	25	45	74	90	85

4.2Generation of the mathematical model

After measuring the surface hardness, a mathematical relation was developed between input parameters and the responses. Hardness was measured at three points, one of the weld zones represented as H2 and other two at the HAZ on the both regions of the weld.

H1 represents the hardness of the heat affected zone on the SS 304 while H3 represents the same on the SS 316 side. Analysis of variance (ANOVA) analysis of hardness at point H1 is shown in *Table 6*. Analysis of the result shows that regression model is significant as its P-value is 0.018. The most significant factor is interaction of current and speed which is having a P-value as 0.018. After that next significant factor is speed having a P-value of 0.022. The model summary says that R-square value is 0.999 shows that 99.9 % variation is shown by the independent variables. The value of R-square (adjusted) and R-square (predicted) is in close agreement.

Table 6 ANOVA for H1

Source	Degree of freedom (DF)	Adjusted sum of squares (Adj SS)	Adjusted mean squares (AdjMS)	F-Value	P-Value
Regression	6	1748.75	291.458	2331.6	0.016
Current	1	222.37	222.368	1778.9	0.015
Speed	1	276.48	276.48	2211.8	0.014
TA	1	22.85	22.837	182.70	0.047
current×speed	1	351.12	351.125	2809.0	0.012
current×TA	1	28.12	28.125	225.00	0.042
speed×TA	1	3.13	3.125	25.00	0.126
Error	1	0.13	0.125	-	-
Total	7	1748.88	-	-	-

S=0.353553, R-square=99.99%, R-square (adjusted)= 99.94%, R-square(predicted)= 99.43%

H1=-147.7+ 2.0250 current+ 9.575 speed + 1.0444 TA - 0.08750 current × speed- 0.007222 current × TA + 0.00333 speed × TA (1)

H2=-287.8+ 3.2500 current+ 15.375 speed + 1.0778 TA - 0.13250 current × speed - 0.008333 current × TA + 0.0056 speed × TA (2)

ANOVA analysis of hardness at point H2 is shown in *Table 7*. Analysis of the result shows that regression model is significant as its P-value is 0.016. The most significant factor is interaction of current and speed which is having a P-value as 0.012. After that next significant factor is speed having a P-value of 0. 014. The regression equation for the model of hardness at point H1 is shown by Equation 1. The model summary says that R-square value is 0.999 shows that 99.9 % variation is shown by the independent variables. The value of R-square (adjusted) and R-square (predicted) is in close agreement.

ANOVA analysis of hardness at point H3 is shown in *Table 8*. Analysis of the result shows that regression model is significant as its P-value is 0.015. The most significant factor is interaction of Current and speed which is having a P-value as 0.013. After that next significant factor is speed with a P-value of 0.015. The Regression equation for the model of Hardness at point H2 is shown by equation 2. The model summary says that R-square value is 0.999 shows that 99.9 % variation is shown by the independent variables. R-square (adjusted) and R-square (predicted) is in close agreement.

Table 7 ANOVA for H2

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	1406.75	234.458	1875.67	0.018
Current	1	86.33	86.329	690.63	0.024
Speed	1	107.23	107.229	857.83	0.022
TA	1	21.45	21.447	171.57	0.049
current*speed	1	153.12	153.125	1225.00	0.018
current*TA	1	21.12	21.125	169.00	0.049
speed*TA	1	1.13	1.125	9.00	0.205
Error	1	0.13	0.125	-	-
Total	7	1406.88	-	-	-

S= 0.35355, R-square=99.99%, R-square (adjusted)=99.93%, R-square (predicted)=99.59%

Table 8 ANOVA for H3

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	1934.75	322.458	2579.67	0.015
current	1	192.64	192.645	1541.16	0.016

Source	DF	Adj SS	Adj MS	F-Value	P-Value
speed	1	217.12	217.124	1736.99	0.015
TA	1	17.54	17.536	140.29	0.054
current×speed	1	300.12	300.125	2401.00	0.013
current×TA	1	28.12	28.125	225.00	0.042
speed×TA	1	15.13	15.125	121.00	0.058
Error	1	0.12	0.125		
Total	7	1934.88			

S=0.333, R-square=0.999, R-square (adjusted)=99.9%, R-square(predicted)= 99.44%

H3=-252.2 + 3.0250 current + 13.625 speed + 0.9444 TA- 0.12250 current × speed- 0.008333 current × TA + 0.01222 speed × TA (3)

The regression equation for the model of hardness at point H3 is shown by Equation 3.

5.Discussion

ANOVA of the formulated model, has P-value, less than 0.05 that signifies the model. The most important factor is the interaction of welding current with welding speed and then the angle of the torch. Secondly, the high numerical quantity of R square and adjusted R-square also propose the satisfactoriness of the formulated exemplary.

5.1Line plot for hardness along the welded specimen

From the line plot it is observed that hardness of SS 304 is less than SS 316. H1 is the hardest on the heat affected zone of material SS 304 and H3 is the hardness of heat affected zone of material SS 316. On the welded specimen, hardness first increases, then start to decrease. Hardness is maximum at the weld joint as the metal solidifies, it recrystallizes itself into fine grain which makes it hard and brittle in nature.

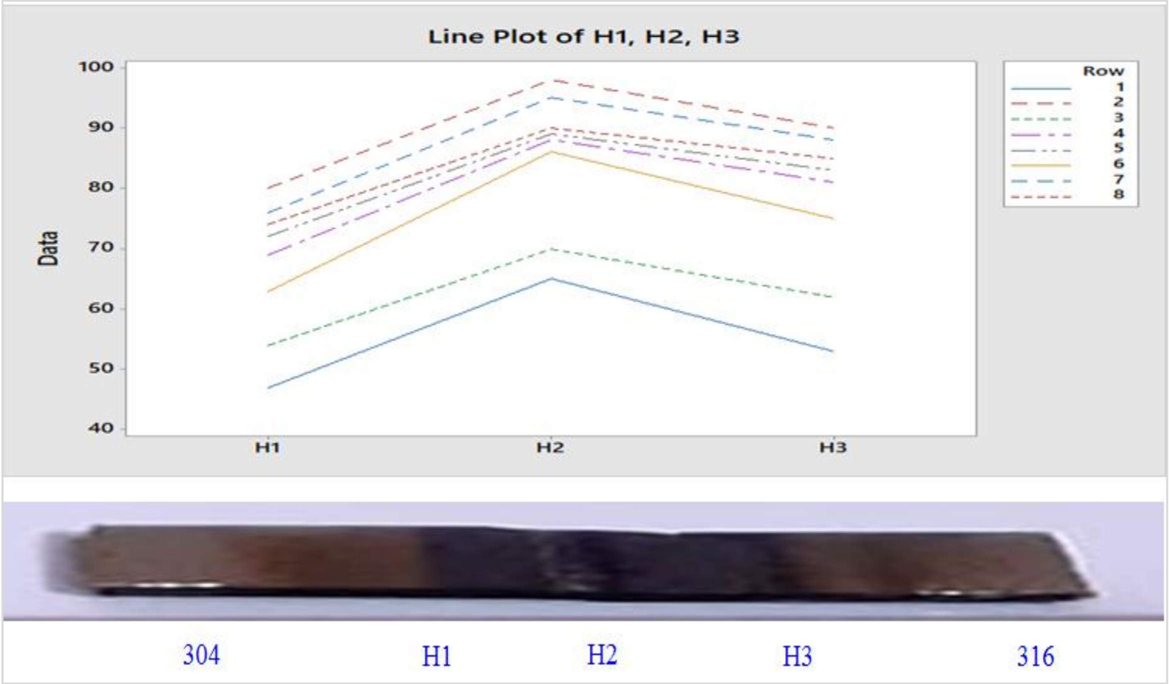


Figure 4 Line plot of hardness along welded specimen for different weld runs

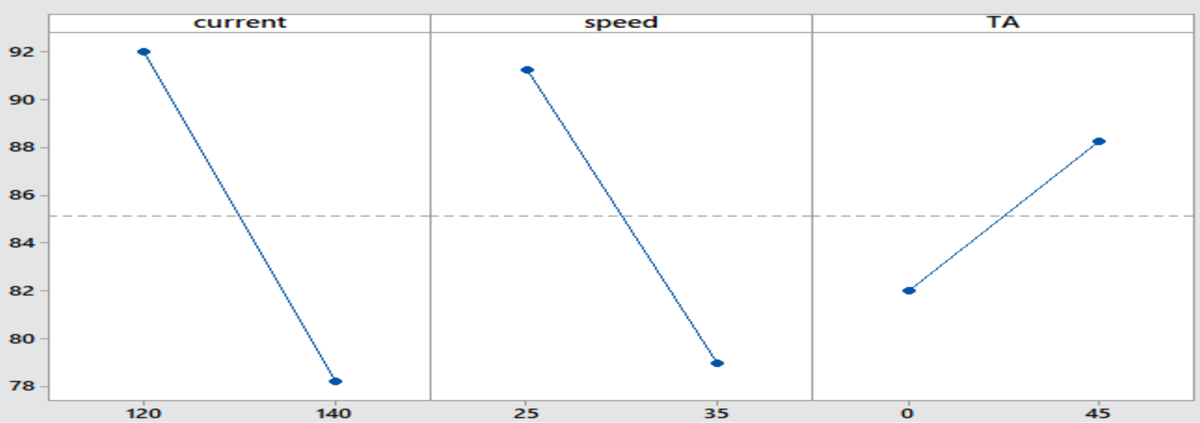
Whereas when move towards the base metal grains gets coarser again resulting decrease in hardness. The line plot of the hardness is shown in *Figure 4*. It is clear from this that the hardness at base metal is less than the weld metal. Position of taking the hardness value from the specimen is also shown in *Figure 4.*

Maximum hardness is observed at the welding current of 120 Amp, 25 cm/min welding speed and 45 ^{0} Torch angle. Minimum hardness is observed at 140 Amp current, 35 cm/min welding speed and 0^{0} torch angle.

5.2Main effect plot of hardness with the input parameters

5.2.1Hardness versus current

An increase in current, hardness starts to decrease because increases in current causes more heat supplied to the base metal which promotes the austenite micro-structure formation that enhance the strength of the metal and decreases the hardness. The result is shown in *Figure 5 (a)*. Hardness plays an important role for mechanical properties of material, if hardness is more material becomes brittle and it cannot be used in many applications. It is also justified by [24, 25] that the value of current can be increased up to a level after that it will decrease the hardness in the TIG welding process.



(a) (b) (c)

Figure 5 Main effect plot of input parameters with hardness

5.2.2Hardness versus speed

An increase in the speed, decreases hardness. The reason is heat input per unit area, per unit time decreases which results in thermal stresses generation, hence less grain boundary and less hardness. The result is shown in *Figure 5 (b)*. Effect of welding speed is explained in the FSW by [26] shows that it has a mixed effect on hardness of the weld zone. In the MIG welding process, hardness increases with the decrease in welding speed. Our results are also in the same line. Mixed effect is also observed by [27] on hardness in MIG welding process for stainless steel alloy-202.

5.2.3Hardness versus torch angle

As the torch angle increases the spread of flame increases the heat affected zone also increases. More stresses causing more grain boundary and make finer grains and ultimately increases the hardness of the metal. The result is shown in *Figure 5(c)*. The angle of the torch was taken as 0^0 and 45^0. Heat input is directly related to the torch angle as we increase the torch angle the heat input will spread over a large area and penetration will decrease.

5.3The combined effect of input parameters on the hardness

5.3.1Hardness vs. speed and current

ANOVA *Table 6, 7, 8* exhibits that interaction of speed and current is significant for all the three cases. *Figure 6* represents the 3D plots of hardness with current and speed. It shows that the decrease in current and speed both will increase the hardness. The maximum hardness can be achieved at current of 120 Amp and a speed of 25 cm/min. The maximum value of hardness achieved as BHN of approximate as 78. The pattern of the plot is also a straight line that indicates that hardness is linearly related to current and speed.

5.3.2Hardness vs. Torch angle and current

ANOVA *Table 6, 7, 8* exhibits that interaction of Torch angle and current is significant for all the three cases. *Figure 7* represents the 3D plots of hardness with current and Torch angle. It shows that the decrease in current and increase in Torch angle will increase the hardness of the joint. The maximum hardness can be achieved at current of 120 Amp and Torch angle of 45^0. The maximum value of hardness achieved as BHN of approximate as 76 to 78. The lower value of current and minimum torch angle gave higher hardness.

5.3.3Hardness vs. Torch angle and speed

ANOVA *Table 6, 7, 8* exhibits that interaction of Torch angle and speed is the most significant factor for the first two cases, whereas in the third case it is also significant with P-value as 0.048. *Figure 8* represents the 3D plots of hardness with speed and Torch angle. It shows that the addition in Torch angle will increase the hardness whereas speed is almost constant. The maximum hardness can be achieved at speed 25cm/min and Torch angle of 45^0. The Maximum value of hardness achieved as BHN of approximate 80. A complete list of abbreviations is shown in *Appendix I.*

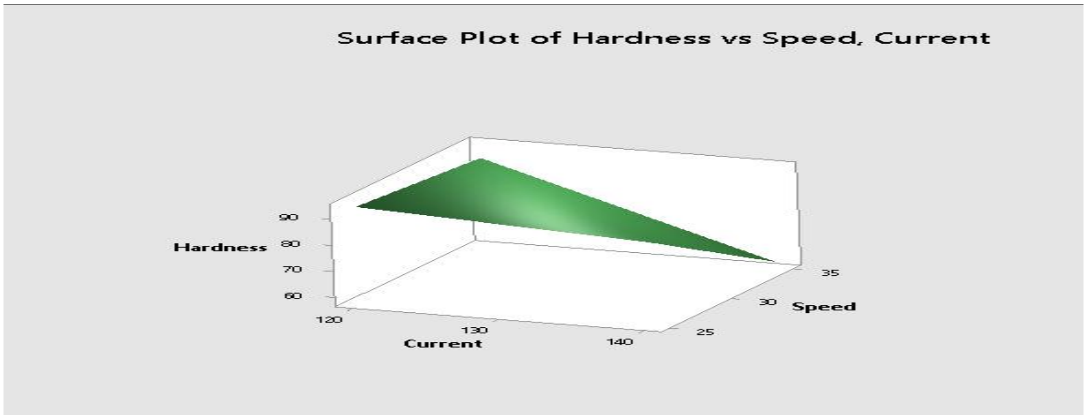


Figure 6 Interaction plot of speed and current with hardness

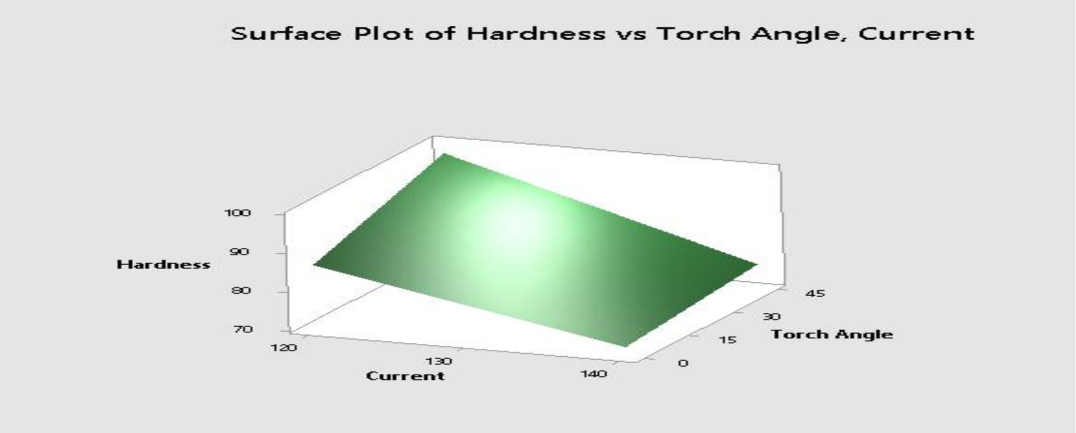


Figure 7 Interaction plot of torch angle and current with hardness

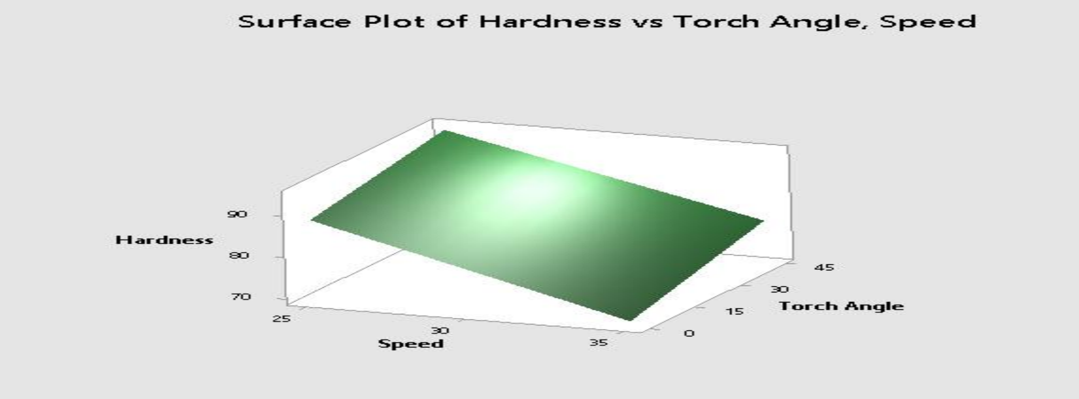


Figure 8 Interaction plot of torch angle and speed with hardness

6.Conclusion

This study is very useful for the industrial application because the fusion of these two metals has better properties against corrosive environment as AISI 304 and AISI 316 contains chromium and nickel. Hardness of the welded joint has a maximum value at the weld zone. It is clear from the study that hardness of these two materials after fusion will increase and maximum hardness can be achieved with the combination torch angle and the welding speed. The approximate value of BHN at this combination is 80. An increase in current will decrease the hardness. So, for the better result current has to kept minimum and the torch angle must also be minimized. Higher hardness, HRB 80 can be achieved at 120 Amp current and 0^0 Torch angle. In case of interaction of welding speed and torch angle, best result of hardness can be achieved at a minimum welding speed and maximum torch angle.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

Author's contribution statement

Deeksha Narwariya: Conceptualization, investigation, data curation, writing-original draft. **Aditya Kumar Rathi:** conceptualization, writing-draft correction, analysis and interpretation of results, writing-review and editing, writing-conclusion.

References

[1] Fei Z, Pan Z, Cuiuri D, Li H, Wu B, Ding D, et al. Effect of heat input on weld formation and tensile properties in keyhole mode TIG welding process. Metals. 2019; 9(12):1-15.

[2] Askeland DR, Fulay PP, Battacharya DK. Essentials of materials science and engineering 2nd ed. S., Australia. 2010.

[3] Jariyaboon M, Davenport AJ, Ambat R, Connolly BJ, Williams SW, Price DA. The effect of welding parameters on the corrosion behaviour of friction stir welded AA2024–T351. Corrosion Science. 2007; 49(2):877-909.

[4] Mvola B, Kah P, Martikainen J. Welding of dissimilar non-ferrous metals by GMAW processes. International Journal of Mechanical and Materials Engineering. 2014; 9(1):1-11.

[5] Devaraj J, Ziout A, Abu QJE. Dissimilar non-ferrous metal welding: an insight on experimental and numerical analysis. Metals. 2021; 11(9):1-31.

[6] Kah P, Shrestha M, Martikainen J. Trends in joining dissimilar metals by welding. In applied mechanics and materials 2014 (pp. 269-76). Trans Tech Publications Ltd.

[7] Shah P, Agrawal C. A review on twin tungsten inert gas welding process accompanied by hot wire pulsed power source. Journal of Welding and Joining. 2019; 37(2):41-51.

[8] Miletić I, Ilić A, Nikolić RR, Ulewicz R, Ivanović L, Sczygiol N. Analysis of selected properties of welded joints of the HSLA steels. Materials. 2020; 13(6):1-25.

[9] Mohyla P, Hajnys J, Sternadelová K, Krejčí L, Pagáč M, Konečná K, et al. Analysis of welded joint properties on an AISI316L stainless steel tube manufactured by SLM technology. Materials. 2020; 13(19):1-14.

[10] Nguyen LT, Hwang JS, Kim MS, Kim JH, Kim SK, Lee JM. Charpy impact properties of hydrogen-exposed 316L stainless steel at ambient and cryogenic temperatures. Metals. 2019; 9(6):1-14.

[11] Gourd LM. Principles of welding technology. London: Edward Arnold; 1986.

[12] Weiss B, Stickler R. Phase instabilities during high temperature exposure of 316 austenitic stainless steel. Metallurgical and Materials Transactions B. 1972; 3(4):851-66.

[13] Li HL, Liu D, Yan YT, Guo N, Feng JC. Microstructural characteristics and mechanical properties of underwater wet flux-cored wire welded 316L stainless steel joints. Journal of Materials Processing Technology. 2016; 238:423-30.

[14] Sharma P, Dwivedi DK. A-TIG welding of dissimilar P92 steel and 304H austenitic stainless steel: mechanisms, microstructure and mechanical properties. Journal of Manufacturing Processes. 2019; 44:166-78.

[15] Sathe SS, Harne MS. Optimization of process parameters in TIG welding of dissimilar metals by using activated flux powder. International Journal of Science and Research. 2013; 4(6):2149-52.

[16] Kulkarni A, Dwivedi DK, Vasudevan M. Dissimilar metal welding of P91 steel-AISI 316L SS with Incoloy 800 and Inconel 600 interlayers by using activated TIG welding process and its effect on the microstructure and mechanical properties. Journal of Materials Processing Technology. 2019; 274(2019):1-14.

[17] Rao VA, Deivanathan R. Experimental investigation for welding aspects of stainless steel 310 for the process of TIG welding. Procedia Engineering. 2014; 97:902-8.

[18] Pasupathy J, Ravisankar V. Parametric optimization of TIG welding parameters using taguchi method for dissimilar joint (Low carbon steel with AA1050). Journal of Scientific & Engineering Research. 2013; 4:25-8.

[19] Badheka VJ, Basu R, Omale J, Szpunar J. Microstructural aspects of TIG and A-TIG welding process of dissimilar steel grades and correlation to mechanical behavior. Transactions of the Indian Institute of Metals. 2016; 69(9):1765-73.

[20] Sathish T, Kumar SD, Muthukumar K, Karthick S. Natural inspiration technique for the parameter optimization of A-GTAW welding of naval steel. Materials Today: Proceedings. 2020; 21:843-6.

[21] Tomaz ID, Colaço FH, Sarfraz S, Pimenov DY, Gupta MK, Pintaude G. Investigations on quality characteristics in gas tungsten arc welding process using artificial neural network integrated with genetic algorithm. The International Journal of Advanced Manufacturing Technology. 2021; 113(11):3569-83.

[22] Mahmood M, Dwivedi VK, Yadav R. Effect of current on the hardness of weld bead generated by TIG welding on mild steel. In advances in industrial and production engineering 2021 (pp. 739-45). Springer, Singapore.

[23] Asibeluo IS, Emifoniye E. Effect of arc welding current on the mechanical properties of A36 carbon steel weld joints. SSRG International Journal of Mechanical Engineering. 2015; 2(9):32-4.

[24] Devanathan C, Shankar E, Sivanand A, Edwin PA. Effect of spindle speed and welding speed on mechanical properties of friction stir welding of AA 6063 with AA 7075. International Journal of Scientific and Technology Research. 2019; 8(10):74-7.

[25] Jassim AK, Ali DC, Laken AH. Effect of metal inert gas welding parameters on the hardness and bending strength of carbon steel plates. In AIP conference proceedings 2021. AIP Publishing LLC.

[26] Bhardwaj B, Singh R, Singh R. To study the effects of welding parameters on MIG welding of stainless steel alloy-202. International Journal of Science Technology & Engineering. 2017; 4(1):132-40.

[27] Huang B, Liu J, Zhang S, Chen Q, Chen L. Effect of post-weld heat treatment on the residual stress and deformation of 20/0Cr18Ni9 dissimilar metal welded joint by experiments and simulations. Journal of Materials Research and Technology. 2020; 9(3):6186-200.

[28] Kulkarni A, Dwivedi DK, Vasudevan M. Microstructure and mechanical properties of A-TIG welded AISI 316L SS-Alloy 800 dissimilar metal joint. Materials Science and Engineering: A. 2020; 790:1-11.

[29] Pahlawan IA, Arifin AA, Marliana E, Irawan H. Effect of welding electrode variation on dissimilar metal weld of 316l stainless steel and steel ST41. In IOP conference series: materials science and engineering 2021 (pp. 1-8). IOP Publishing.

[30] Xu Y, Hou X, Shi Y, Zhang W, Gu Y, Feng C, Volodymyr K. Correlation between the microstructure and corrosion behaviour of copper/316 L stainless-steel dissimilar-metal welded joints. Corrosion Science. 2021; 191:1-12.

[31] Pradhan DK, Sahu B, Bagal DK, Barua A, Jeet S, Pradhan S. Application of progressive hybrid RSM-WASPAS-grey wolf method for parametric optimization of dissimilar metal welded joints in FSSW process. Materials Today: Proceedings. 2022; 50:766-72.

[32] Sun Y, Xue H, Yang F, Wang S, Zhang S, He J, et al. Mechanical properties evaluation and crack propagation behavior in dissimilar metal welded joints of 304 L austenitic stainless steel and SA508 low-alloy steel. Science and Technology of Nuclear Installations. 2022; 2022:1-13.

Deeksha Narwariya M.Tech student of NSUT. She presented 3 papers in the international conference during her M.Tech course. Her area of interest is TIG and MIG Welding Processes Applications of Dissimilar Metal Welding. She is presently perusing her research work outside India.
Email: deeksha098@gmail.com

Aditya Kumar Rathi obtained his B.E degree from, Karnataka University Dharwad in the year 1993 with Mechanical Engineering. He obtained his Master’s degree from Delhi College of Engineering in the year 1999. He joined as Senior Scientific Assistant at Netaji Subhas Institute of Technology in February 1994 and got the promotion as Lectures in year 2002 and subsequently as Assistant Professor in year 2007 and Associate Professor from 2015 onwards, he has a long teaching experience at NSIT about 27 years and published 15 papers in International Journal, 10 papers in International and National Conference. He is a life member of ISTE and IIW. His area of interest is Design and Development of Submerged Arc Welding Fluxes.
Email: aditya.kumar@nsut.ac.in

Appendix I

S. No.	Abbreviation	Description
1	AISI	American Iron and Steel Institute
2	AdjMS	Adjusted Mean Squares
3	Adj SS	Adjusted Sum of Squares
4	ANOVA	Analysis of Variance
5	ANN	Artificial Neural Network
6	BHN	Brinell Hardness Number
7	DF	Degree of Freedom
8	DMW	Dissimilar Metal Welding
9	FSW	Friction Stir Welding
10	FSSW	Friction Stir Spot Welding
11	GA	Genetic Algorithm
12	GTAW	Gas Tungsten Arc Welding
13	HAZ	Heat Affected Zone
14	HRB	Rockwell Hardness at B scale
15	HSLA	High Strength Low Alloy
16	MMA	Manual Metal Arc
17	MAG	Metal Active Gas
18	MIG	Metal Inert Gas
19	RSM	Response Surface Methodology
20	SEM	Scanning Electron Microscope
21	SLM	Selective Laser Melting
22	SWAW	Shielded Metal Arc Welding
23	TiO_2	Titanium Die Oxide
24	TIG	Tungsten Inert Gas
25	TTP	Temperature Time Precipitation
26	V-I	Voltage and Current
27	3D	Three Dimensional

Hindawi
International Transactions on Electrical Energy Systems
Volume 2023, Article ID 4303491, 19 pages
https://doi.org/10.1155/2023/4303491

Research Article

Analysis of Oscillations during Out-of-Step Condition in Power Systems

Zainab Alnassar [1,2] and S. T. Nagarajan [1]

[1] *Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India*
[2] *Department of Electrical Power Engineering, Al-Baath University, Homs, Syria*

Correspondence should be addressed to Zainab Alnassar; zainabalnassar0@gmail.com

Received 27 November 2022; Revised 10 February 2023; Accepted 16 February 2023; Published 28 February 2023

Academic Editor: Davide Falabretti

Power systems interconnected by weak tie lines can be subject to low-frequency oscillations because of disturbances which excites the low-frequency modes of the system; furthermore, these oscillations can be stable or unstable. The latter, if not treated, can cause severe oscillations that divide the network into smaller groups which oscillate against each other, leading to out-of-step (OOS) condition in the network. The detection of OOS condition is a challenge for power system operators in real time as it is difficult with conventional measuring instruments, to identify the instant at which the bus voltage angle between two areas connected by tie line falls out of synchronism. Conventionally, the detection of OOS condition has been carried out with impedance-based relays along with power swing blocking. With the advent of synchrophasor-based measurement units, it is now possible to measure the bus voltage angle in real time leading to direct detection of OOS condition in power system and intentional islanding. In this article, a systematic analytical study and EMT time-domain simulation study have been performed to simulate OOS condition in the power systems, and its detection is based on the voltage angle difference with wide-area measurement systems (WAMSs). The article has been carried out on a single machine infinite bus (SMIB) to monitor generator OOS, Kundur's two-area system to detect interarea OOS, and the IEEE39 bus system to identify OOS condition and with a new algorithm. Time-domain simulation studies carried out with OPAL-RT real-time simulator in HYPERSIM environment corroborates with analytical results.

1. Introduction

Oscillation in the power grid results from the power system's dynamic nature [1]. Among the oscillation modes mentioned in [2], the current attention is drawn to the interarea oscillation mode due to the severity of its results and the fact that it can occur in unexpected lines far from the existing disturbance. The lines in the interarea may not have sufficient damping for these low-frequency oscillations. So, the low-frequency interarea oscillations can cause stress on these weak tie lines and can lead to loss of synchronization by forming coherent groups of generators that oscillate with the same angular speed against others, which in turn leads to cascading failures and ends with network collapse [3, 4]. Maintaining the reliability and security of the power system network requires a basic understanding of oscillations and their dangers, namely, stable and unstable oscillations. Increasing the disturbance severity, which can be either faults, increase or decrease in generation, loss or increase in a large block of load, or changes in network configuration [5], gives rise to the risk of these oscillations leading to so-called out-of-step (OOS) [6].

The current challenge for grid operators is maintaining the stability of the network in the presence of low-frequency oscillations and preventing the OOS condition. Concerning the recent change in power generation toward renewable energy sources (RES), it is worth mentioning that these inverter-based resources, which do not have rotating mass coupled to the grid, lead to a decrease in inertia in the power system. The resulting low inertia in the power system has an impact on the OOS condition [7, 8]. The OOS condition in literature has been addressed in two cases. The first one is for synchronous generators and where the power plant falls out of synchronism after a disturbance contributed by the rotor angle of the generator. Secondly, the OOS condition between two coherent areas contributed by interarea mode. The OOS condition is driven in the power system mainly due to fault disturbances which can excite the generator rotor angle mode or the interarea mode in large interconnected systems. For the faulted system to be stable, the fault should be removed before a time, known as the critical clearing time (CCT). However, when fault clearing time is delayed due to the failure of the breaker, for example, it will lead to out-of-step condition [8].

Upon survey of the literature for OOS, the authors of [5] have proposed a method to protect ultrahigh voltage transmission lines from power oscillation due to wind generation integration and differentiate between symmetrical faults and power oscillation. The authors of [9] have compared the dynamic characteristics of an extensive transmission system with and without photovoltaic penetration, and its stability state has been observed. In [10], out-of-step oscillation has been observed due to sequence faults in a large power system despite clearing the faults at CCT, leading to the need for controlled splitting. Abedini et al. [11] have proposed the free measurement faster-than-real-time (FTRT) method to predict OOS oscillation based on equal area criterion EAC mathematical formulation. In [12], a real-time estimation has been developed to distinguish the OOS oscillation of the generator based on active power derivation. The proposed scheme has relied on the polarity of extracted angular velocity and acceleration data from EAC. In reference to [13], a new EAC approach to detect OOS has been tested by replacing the power-angle curve with the power-time curve.

Traditionally, OOS condition in the power system was detected using an impedance-based distance relay [14]. However, with the implementation of synchrophasor technology on a large scale in power grids, many parameters can be monitored to observe the low-frequency oscillations and detect OOS condition with the help of a phasor measurement unit (PMU) characterized by a phasor (magnitude and angle) output. In [5], an algorithm has been developed based on measuring the angle of the current between sending and receiving ends using the PMU. In [15], the proposed method has used the magnitude of positive sequence voltage obtained from PMU to develop fast and accurate intelligent devices (IDs) to detect symmetrical/asymmetrical faults during power oscillation. The authors of [16] have traced the trajectory of the generators power angle with respect to the frequency in the plane (δ-ω) to implement Zubov's boundary method for stable and OOS oscillation discrimination. In [10], the proposed algorithm can detect out-of-step oscillation in its first cycle and determine its location by obtaining the frequency of the bus voltages using PMU. The feature of the proposed algorithm [17] is independence from the details of the power system and only depends on the phase angle of the terminal voltage of generators collected from PMU.

The latest literature [18–22] have carried out research studies regarding the design of wide-area power system stabilizer (PSS) using PMU for damping interarea oscillations to improve the stability of the system. PSS effectively increases damping of low-frequency oscillations by shifting unstable mode to the left side of the complex plane. However, PSS cannot guarantee the stability of the system in case of large topological changes which in turn leads to OOS condition in the system. In this study, wide area-based OOS condition detection, during the unstable power swing (i.e., the time until which the system remains synchronized from the occurrence of the disturbance in case of an unstable power swing), has been proposed. In this way, this article is different from the damping of oscillation in power systems with PSS.

Though several methods have been proposed in the literature to avoid the mal operation and overcome the difficulties of the traditional impedance-based method with continuously increasing RES in the power system, to avoid the complex calculations required for setting of impedance-based traditional OOS relay, it is worthy to go for a measurement-based method. Thus, the first contribution in this article is an algorithm which is simple and does not need the complex data of the system; it only needs the bus voltage angle obtained directly from PMU without the requirement of any computational estimation. The second contribution is that the presented method can detect both generator mode and interarea mode, as previous studies have been concerned with either one of these modes in their research. Furthermore, a systematic study on the analysis of OOS condition from small to large power systems under different disturbance scenarios in the time domain as well as in the frequency domain (eigenvalues) has been performed. Thirdly, this article shall be helpful for monitor OOS in the power system which can be used to identify unintentional catastrophic tripping of transmission lines due to voltage angle separation between two coherent areas. The proposed OOS detection method can be considered to identify the instant at which the voltage angle separation of a generator (or between two coherent areas) falls out of synchronism, leading to instability in the power system.

To summarize, this article performs an OOS analysis of power oscillation in grid-connected generator and interconnected power systems. An out-of-step tripping algorithm has been developed and implemented in the presence of different disturbance scenarios, using the information output from wide-area measurements. Simulation studies have been performed on SMIB, Kundur's 2-area systems, and IEEE39-bus system in OPAL RT real-time simulator with 50 μ sec time step in HYPERSIM software. Results in the time domain have been verified with eigenvalue analysis.

2. Out-of-Step Background

Out of step is a condition in a power system where the phase angle between two buses in an interconnected power system or rotor angle of a generator and its generator bus exceeds 180 degrees [23]. OOS condition can occur between the generator and the network which is known as a local mode with frequency oscillation in the range (0.7–2) Hz or between coherent areas causing interarea oscillation mode in the network transmission lines with low-frequency (0.1–0.8) Hz [24]. The generator falls out of synchronism with the generator bus in case of OOS condition, and in case of an interconnected power system, the tie line oscillations shall not be stable. This article has focused on interarea OOS condition and on generator as well. The basics of the OOS condition can be analysed with steady-state power transfer equation, swing equation, and modal values explained as follows.

2.1. Steady-State Power Transfer. The transmitted active power through a transmission line in a grid, as shown in Figure 1 [25], is given as follows:

$$P_e = \frac{V_s.V_R}{X_s}. \sin \delta. \tag{1}$$

In equation (1), V_S and V_R are sending and receiving ends voltage of the line, δ is the voltage angle difference ($\delta_s - \delta_R$), and X_S is the equivalent reactance of the transmission line. Typically, there is a voltage angle difference between two ends of any transmission line to achieve the power transfer principle as a power-angle ($P-\delta$) characteristic [25]. However, if this angle difference oscillates under exposure to any disturbance, it should not reach the value of 180°. Otherwise, the oscillation will not dampen out, and this angle difference will keep increasing, leading to a loss of the synchronism of these two areas, according to equal area criterion (EAC) [26].

2.2. Swing Equation. In case of a disturbance to the generator connected to the bus, the rotor angle of the generator oscillates according to the swing equation [27]:

$$\frac{H}{\pi . f_0} \frac{d^2\delta}{\mathrm{d}t^2} = P_i - P_t, \tag{2}$$

where H is the area inertia per unit, δ is the electrical power angle in radians, which will vary with time, and P_i is the area input power and P_t is the electrical transmitted power. OOS condition occurs when the rotor angle oscillations cross 180 degrees away from the bus voltage angle to which it is connected.

2.3. Modal Values. Modes are the roots of the characteristic equation of the power system matrix, and these can be determined by eigenvalue analysis. With the use of eigenvalues, the nature of these oscillations mode can be indicated as follows [1]:

$$\lambda = \sigma \pm j\omega. \tag{3}$$

The damping coefficient is given as follows:

$$\tau = \frac{-\sigma}{\sqrt{\sigma^2 + \omega^2}}. \tag{4}$$

Moreover, the frequency of oscillations is given as follows:

$$f = \frac{\omega}{2\pi}. \tag{5}$$

The positive real part of the eigenvalue indicates an unstable system with a negative damping coefficient, and the imaginary part decides the oscillation mode.

The OOS condition due to this small signal stability cannot be identified easily with conventional measurements. Traditionally impedance-based distance relays with blinders were used to identify the OOS condition due to the interarea mode in the system.

2.4. Wide Area Measurement System (WAMS). With the advancement of synchrophasor technology, an enormous number of research studies have been directed to the wide-area measurement system (WAMS). WAMS is a network of PMUs connected by a communication system. All PMUs of various substations are connected to one phasor data concentrator (PDC) [28], which is a central computer that collects information from all PMUs and stores it for later use to predict how the future of the network will be. PDC can monitor the system in real time to warn the operators or send a trigger signal in case of any disturbance. The main feature of this technology is the GPS synchronization with a phasor measurement unit (PMU) to give output with time-stamped measurements [23]. PMU is a measurement device that can report data at various data rates (10, 30, 60, and 120 frames per second), to convert the input instantaneous values to the time-stamped samples as represented in Figure 2.

2.5. Traditional Algorithm. Traditionally, out-of-step condition was detected in the power system with impedance measurement, based on distance relay with a "blinders scheme" [29]. Power swing blocking (PSB) and out-of-step tripping (OOT) functions have been added to the impedance relay. In this scheme, if the impedance trajectory during the fault or oscillation travels from the right outer blinder to the right inner blinder (ΔZ) during time $\Delta t > \Delta t_{\text{setting}}$, the outer zone Z3 function will be blocked from tripping as the distance relay distinguishes stable power swing from fault. While OOT function permits the distance relay to trip immediately when the impedance trajectory leaves the right inner blinder as the relay identifies the OOS condition, as represented in Figure 3.

As aforementioned, blinders' scheme requires proper settings for ΔZ and $\Delta t_{\text{setting}}$ along with blinder locus depending on system configuration which may change at any time because of any contingency. Therefore, to implement this multiple functionality scheme, a thorough study of system stability and detailed analysis of the system in presence of different types of disturbances has to be carried out precisely which is difficult in a larger system. Otherwise, maloperation of relay may occur causing undesired triggering.

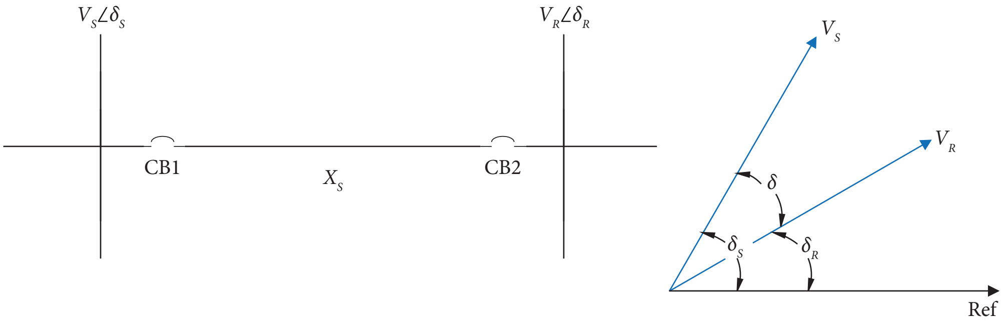


FIGURE 1: Sending and receiving ends of the transmission line.

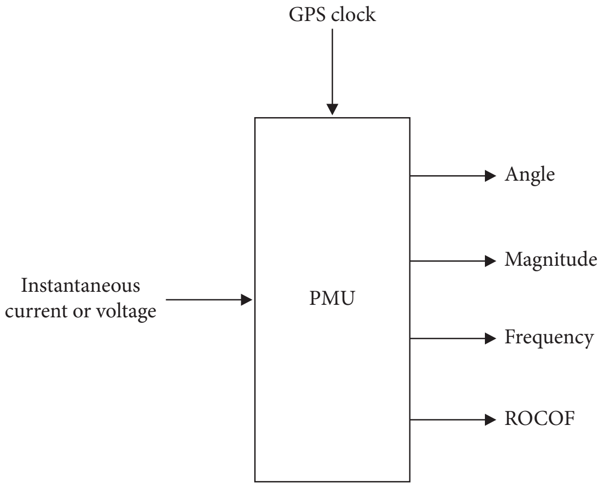


FIGURE 2: PMU input and output.

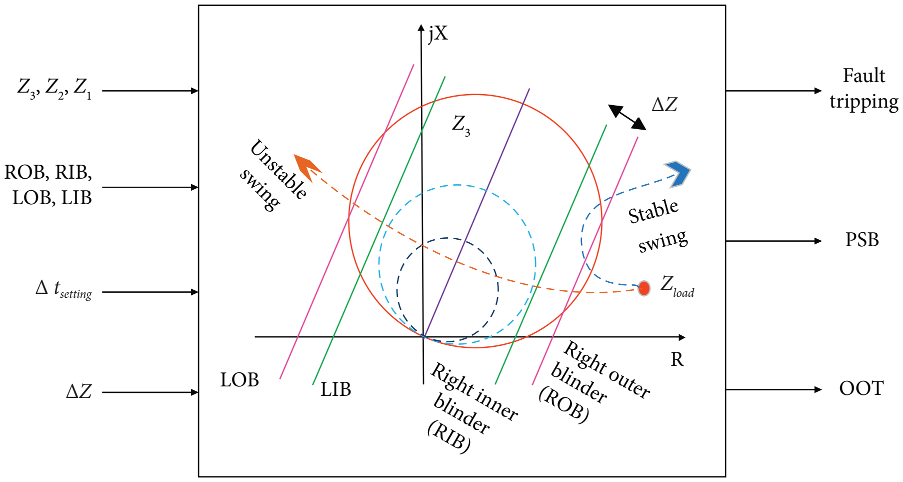


FIGURE 3: Blinder scheme.

2.6. Proposed Algorithm. Taking advantage of the PMU data stream at substations, the bus voltage angle can be measured from two PMUs at the ends of the tie line. The test configuration is shown in Figure 4.

The voltage angle difference $(\delta_s - \delta_R)$ has been implemented to trip at out-of-step condition after collecting PMUs data [30], as shown in Figure 5.

The voltage angle difference tool of the tie line has been carried out using MATLAB/Simulink and has been imported to HYPERSIM for observation.

Figure 6 demonstrates the proposed relay algorithm to detect OOS condition based on the voltage angle difference between the buses at both ends of the tie line. The trip decision is taken when the voltage angle difference is equal to

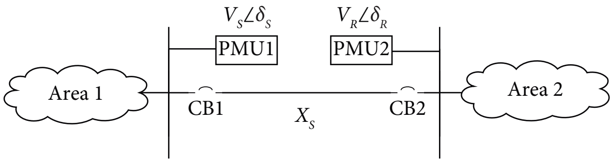


Figure 4: PMUs on buses on both ends of the tie line.

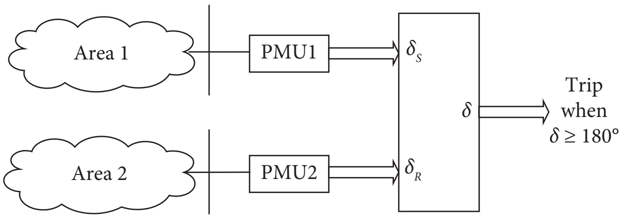


Figure 5: Implementation of OOS tripping.

or greater than 180°. The main advantages of the proposed approach are the simplicity, measurement based, and independent of the system topology.

3. Case Study

The study has been carried out on OPAL-RT real-time simulator using HYPERSIM environment, which is software that helps to simulate and test power systems with realistic visualization. PMU is a measurement device that can be utilized to obtain phasor quantities, which has been simulated in Simulink and imported into HYPERSIM, to collect bus voltage angle. Furthermore, HYPERSIM can provide the damping coefficient and the frequency of oscillations of the system from which eigenvalue can be calculated.

The analysis has been carried out on three configuration systems to monitor the behaviour of system stability under different disturbance scenarios. First, the effect of changing the generated power of SMIB on CCT, considering a three-phase fault, has been simulated at the sending end of one of the lines. Second, on Kundur 2 area system, three scenarios have tested changing inertia, load variation, and changing network configuration. Lastly, interarea oscillation has been created by 2 three-phase faults at different locations on the 39-bus system to observe the behaviour of out-of-step condition in the extensive system.

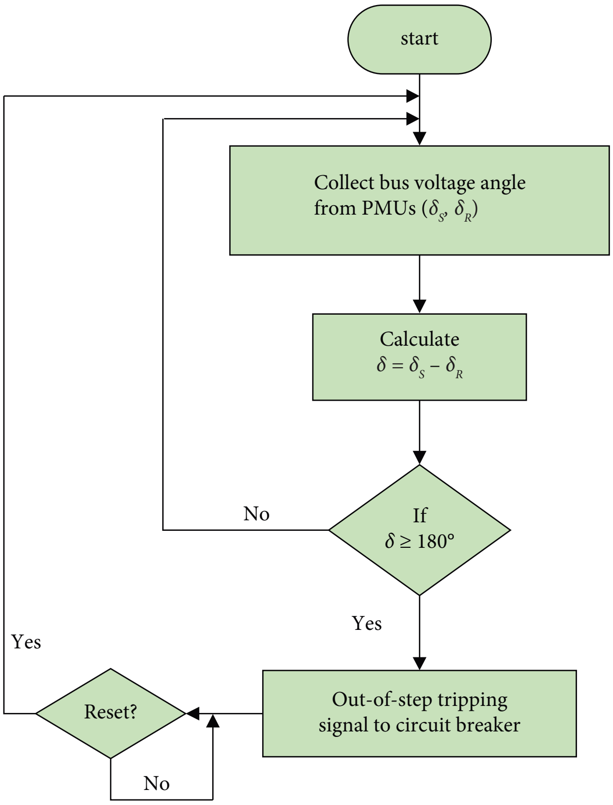

start
Collect bus voltage angle from PMUs (δ_S, δ_R)
Calculate $\delta = \delta_S - \delta_R$
If $\delta \geq 180°$
No
Yes
Out-of-step tripping signal to circuit breaker
Reset?
Yes
No

Figure 6: Proposed algorithm based on voltage angle difference.

3.1. Case Study I: Single Machine System (Generator Mode). The study system in this case is a 60 Hz single machine connected to an infinite bus through two transmission lines (Figure 7). The parameters of SMIB are given in Table 1. A three-phase fault has been applied on line 2, at the sending end of the line at 1 sec. The fault has been cleared by tripping of the lines from both ends. The time of clearing the fault has been varied according to the different scenarios of power transmission. The OOS condition, at which the generator falls out of synchronism with a phase angle difference of 180 degrees, has been detected with PMU measurements from bus 2 and bus 3. The relay algorithm developed has been used to detect the OOS condition and trip the generator. The results have been analytically verified with the calculated values of CCT.

The CCT has been calculated as given in [31] using critical clearing angle (CCA) δ_c with the following equations:

$$\begin{aligned}\cos\delta_c &= \frac{P_0\left(\delta_{\max}-\delta_0\right)+P_{3\max}\cos\delta_{\max}-P_{2\max}\cos\delta_0}{P_{3\max}-P_{2\max}},\\ \text{du ring fault}\, P_{2\max}\sin\delta &= \frac{V_sV_R}{X_{df}}\sin\delta,\\ \text{after fault} P_{3\max}\sin\delta &= \frac{V_sV_R}{X_{af}}\sin\delta.\end{aligned}\tag{6}$$

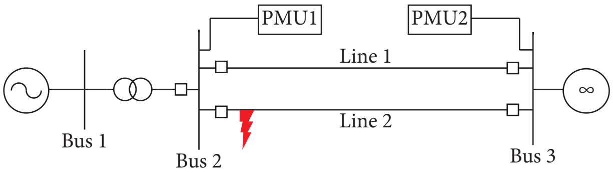


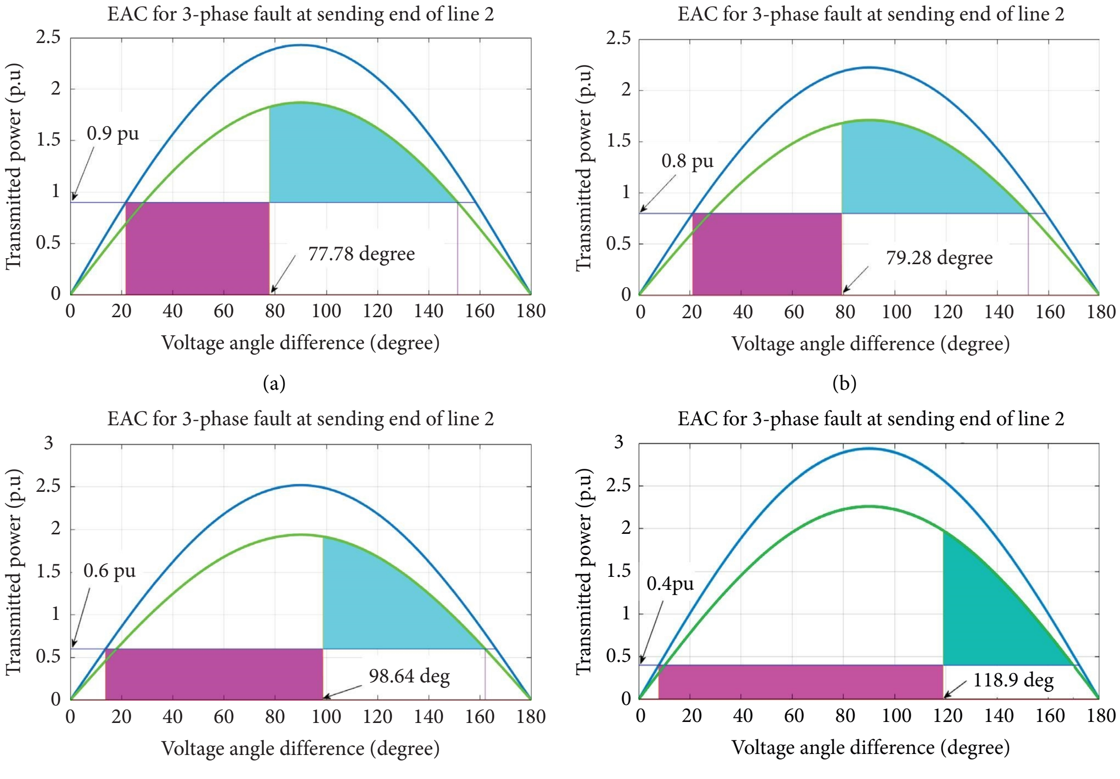
Figure 7: SMIB with fault on line 2.

Table 1: SMIB parameters.

Device	Parameter (p.u)
Transformer imp	0.05
Generator (X'_d)	0.3
Generator (H)	6.112
Line 1 = line 2	0.3
$X_{\text{before fault}}$	0.5
$X_{\text{during fault}}$	inf
$X_{\text{after fault}}$	0.65
Network	Inf. bus voltage = 1

Table 2: Calculated and simulated CCT values with changing power.

Scenarios	Power flow (pu)	CCT (sec) In EAC and time domain
Scenario1	$0.9 + 0.430j$	0.265
Scenario2	$0.8 + 0.074j$	0.287
Scenario3	$0.6 + 0.450j$	0.400
Scenario4	$0.4 + 0.910j$	0.561



(a)

(b)

(c)

(d)

Figure 8: SMIB equal area criteria for different values of active power: (a) 0.9 pu, (b) 0.8 pu, (c) 0.6 pu, and (d) 0.4 pu.

X_{df} = inf because the fault occurs on sending end. X_{af} = the reactance of line 1.

$$t_c = \sqrt{\frac{2H(\delta_c - \delta_0)}{\pi f_0 P_0}}. \tag{7}$$

OOS condition has been evaluated in the power angle domain in the light of the same fault applied at 1 second for all generating scenarios in Table 2. To verify the OOS condition at CCT, equal area criteria have been plotted for all scenarios observing CCA for corresponding generated power.

Figure 8 demonstrates equal area criteria with the value of CCA required to get CCT. Table 2 shows the CCT from the time of the fault. The table illustrates that decreasing generated active power increases CCT. Thus, clearing time (CT) of the fault on or before these CCT will maintain the system's stability as verified in time domain (Figures 9 and 10).

Traditional transmission line relays are coordinated in such a way as to clear the fault at a predetermined time. Moreover, the integration of renewable energy and its generation relies on the renewable sources. This may cause instability issue and out-of-step condition when, for example, the relay has already been set to operate at 0.4 seconds (scenario 3 in Table 2) from the time of the fault.

Now, when the generated power raises up to 0.8 pu, which requires lesser time to clear the same fault, for the system to be stable, the transmitted power oscillates continuously, as in Figure 10.

To capture the previous said phenomenon in WAMS, bus voltage angle difference is measured through PMU at Bus 2 and Bus 3. Figure 11 plots stable and unstable bus voltage angle difference oscillations due to OOS condition. If there is any delay in clearing time (CT) of the fault, the system will be out-of-step, and there will be a loss synchronization when the voltage angle difference goes beyond 180° as in Figure 11, which compares clearing the fault on and after CCT.

To verify the results in small signal stability criteria, eigenvalues have been calculated using damping ratio and oscillation frequency, which have been obtained from the model identification method in SCOPE VIEW software from Figures 9 and 10. The stability conditions have been evaluated by using eigenvalues in case of fault clearing time (CT) equal to CCT and greater than CCT.

As shown in Table 3 the real parts of the eigenvalues are negative when clearing the fault happens at CCT, indicating a stable system. Unlikely, when CT > CCT, the real parts of the eigenvalues are positive which denotes the OOS condition for different scenarios considered.

From this case study, it can be inferred that the OOS condition in generator detected with phase angle difference from PMU measurements in real time confirms to the EAC criterion, which is difficult to implement in generators otherwise.

3.2. Case Study II: Kundur Two Area System (Interarea Mode).

The second case study has been carried out on Kundur's two area system. It is a 50 Hz power system consisting of two symmetrical areas, as shown in Figure 12. Each area has two generators and five buses in addition to the two middle lines with a bus; thus, there are 11 buses. Two loads are connected with buses 4 and 6, and detailed parameters are in [30]. The transmitted power from bus 4 to 6 has been measured, and two PMUs have been connected with bus 3 and bus 7. Thus, the tie lines' transmitted power and voltage angle difference have been observed to identify out-of-step conditions during the scenarios of varying loads, inertia, and network configuration.

3.2.1. Scenario I: Load Variation.

In this scenario, the OOS condition has been simulated in two area systems by varying the load in the system so that the interarea mode oscillations are excited. The initial condition for tie line power transfer of 410 MW has been set and load 2 on bus 6 has been varied by increasing the load by 30% and 50% at 1 sec. Figure 13 shows how the power oscillations due to interarea mode for the two cases. It can be observed that the system remains stable in case of a 30% change where as it turns unstable for 50% change.

Therefore, the OOS condition for 50% change has to be detected in real time so as to prevent the collapse of the system. Figure 14 represents the effect of load variation, on the voltage angle difference between bus 7 and bus 3, measured through PMU for both cases. It can be observed that, for the unstable case at 180-degree separation, two areas lose their synchronization and run under out-of-step conditions. Eigenvalue analysis for this scenario has been performed and tabulated in Table 4 to verify the time-domain results.

3.2.2. Scenario II: Change in Inertia.

In this section, the effect of inertia has been observed by applying a three-phase fault on one of the lines between bus 5 and bus 6 at 1 sec; then, the fault has been cleared at 1.08 sec [30]. Under this condition, two subscenarios have been studied, giving the inertia of generator 3 two values (6.175 and 3.175). Figure 15 shows a comparison between stable and unstable transmitted power during a change of inertia.

Low inertia increases the oscillation of the generator rotor, consequently causing unstable power flow in the interconnected system and increasing the bus voltage angle difference of the tie lines, as shown in Figure 16.

It can be noted from equation (7) that CCT relies on system inertia and decreasing inertia requires decreasing the time of clearing the fault [8]. Noting that, in this article, in both scenarios, clearing the fault has been carried out at 1.08 sec, causing the out-of-step condition in case of inertia 3.175. Therefore, under the same fault condition, low inertia causes OOS condition, which has to be considered to protect

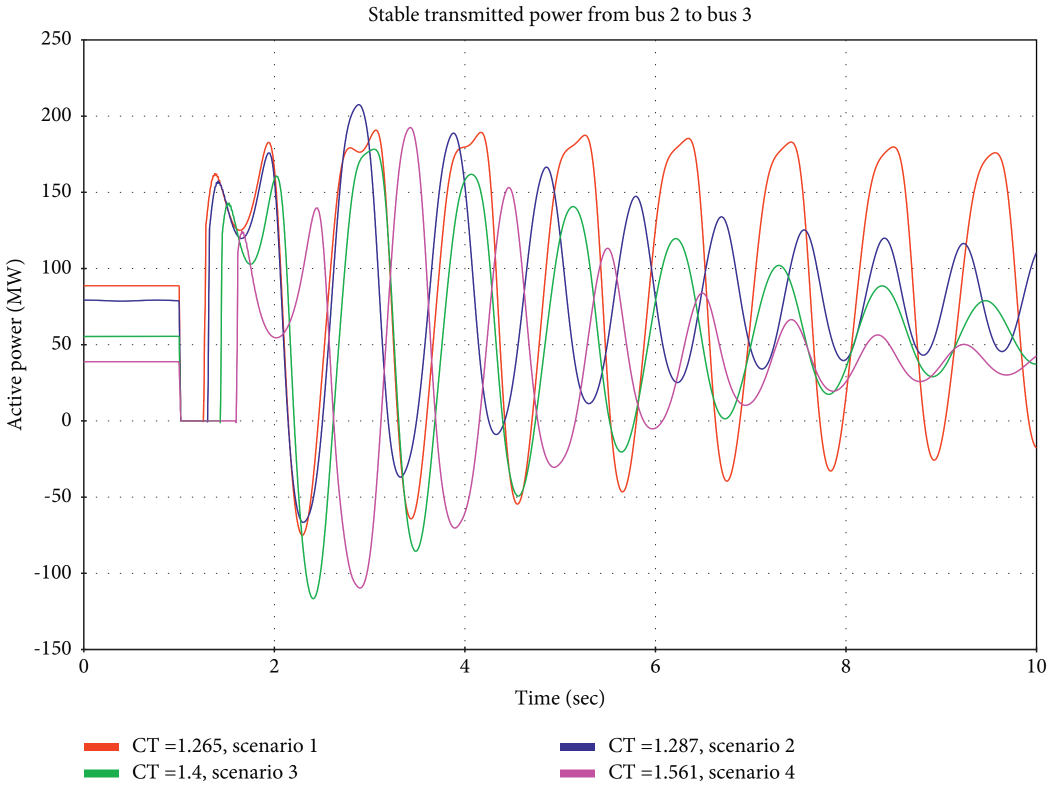


Figure 9: Transmitted power from bus 2 to bus 3 for different generated power when CT = CCT.

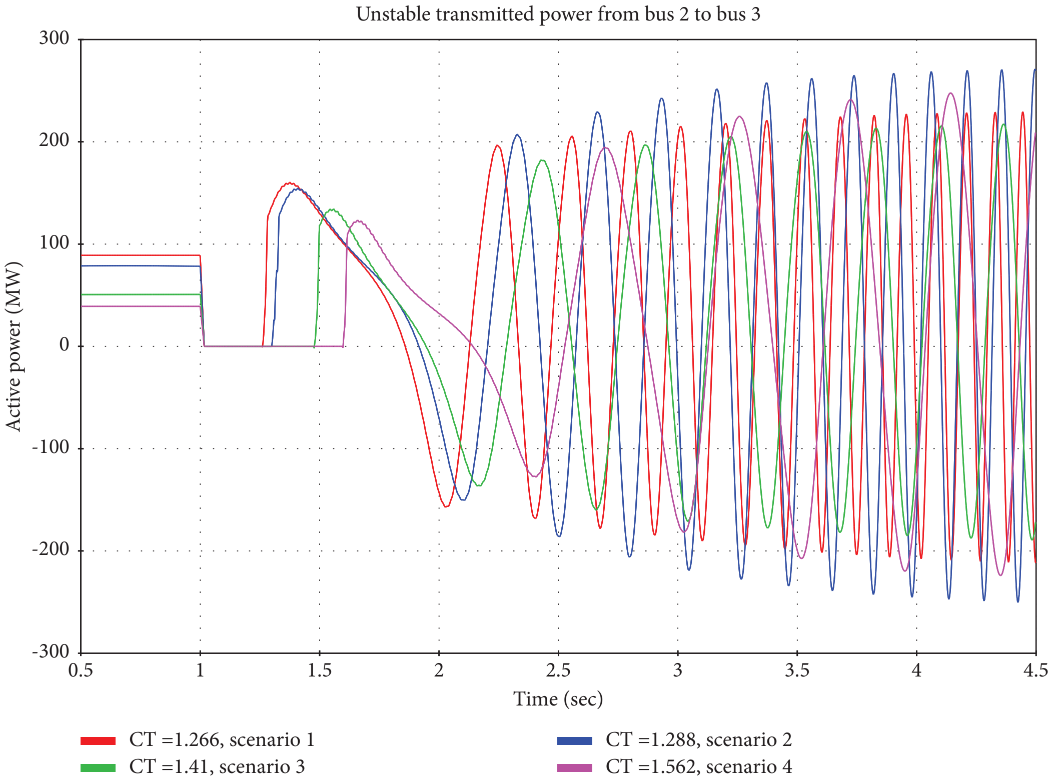


Figure 10: Transmitted power from bus 2 to bus 3 for different generated power when CT > CCT.

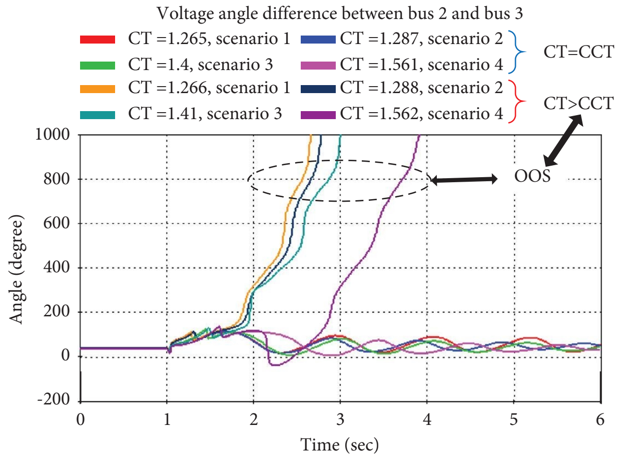


Figure 11: Voltage angle difference between bus 2 and bus 3 for different generated power in cases CT = CCT and CT > CCT.

Table 3: SMIB eigenvalues.

Scenario Power flow (SMIB)	$\lambda = \sigma \pm j\omega$ (stable) (CT = CCT)	$\lambda = \sigma \pm j\omega$ (unstable) (CT > CCT)
$0.9 + 0.43j$	$-0.3474 + 4.950j$	$0.2287 + 87.96j$
$0.8 + 0.074j$	$-0.335 + 4.390j$	$0.2243 + 65.97j$
$0.6 + 0.45j$	$-0.7100 + 4.390j$	$0.0591 + 29.53j$
$0.4 + 0.91j$	$-0.6500 + 4.020j$	$0.056 + 16.960j$

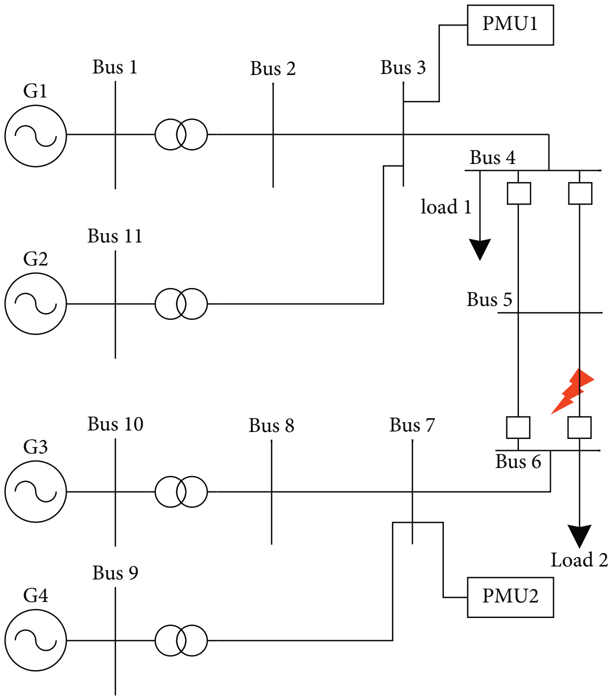


Figure 12: Kundur two area system.

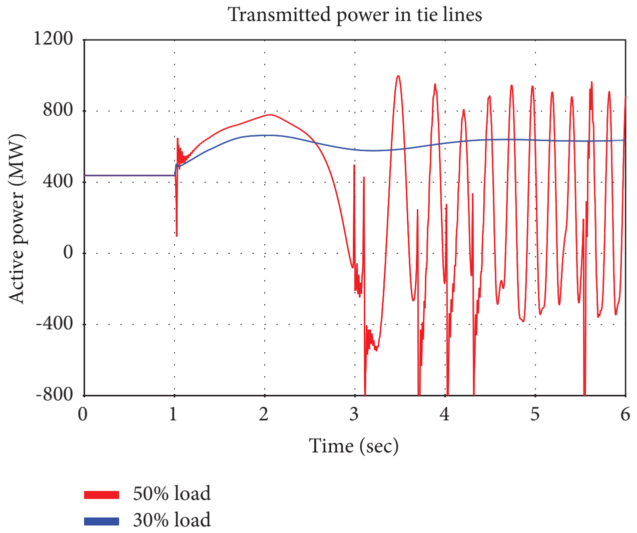


Figure 13: Power oscillations in tie lines for cases of 30% and 50% load increase.

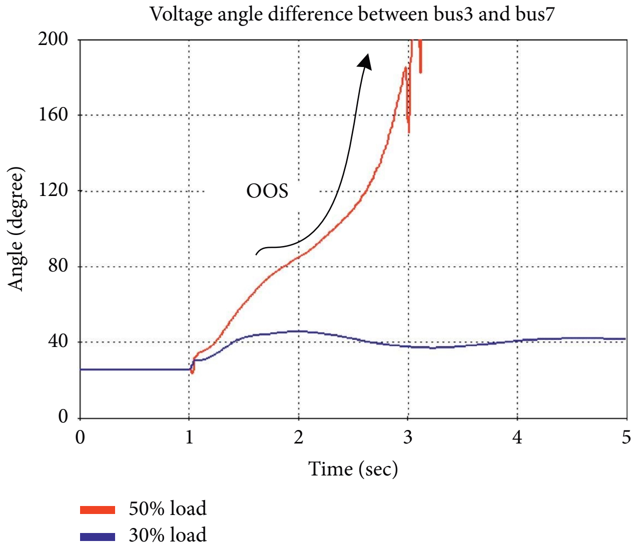


Figure 14: Voltage angle difference between bus 3 and bus 7 in cases of 30% and 50% loading increase.

Table 4: Two-area system eigenvalues.

Scenario	$\lambda = \sigma \pm j\omega$ (stable)	$\lambda = \sigma \pm j\omega$ (unstable)
Change inertia	$H = 6.175$ $-0.240 + 3.080j$	$H = 3.175$ $0.0704 + 17.59j$
Change load	30% variation $-0.5650 + 2.390j$	50% variation $0.188 + 37.57j$
Network change	AG fault $-0.2521 + 1.696j$	ABCG fault $0.0277 + 15.4j$

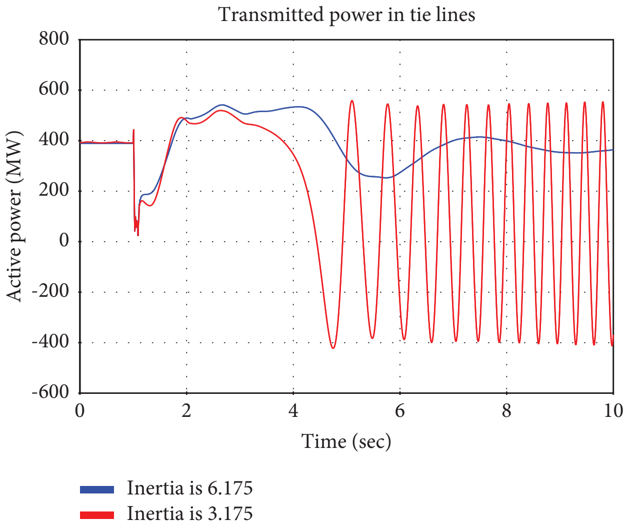


Figure 15: Effect of changing inertia on transmitted power.

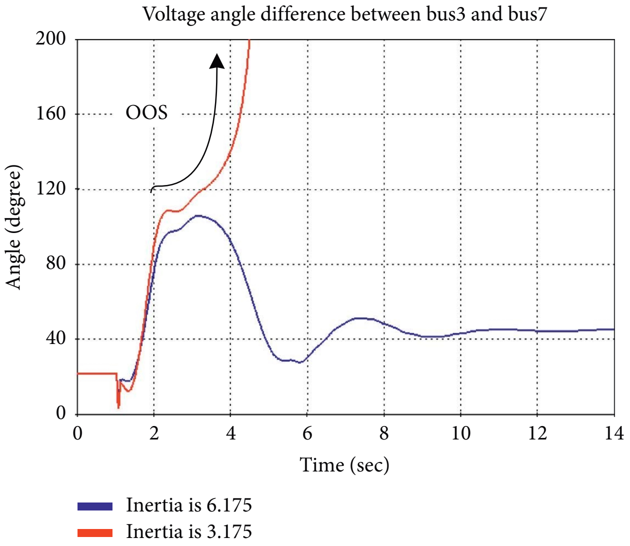


Figure 16: Effect of changing inertia on voltage angle difference.

the grid from this condition and its results. The results in this scenario have been verified by the eigenvalues calculated and tabulated in Table 4.

3.2.3. Scenario III: Change in Network. In this section, the changing network is represented by losing one of the lines between bus 4 and bus 6 due to a 3-phase fault and a line to ground fault and cleared at 1.07 sec, as shown in Figure 17.

The tie line power remains stable after removing the line to the ground fault. However, in the case of the 3-phase fault, caused out-of-step between area 1 and area 2 and bus voltage angle difference to cross 180°, as given in Figure 18, eigenvalue analysis has been carried out for both conditions, and Table 4 demonstrates unstable condition with positive real parts of the eigenvalues.

Disconnecting any electrical element from the network under any condition may expose it to the consequence of OOS. In such an unavoidable case, a protection system from OOS has to be built into the power system to ensure the security.

Table 4 presents positive real part eigenvalues in case load = 50%, $H = 3.175$, and loss line of the network, indicating OOS condition.

In all scenarios when an OOS condition exists because of any of the disturbances mentioned previously in the system, splitting action should be taken, by creating two independent islands to protect the system from collapsing.

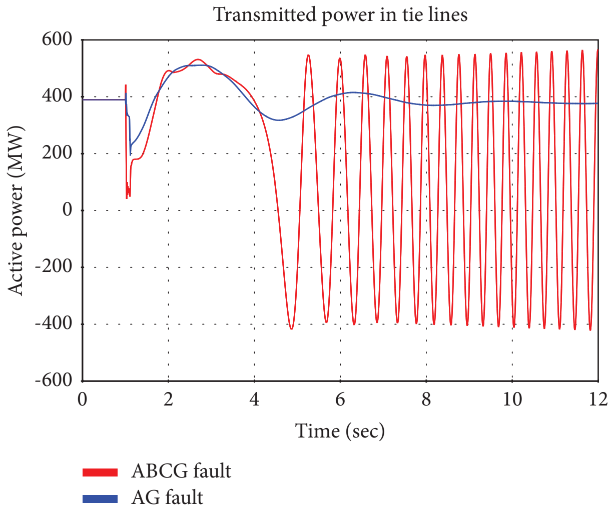


Figure 17: Effect of changing network configuration on transmitted power.

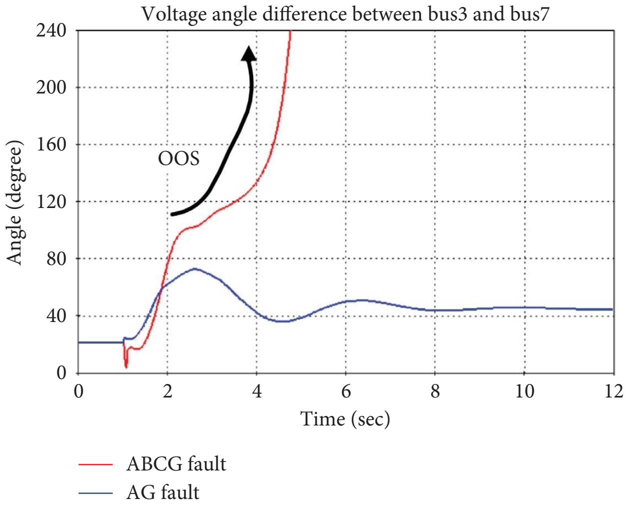


Figure 18: Effect of changing network configuration on voltage angle difference.

This solution has been analysed on the extensive system as follows using the proposed voltage angle difference method.

3.3. Case Study III: IEEE39 bus System. This section aims to analyze out-of-step oscillation in the IEEE39 bus system and identify the coherent group. The system has a 60 Hz frequency and consists of 10 generators, 12 transformers, and 34 transmission lines (Figure 19), and the detailed parameters are in [32]. Two scenarios have been simulated: first a three-phase fault and second a three-phase faults at two locations simultaneously in the system.

3.3.1. Scenario I. First, a fault was applied on line 21-22 at 1.0 sec and cleared by tripping the line at 1.12 sec. All generator's frequency has been measured, which is the best parameter that can be monitored to check if there is an oscillation in the system. Stable oscillations have been observed for this scenario as shown in Figure 20. Therefore, no OOS condition has been detected in the system. The eigenvalue calculation given in Table 5 verifies the stable condition.

3.3.2. Scenario II. In the second scenario along with the previous fault on line 21-22, one more fault on line 26–29 at 1 sec and has been applied and cleared by isolating that line at 1.1 sec. The response of the system in terms of generator frequency is shown in Figure 21 in three parts by grouping the coherent oscillations. It can be observed that the generators frequency is no longer stable leading to OOS condition in the system. Eigenvalue in Table 6 confirms the time-domain response.

3.3.3. Comparison of Scenarios. After finding out coherent generators, tie lines have to be identified; thus, in time-

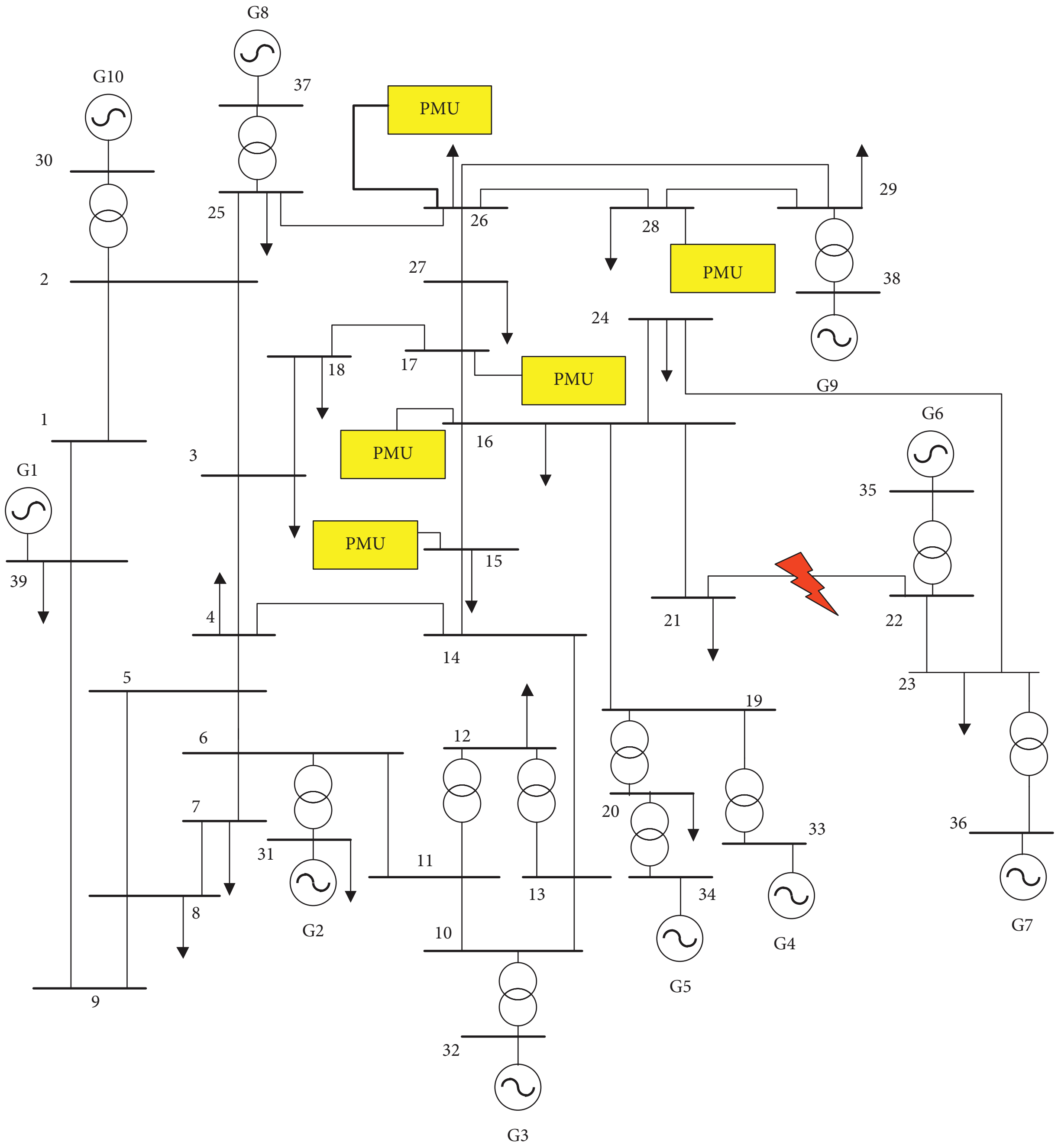


Figure 19: IEEE39-bus system.

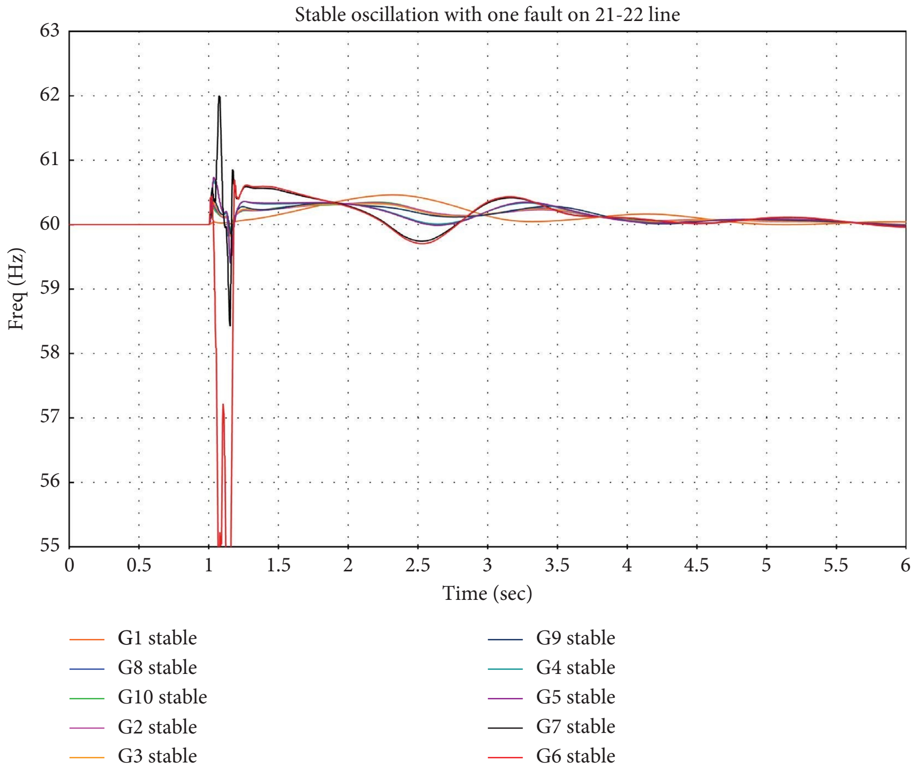


Figure 20: Stable generators' frequency.

Table 5: 39-bus system negative real part of eigenvalues.

Study case	$\lambda = \sigma \pm j\omega$ (stable)
39-bus system	Single fault $-0.3850 + 1.82j$

domain simulation, the oscillation frequency of system buses has been observed to monitor the coherency of buses as well. Along with that, the bus voltage angles difference across the lines connecting the coherent groups have been observed. The lines (15-16, 16-17, and 26–28) are going out of step, as shown in Figure 22.

Figure 22 compares both the scenarios transmitted active power and voltage angle difference of the three tie lines. In the case of a single fault, the power and angle difference for all the tie lines become stable after oscillating for some time. However, in the case of two simultaneous faults, the tie line bus voltage angle difference exceeds 180 degrees leading to OOS condition. Keeping the system operating in this condition will cause significant damage to the network and collapse of the system, so separating actions should be taken immediately when the voltage angle difference reaches 180° to maintain the security and reliability of the system. Thus, the three tie lines have been tripped with the developed algorithm, creating three stable islands with continuing the supply to the loads as in Figure 23.

This solution protects the system from collapsing and prevents any power outage for the consumers. Figure 24(b)

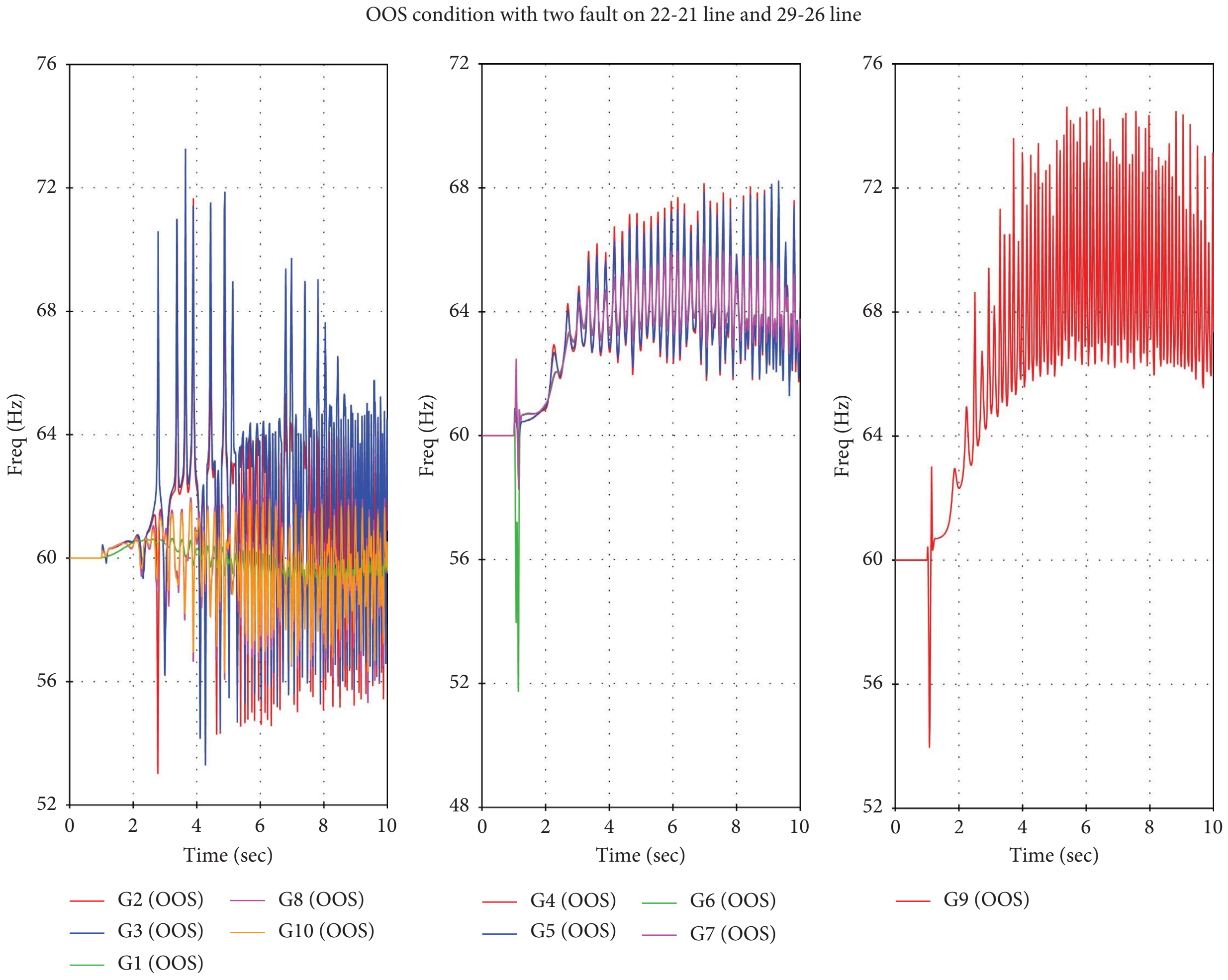


Figure 21: group of generator oscillates with the same frequency.

Table 6: 39-bus system eigenvalues (OOS).

Study case	$\lambda = \sigma \pm j\omega$ (unstable)
39-bus system	Two faults $0.1327 + 33.175j$

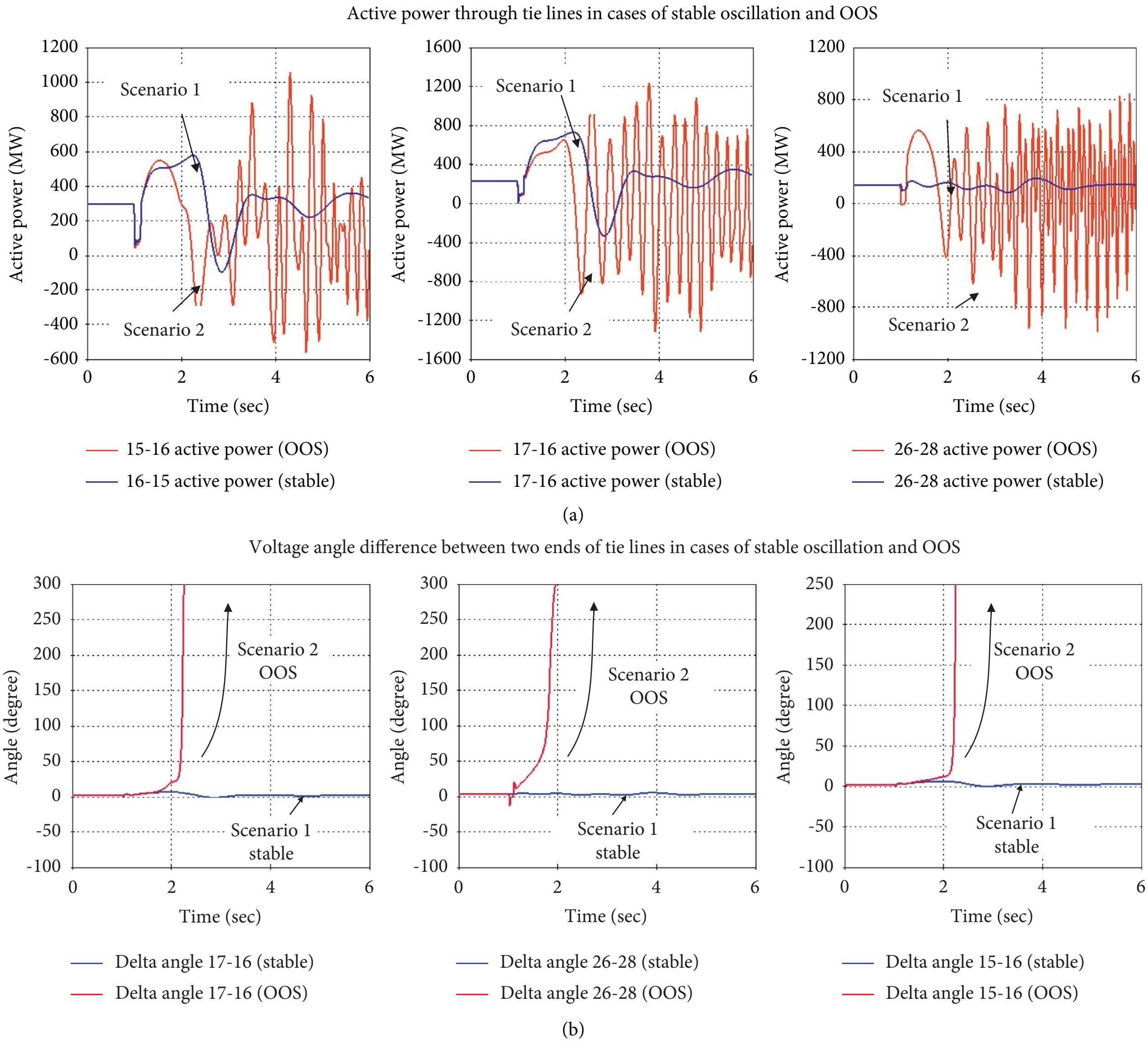


Figure 22: Comparison between stable and out-of-step conditions: (a) tie lines' active power and (b) voltage angle difference.

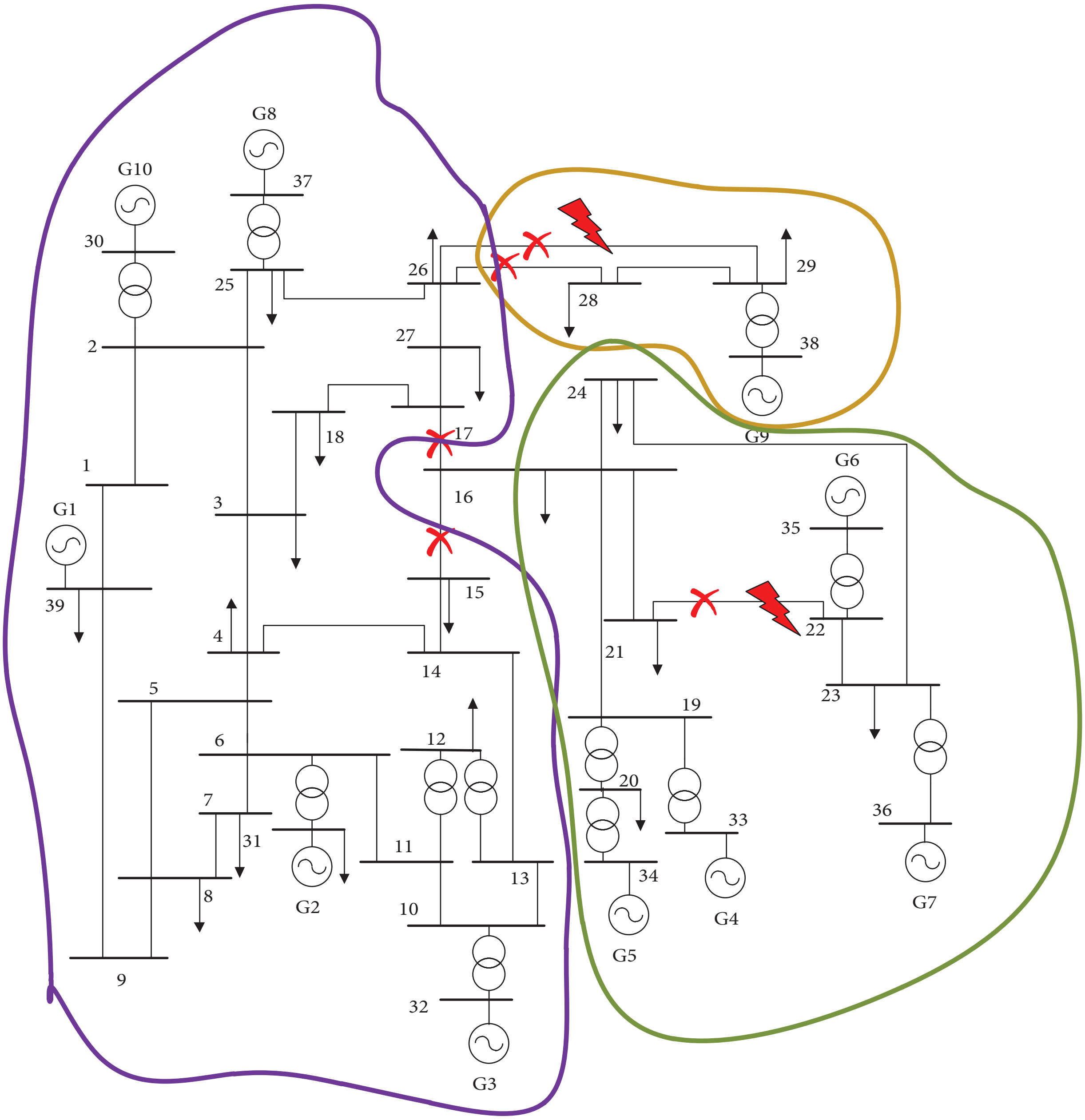


Figure 23: Separating the 39-bus system into three islands.

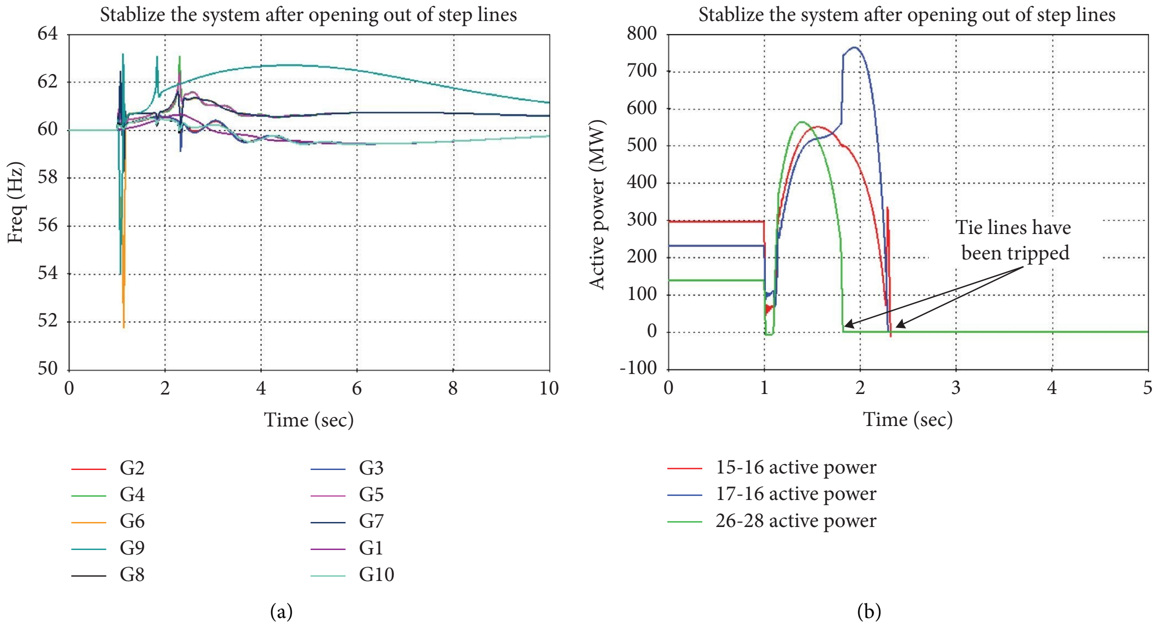


Figure 24: Stable islands after separation.

shows that, after opening, the three tie lines as the figure illustrate that no power flow through these lines. Consequently, the generators' frequency has stopped oscillating and going to be stable after a few second (Figure 24(a)).

4. Conclusions

Power swings in the power system are triggered due to faults and topological changes in the system. Power swing strigger oscillations in the system which can be stable or unstable. The unstable oscillations lead to a condition where a generator or two coherent areas oscillate against each other leading to a 180-degree phase angle separation named as out-of-step condition. The tripping of a generator or tie lines is essential for maintain the stability of system. A detailed investigation of OOS condition has been carried out in time domain simulation on three benchmark systems for generator mode and interarea mode, by simulating the systems in real-time simulator platform from OPAL-RT with HYPERSIM environment. The results of the article were compared with small signal stability analysis by eigenvalue technique.

Furthermore, a PMU-based OOS condition detection with bus voltage angle has been considered in this work which can be easily implemented in WAMS. This method is better than the conventional impedance-based detection of OOS condition which is prone to maloperation during power swings in the power system. Also, this method is independent of the network variations in the power system. The results from this article in the time domain corroborate with the analytical work considered in the article.

Data Availability

The data used to support the findings of the study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research work was entirely the product of the authors' initiatives.

References

[1] M. Jan-E-Alam, "A study on the presence of inter-area oscillation mode in Bangladesh power system network," *Journal of Electrical Engineering*, vol. 36, no. 2, pp. 16–21, 2011.

[2] G. R. Gajjar and S. Soman, "Power system oscillation modes identifications: guidelines for applying TLS-ESPRIT method," *International Journal of Emerging Electric Power Systems*, vol. 14, no. 1, pp. 57–66, 2013.

[3] Y. Chompoobutrgool and L. Vanfretti, "Identification of power system dominant inter-area oscillation paths," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2798–2807, 2013.

[4] A. Fathy and M. M. M. El-Arini, "Identification of coherent groups of generators based on fuzzy algorithm," in *Proceedings of the 14th International Middle East Power Systems Conference (MEPCON'10)*, pp. 19–21, Cairo University, Egypt, December 2010.

[5] J. T. Rao, B. R. Bhalja, M. v. Andreev, and O. P. Malik, "Synchrophasor assisted power swing detection scheme for wind integrated transmission network," *IEEE Transactions on Power Delivery*, vol. 37, no. 3, pp. 1952–1962, 2022.

[6] P. K. Bera and C. Isik, "Identification of stable and unstable power swings using pattern recognition," in *Proceedings of the 2021 IEEE Green Technologies Conference (GreenTech)*, pp. 286–291, Denver, CO, USA, April 2021.

[7] A. Haddadi, I. Kocar, U. Karaagac, H. Gras, and E. Farantatos, "Impact of wind generation on power swing protection," *IEEE*

Transactions on Power Delivery, vol. 34, no. 3, pp. 1118–1128, 2019.

[8] M. Amroune and T. Bouktir, "Effects of different parameters on power system transient stability studies," *Journal of Advanced Sciences & Applied Engineering*, vol. 1, no. 1, pp. 28–33, 2014.

[9] S. Eftekharnejad, V. Vittal, G. T. Heydt, B. Keel, and J. Loehr, "Impact of increased penetration of photovoltaic generation on power systems," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 893–901, 2013.

[10] S. Zhang and Y. Zhang, "Characteristic analysis and calculation of frequencies of voltages in out-of-step oscillation power system and a frequency-based out-of-step protection," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 205–214, 2019.

[11] M. Abedini, M. Davarpanah, M. Sanaye-Pasand, S. M. Hashemi, and R. Iravani, "Generator out-of-step prediction based on faster-than-real-time analysis: concepts and applications," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4563–4573, 2018.

[12] M. R. Nasab and H. Yaghobi, "A real-time out-of-step protection strategy based on instantaneous active power deviation," *IEEE Transactions on Power Delivery*, vol. 36, no. 6, pp. 3590–3600, 2021.

[13] S. Paudyal, G. Ramakrishna, and M. S. Sachdev, "Application of equal area criterion conditions in the time domain for out-of-step protection," *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 600–609, 2010.

[14] D. Kang and R. Gokaraju, "A new method for blocking third-zone distance relays during stable power swings," *IEEE Transactions on Power Delivery*, vol. 31, no. 4, pp. 1836–1843, 2016.

[15] S. Chatterjee, "Identification of faults during power swing: a PMU based scheme," in *Proceedings of the 2019 8th International Conference on Power Systems (ICPS)*, pp. 1–5, Jaipur, India, December 2019.

[16] J. R. A. K. Yellajosula, Y. Wei, M. Grebla, S. Paudyal, and B. A. Mork, "Online detection of power swing using approximate stability boundaries," *IEEE Transactions on Power Delivery*, vol. 35, no. 3, pp. 1220–1229, 2020.

[17] M. R. Salimian and M. R. Aghamohammadi, "Intelligent out of step predictor for inter area oscillations using speed-acceleration criterion as a time matching for controlled islanding," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2488–2497, 2018.

[18] L. Zacharia, M. Asprou, and E. Kyriakides, "Wide area control of governors and power system stabilizers with an adaptive tuning of coordination signals," *IEEE Open Access Journal of Power and Energy*, vol. 7, no. 1, pp. 70–81, 2020.

[19] R. T. Elliott, P. Arabshahi, and D. S. Kirschen, "A generalized PSS architecture for balancing transient and small-signal response," *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 1446–1456, 2020.

[20] I. Zenelis, X. Wang, and I. Kamwa, "Online PMU-based wide-area damping control for multiple inter-area modes," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5451–5461, 2020.

[21] X. Zhou, S. Cheng, X. Wu, and X. Rao, "Influence of photovoltaic power plants based on VSG technology on low frequency oscillation of multi-machine power systems," *IEEE Transactions on Power Delivery*, vol. 37, 2022.

[22] J. Qi, Q. Wu, Y. Zhang, G. Weng, and D. Zhou, "Unified residue method for design of compact wide-area damping controller based on power system stabilizer," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 2, pp. 367–376, 2020.

[23] N. Hatziargyriou, J. Milanovic, C. Rahmann et al., "Definition and classification of power system stability - revisited & extended," *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 3271–3281, 2021.

[24] M. Klein, G. J. Rogers, and P. Kundur, "A fundamental study of inter-area oscillations in power systems," *IEEE Transactions on Power Systems*, vol. 6, no. 3, pp. 914–921, 1991.

[25] D. A. Tziouvaras and D. Hou, "Out-of-step protection fundamentals and advancements," in *Proceedings of the 57th Annual Conference for Protective Relay Engineers*, pp. 282–307, College Station, TX, USA, April 2004.

[26] S. Zhang and Y. Zhang, "A novel out-of-step splitting protection based on the wide area information," *IEEE Transactions on Smart Grid*, vol. 8, no. 1, pp. 41–51, 2016.

[27] P. Kundur, *Power System Stability and Control*, McGraw-Hill, New York, NY, USA, 1994.

[28] A. Ahlawat, A. Goyal, S. K. Mishra, and S. T. Nagarajan, "A laboratory setup for synchrophasor applications," in *Proceedings of the 2020 IEEE 17th India Council International Conference*, New Delhi, India, December 2020.

[29] M. McDonald and D. Tziouvaras, "Power swing and out-of-step considerations on transmission lines," *IEEE PSRC WG D6, A report to the Power System Relaying Committee of the IEEE Power Engineering Society*, 2005.

[30] S. K. Yadav and S. T. Nagarajan, "Study on impact of power system inertial stability by renewable energy sources," in *Proceedings of the 2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP)*, pp. 1–6, Hyderabad, India, July 2022.

[31] H. Saadat, *Power System Analysis*, McGraw-Hill, New York, NY, USA, 1999.

[32] M. Pai, *Energy Function Analysis for Power System Stability*, Springer, New York, NY, USA, 1989.

ANFIS (Adaptive Neuro-Fuzzy Inference System) based on Microgrid's Reliability and Availability

Geeta Yadav
Department of Electrical and Electronics Engineering,
Manav Rachna International Institute of Research and Studies,
Faridabad, India
https://orcid.org/0000-0002-3419-0231

Dheeraj Joshi
Department of Electrical Engineering,
Delhi Technological University,
Delhi, India
joshidheeraj@dce.ac.in

Leena G
Department of Electrical and Electronics Engineering,
Manav Rachna International Institute of Research and Studies,
Faridabad, India
leenag.fet@mriu.edu.in

M K Soni
IIMT Group of Colleges,
Noida, India
dr_mksoni@hotmail.com

2022 IEEE 10th Power India International Conference (PIICON) | 978-1-6654-5930-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/PIICON56320.2022.10045275

Abstract— **The applicability of machine learning algorithms used to solve microgrid optimization is investigated in this paper. This paper's main objective is to build a microgrid model to achieve maximum reliability and availability using renewable resources that cater to users' needs with different demands and supplies. The model generated from the adaptive neuro-fuzzy inference system (ANFIS) is used to get the optimum reliability and availability strategy to achieve the user expectations and needs of future microgrids. The ANFIS model is trained with different data sets from Markov modeling. The dataset is divided into three sections, 40% of the data is used to train the model, testing is performed with 40%, and the last 20% of the information is checked. Implementation results show that ANFIS models emulate Markov modeling methods and artificial neural networks model and enhance reliability and availability. Additionally, the ANFIS model is better than the Markov and artificial neural networks models regarding the impact of individual failure rate optimization from sub-models. Comparison is shown with Markov modeling, genetic algorithm, artificial neural networks, and fuzzy system. The optimization toolbox Matlab is used for ANFIS.**

Keywords—ANFIS, Markov Modeling, Microgrid, submodels, artificial neural networks, fuzzy systems.

I. INTRODUCTION

Multiple cases of power crises or power failures occur worldwide due to various reasons affecting the services of power utilities and reducing the reliability indices. The power services or multiple interruptions also affect the consumers and economic eruption [1]. Every state or country checks the reliability matrices to assess the consumer's satisfaction or complaints. As per statistics, the fall in the electricity demand-supply increased by 1.88 billion units or 1.6% during April in India. Similarly, there was a shortage of 623 million units in march, and the primary reason was coal shortage, although multiple actions have been taken to increase coal production by 27.2% [2].

Everyone needs a permanent solution to this problem; the answer is Microgrid. Natural disasters, climatic change, animal attacks or tree drops, lightning effect, or any shot of attacks and load management highly impact electricity management [3]. Microgrid handles the distribution of energy locally as well as in islanded mode. The microgrid is economically efficient and has clean penetration of renewable resources [4]. Microgrids face multiple challenges as component failures need to be considered, and every component fails at a different rate and another state. Reliability and availability assessment is highly required in the design and operational phase [5].

The reliability is assessed by considering the model as a system, and every component's failure is considered along with maintenance. The performance indices considered are the mean time to failure and mean time to repair for microgrid reliability and availability design. The fuzzy reliability model is regarded as the uncertainty present in failure rate is of deep concern, and artificial neural networks learn themselves and perform the following action correctly [6]. In ANFIS, both scenarios are considered. ANFIS is Sugeno fuzzy model [8]. ANFIS provides generalization capabilities and shares the robustness of results. ANFIS takes crisp input and generates membership functions and fuzzy rules. The output is also brittle, even in complex structures and gradient learning [9].

The Sugeno model computes the ANFIS-based microgrid reliability and availability, performed for large states and considered initials. Section 2 explains the research method used, and Section 3 contains the results and discussion. Section 4 has a conclusion.

II. RESEARCH METHOD

A. Overview of ANFIS Microgrid Model

There are three steps in implementing the ANFIS model training, testing, and checking. Still, before starting the system, the prepared dataset should be ready, obtained from Markov modeling, genetic algorithm, and artificial neural networks [10]. These datasets are obtained from different values of failure rates, repair rates, and coverage factors. The training should be completed once the epochs reach the threshold values [10]. The datasets are prepared in offline mode using Markov modeling [11]. The Matlab code is used to get the reliability and availability values based on different failure and repair rates. These data are divided into three parts. The first 40% of datasets are used for training the model. As shown in Fig. 1, the microgrid model is segregated into multiple components like PV panels, Converters, transformers, load, etc. The overall reliability of the microgrid is assessed, and the mean time to failure is calculated. These performance indices are optimized using the ANFIS model.

978-1-6654-5930-3/22/$31.00 ©2022 IEEE

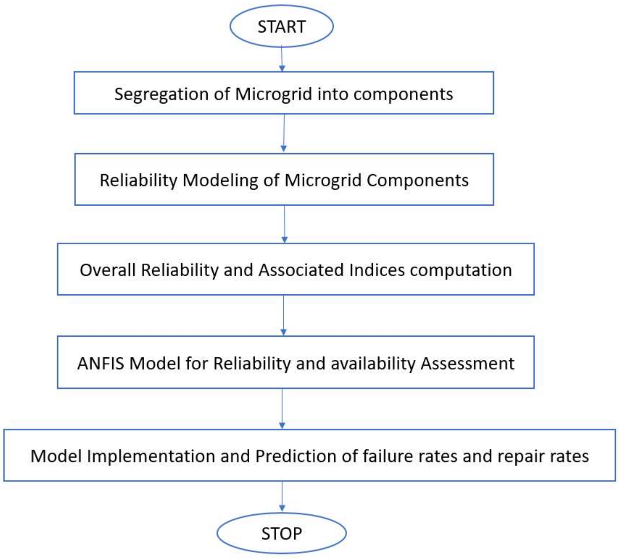

START
Segregation of Microgrid into components
Reliability Modeling of Microgrid Components
Overall Reliability and Associated Indices computation
ANFIS Model for Reliability and availability Assessment
Model Implementation and Prediction of failure rates and repair rates
STOP

Fig. 1 ANFIS framework for the proposed model

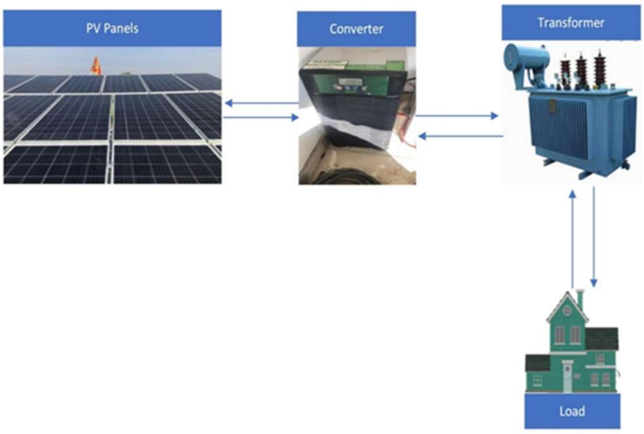


Fig. 2 Microgrid Proposed model

The reliability and Availability assessment performed in the Fig 1 is based on the block diagram of the Microgrid in Fig 2 and elaborated in the flow chart in Fig 2, where the dataset gathered before training is already available to us using applied reliability and availability assessment techniques discussed above [11]. The PV panels are connected with converters, then with transformer and load. The failure rate and repair rate of every component is considered.

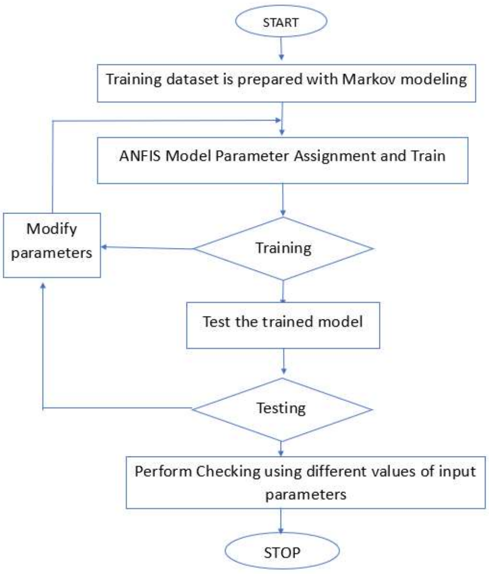

START
Training dataset is prepared with Markov modeling
ANFIS Model Parameter Assignment and Train
Modify parameters
Training
Test the trained model
Testing
Perform Checking using different values of input parameters
STOP

Fig. 3 Proposed method flow chart

B. *Training dataset preparation using Markov Modeling:*

The training dataset is prepared from the data collected from Markov modeling. The reliability computed versus time from Markov modeling is shown below in Fig 4.

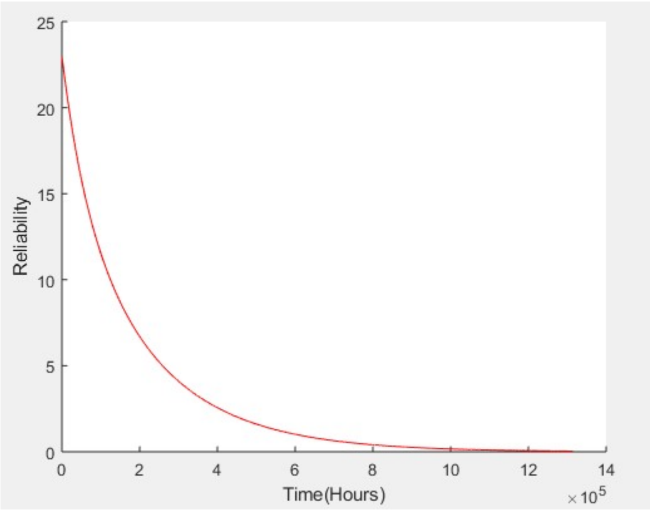


Fig .4 Reliability concerning the time

Markov reliability model is computed from multiple states, which change from state i to state j in stipulate time [12]. The state transition matrix is evaluated based on transition among states, as shown in Fig 5.

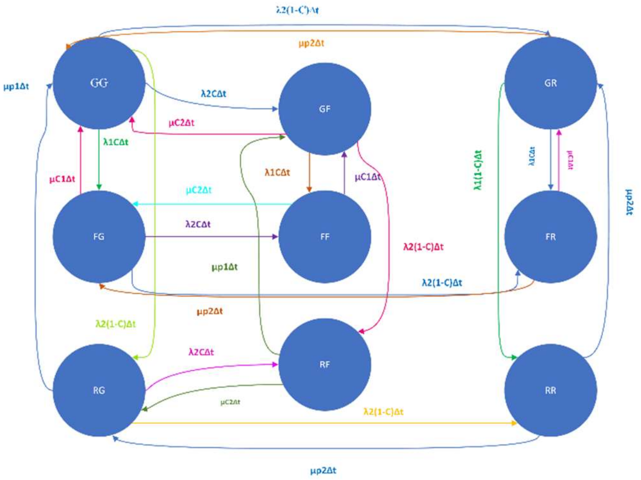


Fig. 5 Transition diagram of the proposed model

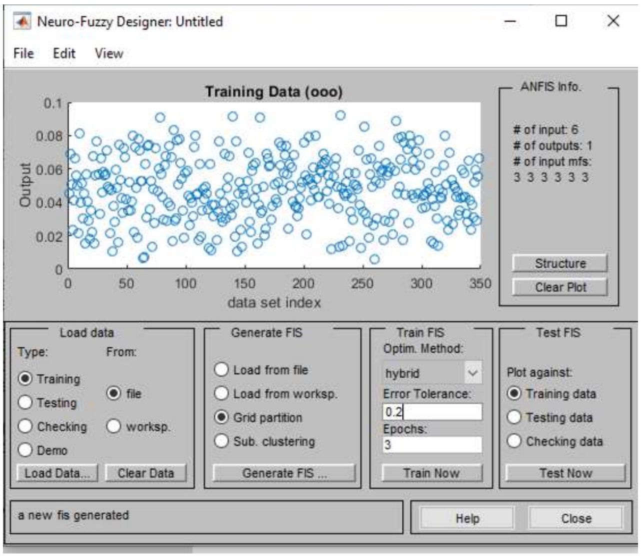
The state transition matrix is formed from this state transition diagram. P is denoted as the probability of a microgrid where poo is the initial state, and different shapes are added later. Where n is the total number of components and m is the total number of states. The matrix is of n X m dimensions.

$$P = \begin{bmatrix} p_{00} & \cdots & p_{0n} \\ \vdots & \ddots & \vdots \\ p_{n0} & \cdots & p_{nm} \end{bmatrix} \quad (1)$$

The availability is assessed using the below by taking the differential of the above probability functions, and steady states are considered zero. The adding remaining states to computed the overall availability of the microgrid using different differential equations [13].

III. Results and Discussions

The ANFIS data has six inputs, including failure and repair rates, and two outputs, reliability and availability. The input and output data are normalized in the state transition model. Select the training as the failure rates of PV panels, Converters, transformers, and repair rates for each component and upload the training matrix using an M-file in Matlab. As shown in Fig 5, the dataset is in circles in the ANFIS model. The fuzzy memberships are created using the Sugeno model. The inputs are considered low, medium, and high.


Neuro-Fuzzy Designer: Untitled

ANFIS Info.

# of input: 6
# of outputs: 1
# of input mfs:
3 3 3 3 3 3

Load data — Type: ● Training ○ Testing ○ Checking ○ Demo — From: ● file ○ worksp.

Generate FIS — ○ Load from file ○ Load from worksp. ● Grid partition ○ Sub. clustering

Train FIS — Optim. Method: hybrid — Error Tolerance: 0.2 — Epochs: 3

Test FIS — Plot against: ● Training data ○ Testing data ○ Checking data

a new fis generated

Fig. 6 ANFIS training dataset plotted

The ANFIS model is quite complex, as shown in Fig 7. 729 rules are created along with the membership function.

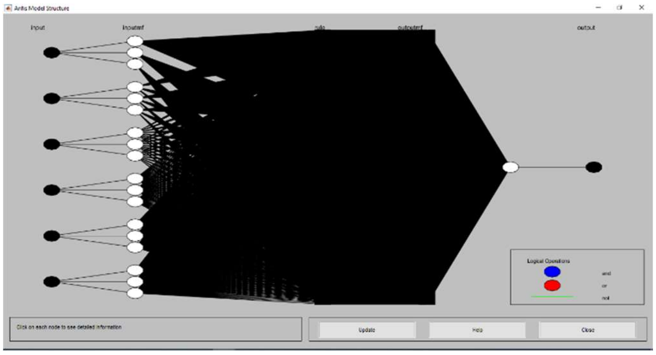
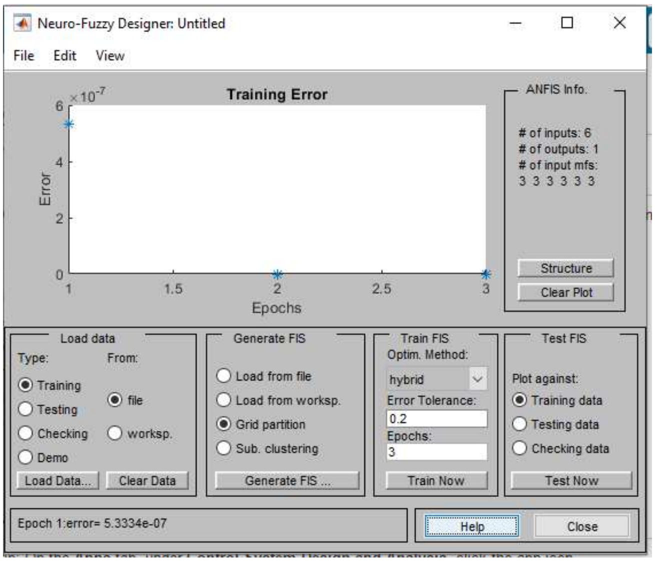


Fig. 7 ANFIS results

The ANFIS model is simulated for 350 values of the dataset, and the error received after training of 40% of data is 0.00000053334, which is significantly less. It means the error received after computation is negligible. The system is trained thoroughly, and there are 729 rules created. The model is taught multiple times with different datasets to get the minimum error, and the model should predict accurate results [15]. The error is shown in Fig. 8. The ANFIG information is also present in the command window of Matlab. The total number of nodes is 1503, and the linear parameters are 729. The number of non-linear parameters is 54. Several training data pairs are around 350, and the total number of fuzzy rules generated is 729. The different parameter is changed, and the same will be reflected in the command prompt. The epoch time or the number of iterations is increased or decreased based on the results obtained. The viability of the results is also mentioned in the command prompt as a training error, and the testing error is discussed over there, as presented in Fig. 11.


Neuro-Fuzzy Designer: Untitled

ANFIS Info.

# of inputs: 6
# of outputs: 1
# of input mfs:
3 3 3 3 3 3

Load data — Type: ● Training ○ Testing ○ Checking ○ Demo — From: ● file ○ worksp.

Generate FIS — ○ Load from file ○ Load from worksp. ● Grid partition ○ Sub. clustering

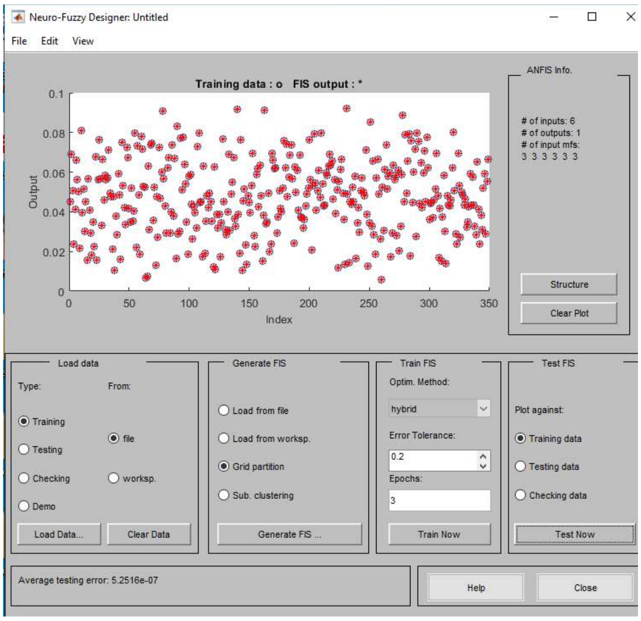
Train FIS — Optim. Method: hybrid — Error Tolerance: 0.2 — Epochs: 3

Test FIS — Plot against: ● Training data ○ Testing data ○ Checking data

Epoch 1:error= 5.3334e-07

Fig. 8 ANFIS results - Epoch error

The Average testing error is also computed after training which is shown in Fig 9 as: 0.000000525



# of inputs: 6
# of outputs: 1
# of input mfs:
3 3 3 3 3 3

Type: ● Training ○ Testing ○ Checking ○ Demo

From: ● file ○ worksp.

○ Load from file
○ Load from worksp.
● Grid partition
○ Sub. clustering

Optim. Method: hybrid

Error Tolerance: 0.2

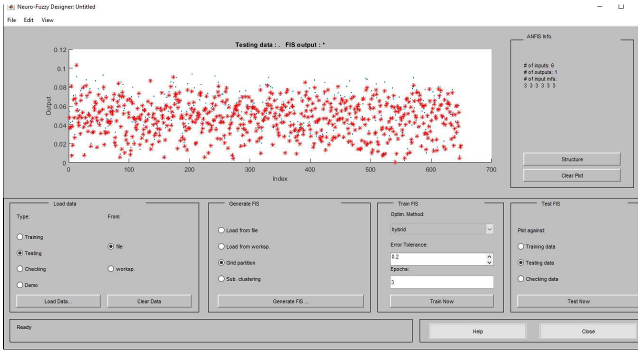
Epochs: 3

Plot against: ● Training data ○ Testing data ○ Checking data

Average testing error: 5.2516e-07

Fig. 9 ANFIS results

After recursive training of the system, the testing is performed with another dataset of around 40%. The red points are the trained, and the blue is the actual results, as depicted in Fig 10.



# of inputs: 6
# of outputs: 1
# of input mfs:
3 3 3 3 3 3

Type: ○ Training ● Testing ○ Checking ○ Demo

From: ● file ○ worksp.

○ Load from file
○ Load from worksp.
● Grid partition
○ Sub. clustering

Optim. Method: hybrid

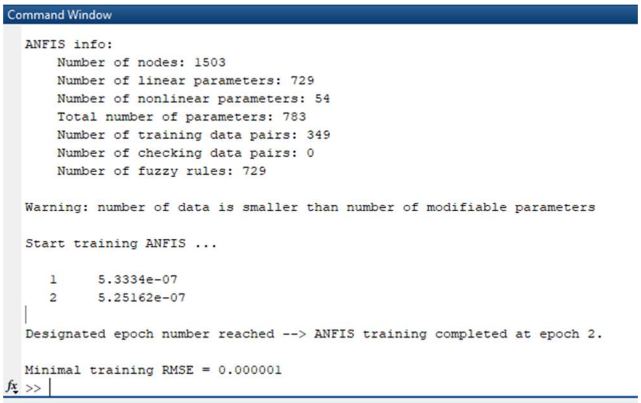
Error Tolerance: 0.2

Epochs: 3

Plot against: ○ Training data ● Testing data ○ Checking data

Ready

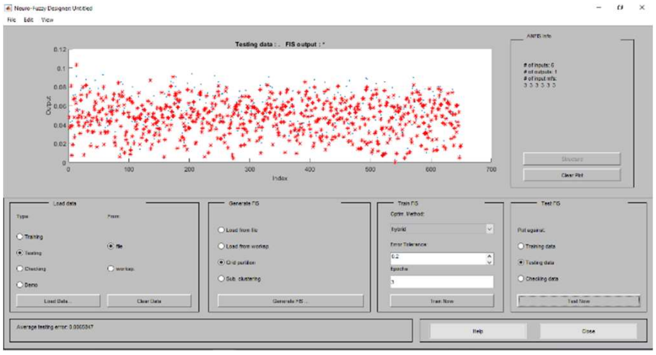
Fig. 10 ANFIS results



```
ANFIS info:
    Number of nodes: 1503
    Number of linear parameters: 729
    Number of nonlinear parameters: 54
    Total number of parameters: 783
    Number of training data pairs: 349
    Number of checking data pairs: 0
    Number of fuzzy rules: 729

Warning: number of data is smaller than number of modifiable parameters

Start training ANFIS ...

   1      5.3334e-07
   2      5.25162e-07
|
Designated epoch number reached --> ANFIS training completed at epoch 2.

Minimal training RMSE = 0.000001
fx >>
```

Fig. 11 ANFIS results

The ANFIS result after training and testing is shown below in Fig 12. The red points are training, and testing points, but the blue points are the actual values that are very near to trained values.



# of inputs: 6
# of outputs: 1
# of input mfs:
3 3 3 3 3 3

Type: ○ Training ● Testing ○ Checking ○ Demo

From: ● file ○ worksp.

○ Load from file
○ Load from worksp.
● Grid partition
○ Sub. clustering

Optim. Method: hybrid

Error Tolerance: 0.2

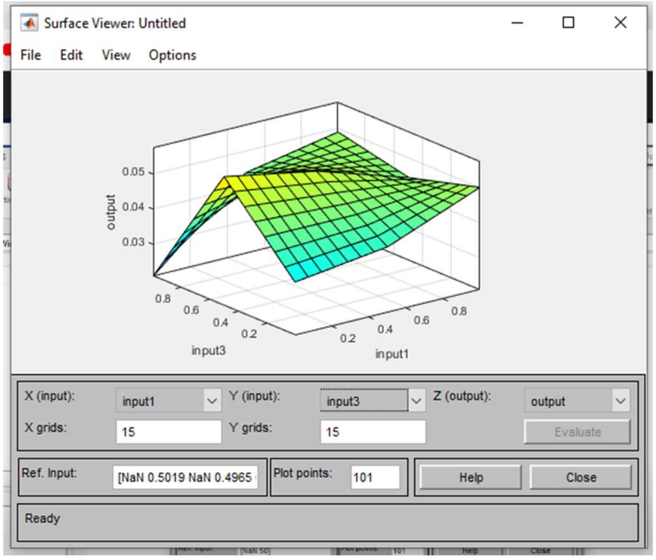
Epochs: 3

Plot against: ○ Training data ● Testing data ○ Checking data

Average testing error: 0.0000347

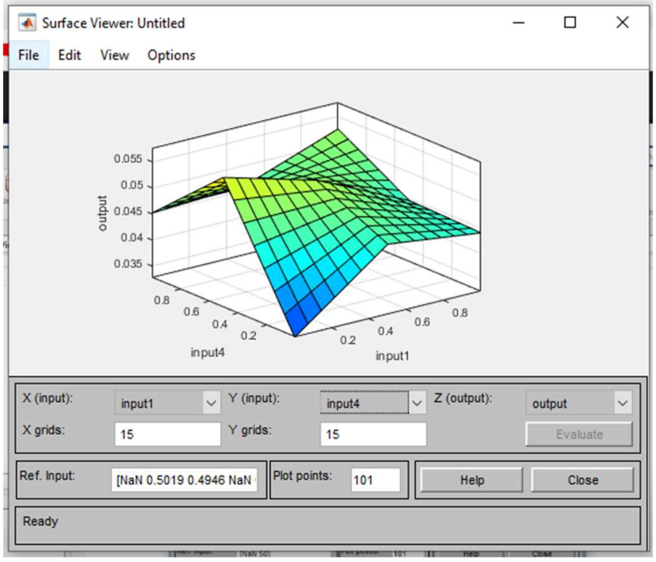
Fig. 12 ANFIS results

The overall reliability depends on the multiple failure rates and repair rates. So, the reliability relationship with every parameter can also be done through ANFIS modeling. Here the failure rates are input one and input three, plotted with output where other inputs are taken as negligible, as in Fig 13, Fig. 14, and Fig. 15.



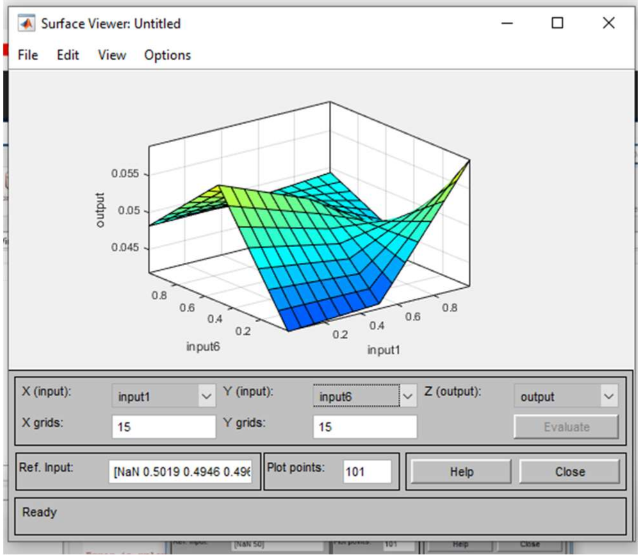
X (input): input1
Y (input): input3
Z (output): output
X grids: 15
Y grids: 15
Ref. Input: [NaN 0.5019 NaN 0.4965
Plot points: 101
Ready

Fig. 13 ANFIS results



X (input): input1
Y (input): input4
Z (output): output
X grids: 15
Y grids: 15
Ref. Input: [NaN 0.5019 0.4946 NaN
Plot points: 101
Ready

Fig. 14 ANFIS results



X (input): input1 Y (input): input6 Z (output): output

X grids: 15 Y grids: 15

Ref. Input: [NaN 0.5019 0.4946 0.49€ Plot points: 101

Ready

Fig. 15 ANFIS results

The comparison of different methods is shown below in Table 1. The deviation from the actual result is shown as an error in Table 1.

TABLE I. ERROR WHILE EVALUATING RELIABILITY

	Error while evaluating Reliability		
	Markov Modeling	***Artificial Neural Networks***	***ANFIS***
Error	0.03	8.4e-11	0.0000005

IV. CONCLUSION

Microgrids' reliability and availability assessment can predict the remnant life of any subcomponent. The statistics study of failures that occurred and evaluation of reliability using the Matlab optimization toolbox are conducted. Also, failure rates, repair rates, and errors are optimized for reliability. The ANFIS model was executed using different failure and repair rates of the subcomponents and general illations. The design of other components is predicted with precise failure rate values, and repair rates are expected in the ANN model with minimum error.

As a future enhancement, more examples of the dataset can be used to precisely predict the component's best possible failure rate and repair rate with different configurations and the addition of communication devices like wifi, Zigbee, etc.

REFERENCES

[1] B. Amam and Z. Li., "Electrical power crisis solution by the developing renewable energy based power generation expansion," Energy Reports, vol. 6, pp. 480-490, 2020. Doi: 10.1016/j.egyr.2019.11.106.

[2] P. Balasubramanian and P. Karthickumar, "Indian energy crisis - A sustainable solution," *IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012)*, 2012, pp. 411-415.

[3] S. Leonori, A. Rizzi, M. Paschero and F. M. F. Mascioli, "Microgrid Energy Management by ANFIS Supported by an ESN Based Prediction Algorithm," *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1-8, doi: 10.1109/IJCNN.2018.8489018.

[4] G. Yadav, D. Joshi, L. Gopinath, and M K Soni, "Reliability and Availability Optimization of Smart Microgrid Using Specific Configuration of Renewable Resources and Considering Subcomponent Faults", *Energies,* vol 15, issue no. 16: 5994. doi: https://doi.org/10.3390/en15165994

[5] S. Y. Shirmardi, M. Joorabian, H. Barati, "Flexible-reliable operation of green microgrids including sources and energy storage-based active loads considering ANFIS-based data forecasting method", Electric Power Systems Research, Vol 210, 2022, doi: https://doi.org/10.1016/j.epsr.2022.108107

[6] P. Binh and T. Q. D. Khoa, "Application of Fuzzy Markov in calculating reliability of power systems," *2006 IEEE/PES Transmission & Distribution Conference and Exposition: Latin America, 2006,* pp. 1-4, doi: 10.1109/TDCLA.2006.311384.

[7] A. Kumar and P. Kumar, "Application of Markov process/mathematical modelling in analysing communication system reliability" , *International Journal of Quality and Reliability Management*, vol. 37, no. 2, 2020.

[8] M. Manohar, E. Koley and S. Ghosh, "A wavelet and ANFIS based reliable protection technique for Microgrid," *2019 8th International Conference on Power Systems (ICPS)*, 2019, pp. 1-6, doi: 10.1109/ICPS48983.2019.9067617.

[9] Y. K. Semero, D. Zheng and J. Zhang, "A PSO-ANFIS based Hybrid Approach for Short Term PV Power Prediction in Microgrids", Electric Power Components and Systems, vol 46, number 1, pp 95-103, 2018, doi: https://doi.org/10.1080/15325008.2018.1433733

[10] A. Abdelsamad and D. Lubkeman, "Reliability Analysis for a Hybrid Microgrid based on Chronological Monte Carlo Simulation with Markov Switching Modeling," *2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2019, pp. 1-5, doi: 10.1109/ISGT.2019.8791611.

[11] L. Wang and S. Pang, "An Implementation of the Adaptive Neuro-Fuzzy Inference System (ANFIS) for Odor Source Localization," *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 4551-4558, doi: 10.1109/IROS45743.2020.9341688.

[12] A. Kumar, P. Kumar, "Application of Markov process/mathematical modelling in analysing communication system reliability," *International Journal of Quality and Reliability Management*, vol. 37, no. 2, 2020.

[13] D. Cevasco, S. Koukoura, A. J. Kolios, "Reliability, availability, maintainability data review for the identification of trends in offshore wind energy applications", *Renewable and Sustainable Energy Reviews*, vol. 136, 2021.

[14] S. K. Dubey, B. Jasra, "Reliability assessment of component based software systems using fuzzy and ANFIS techniques", *Int J Syst Assur Eng Manag* , pp. 1319–1326, 2017. doi: https://doi.org/10.1007/s13198-017-0602-z.

Hindawi
Journal of Electrical and Computer Engineering
Volume 2023, Article ID 8669796, 14 pages
https://doi.org/10.1155/2023/8669796

Research Article

Application of Fuzzy-RBF-CNN Ensemble Model for Short-Term Load Forecasting

Mohini Yadav,[1] Majid Jamil,[1] Mohammad Rizwan,[2] and Richa Kapoor[3]

[1] *Department of Electrical Engineering, Jamia Millia Islamia, New Delhi 110025, India*
[2] *Department of Electrical Engineering, Delhi Technological University, New Delhi 110042, India*
[3] *Department of Electrical and Electronics, Hindustan College of Science and Technology, Mathura 281122, India*

Correspondence should be addressed to Mohini Yadav; mohiniyadav565@gmail.com

Received 19 October 2022; Revised 5 February 2023; Accepted 14 February 2023; Published 3 March 2023

Academic Editor: B. Rajanarayan Prusty

Accurate load forecasting (LF) plays an important role in the operation and decision-making process of the power grid. Although the stochastic and nonlinear behavior of loads is highly dependent on consumer energy requirements, that demands a high level of accuracy in LF. In spite of several research studies being performed in this field, accurate load forecasting remains an important consideration. In this article, the design of a hybrid short-term load forecasting model (STLF) is proposed. This work combines the features of an artificial neural network (ANN), ensemble forecasting, and a deep learning network. RBFNNs and CNNs are trained in two phases using the functional link artificial neural network (FLANN) optimization method with a deep learning structure. The predictions made from RBFNNs have been computed and produced as the forecast of each activated cluster. This framework is known as fuzzy-RBFNN. This proposed framework is outlined to anticipate one-week ahead load demand on an hourly basis, and its accuracy is determined using two case studies, i.e., Hellenic and Cretan power systems. Its results are validated while comparing with four benchmark models like multiple linear regression (MLR), support vector machine (SVM), ML-SVM, and fuzzy-RBFNN in terms of accuracy. To demonstrate the performance of RBF-CNN, SVMs replace the RBF-CNN regressor, and this model is identified as an ML-SVM having 3 layers.

1. Introduction

Load forecasting [1] is a technique to predict the future load demands to attain equilibrium between the energy supply and consumption while considering different variables such as historical loads and weather conditions (temperature, pressure, humidity, etc.). With time series variations of nonlinear loads and different seasonal conditions, it is difficult to achieve short-term load forecasts with accuracy as the main consideration in different economical factors [2].

Accurate load forecasting will reduce the cost of electricity generation and improve trading advantages [3], especially during peak hours. Load forecasting shows the significance of a continuous demand rise, which leads to two prediction terms: overestimated and underestimated demands. This overestimated prediction incurs extra charges on the production cost of generation to maintain the large storage reserves as a backup supply. On the other hand, the underestimation affects the demand response. Accurate load forecasting plays a vital role in an industrial sector, a deregulated economy, and the coordination of the functioning of the entire power network, especially in smart grids [4–11]. Different energy companies try to calculate the forecast demand [12] more precisely and efficiently using an artificial neural network (ANN). As per the Mckinsey Global Institute, artificial intelligence (AI) should be considered for the prediction of power demand and supply [13]. As an example, the UK-based National Grid and the Google-owned team "DeepMind" [14, 15] enable the future demand forecast using smart meters and AI techniques. Furthermore, retailers also calculate the energy prices based on the forecast. With the new exposure to smart metering, many researchers have been exploring STLF in the residential sector [16–19].

Previously, the studies were based on manual methods that rely on particular datasets, which is not sustainable for utilities as the complexity of the network increases. For STLF, the researchers have used different regression methods [20] and presented exponential stabilities in mixed mode-dependent time delays [21], neural networks [22], clustering approaches [23], and support vector machine-based methods [24]. In the last years, researchers also used the Kalman filtering approach in forecasting applications and discussed the reduced order filtering approaches [25, 26]. All these methods are simple, but they are unable to reach to higher accuracy level when compared with artificial intelligence techniques. The authors [27] have presented the switched delay PSO technique for STLF.

Better performance has been achieved with the integration of different computational methods. Some references are briefly explained here. A multitask regression approach is used to predict the load recorded from residential and industrial smart meters [28]. Feature selection is done using additive models [29]. Recursive support vector regression (R-SVR) is integrated with empirical mode decomposition (EMD), where EMD is categorized as principal and behavioral, and then R-SVR is employed to provide prediction [30]. RBFNN and neural-based fuzzy models are utilized to tune their parameters using a penalty function [31]. To overcome the load forecasting problem, a fuzzy neural network is developed using a bilevel optimization algorithm [32].

It seems [33–35] that RBFNN presents an incredible response for the evaluation of demand forecasting. Although its prediction is based on the width of the radial basis function in the hidden layer, in the case of highly complicated problems like load forecasting, it deals with the "curse of dimensionality," as usual for every kernel approach. To ease this drawback, a composite RBF is accompanied by a deep learning (DL) approach. DL includes an automation process with a decision-making feature [36]. DL has a more complex architecture than other neural network architectures, including more layers and mathematical formulation. CNN used in this paper is one of the techniques of DL. DL serves as a great application in load forecasting [37]. Long-short-term memory recurrent neural network (LSTM-RNN) [38] and 3-filter CNN [39] are showing promising results in this field. CNN selects the input feature when the input is transformed into a graphical representation [40]. The input is accumulated into clusters by applying the k-means technique [41], subsequently, CNN is used for prediction purposes in each cluster [42]. The authors [43] have trained the neural network with a deep learning process, including convolution and pooling techniques, to extract the feature maps and forecast the load demand data for days ahead.

The main contribution of this study is to generate ensemble predictions from multiple local regressors, and this regression variable activates the forecast process using the data clustering method to assign the input to different clusters. The presented work defines its novelty that comes with an introduction of:

(1) New technique based on the fuzzy clustering approach, which clusters the input vector to generate ensemble predictions.

(2) Creative neural network architecture consists of RBF, convolution, and pooling in a fully connected two-layer network. This is termed as fuzzy-RBF-CNN. This proposed approach divides the input dataset into input subsets, which are used to develop the ensemble of RBF-CNN regressors. Each of these regressors is trained in two phases using the functional link artificial neural network (FLANN) optimization method with a deep learning framework.

(3) Using the RBFNN training procedure, RBF-CNN regressor widths and RBF centers are optimized. The above-mentioned hidden layer outputs from RBFNN are received by the corresponding CNN, which then executes the load prediction. In this way, the final prediction is determined as an average of ensemble load predictions. CNN is used to extract the input feature when input is transformed into a graphical representation. CNN used in this paper is one of the techniques of DL. DL serves as a great application in load forecasting.

(4) In the case of highly complicated problems like load forecasting, RBFNN deals with the "curse of dimensionality," which is usual for every kernel approach. To ease this drawback, a composite RBF is accompanied by a deep learning approach. RBFNN presents an incredible response for the evaluation of demand forecasting. Although its prediction is based on the width of the radial basis function in the hidden layer.

In this paper, a fuzzy-based prediction framework integrated with a deep learning network has been presented for STLF. This hybrid approach can capture hidden characteristics of load pattern and gain accuracy in results of load forecasting. On the basis of the obtained results and complete analysis, the following conclusions are being drawn: firstly, in comparison to the LSTM method (generally for RBFNN second layer) activations performed by CNN on RBF give around 9% improvement in forecasting accuracy. It indicates that higher forecast accuracy is attained by RBF-CNN regressors. Secondly, the application of CNN on the RBFNN hidden layer gives high robustness. Third, the proposed model (Fuzzy-RBF-CNN) performs better than ML-SVM and results in a 14% improvement on average. Fourth, in comparison of MAPEs of 24 h-SVM and fuzzy-RBFNN, the fuzzy clustering approach is more successful, as it provides 39% and 34% better performance with reference to 24 h-SVM. Thus, it shows the effectiveness of the fuzzy clustering method and the improvement in RBFNN response by CNN.

The rest part of this paper is explained in given Sections: Section 2 presents the relevant and recent literature for study. Section 3 explains the proposed hybrid model. Section 4 discusses the complete training procedure of the proposed model structure. Section 5 illustrates the case study in both interconnected and isolated power systems. Section 6 shows the outcome of training the system with different methods, and Section 7 concludes the proposed work.

2. Related Work

STLF defines the load prediction horizon from an hour to one week, which is significant for large-scale decision-making operations of power grids where group of countries have a single power system, such as the European Union. To clearly understand the new approach for the STLF model, the whole literature involves both the statistical model and machine learning models. Both of these models are subdivided into individual models and hybrid models. Hybrid models involve feature extraction, forecasting, and different optimization approaches, unlike individual models that involve only forecasters.

2.1. Individual Forecasting Models. In this STLF, the forecaster predicts the load consumption. Distributed methods are proposed [44] to forecast load using weather information. Auto regression integrated moving average (ARIMA) and Grey [45] are individual forecasting models used for subnetworks (which are formed by dividing power systems as per weather conditions). To determine the performance of adapted methods with respect to traditional models, two performance metrics, i.e., the root mean square error (RMSE) and the mean square error (MSE), are used. Here, the MSE checks whether the value of the forecast is close to the actual value, and its lower value indicates a better fit. RMSE will exaggerate the large errors, which is helpful when compared with other methods. A deep recurrent neural network (DRNN) is used to predict energy consumption [46]. This method outperforms other methods in terms of RMSE like convolution RNN, ARIMA, and SVR by 5.9%, 18.5%, and 12.1%, respectively. In [47], the author has proposed LSTM-RNN that mainly concentrates on accuracy while ignoring the convergence rate and calculation complexity. In [48], the author has tested the recency effect experienced in LF to improve prediction accuracy at the level of high model complexity. The author has presented a long-term forecasting model by conducting an analysis on western US energy utility. It is done for both peak and normal load usage. These aforementioned techniques are more robust with fast convergence but lag in the accuracy level of the forecast.

2.2. Hybrid Forecasting Model. In these models, techniques of feature extraction and optimization are used along with forecasters to improve forecasting accuracy. The author in [49] analyzes the complete power system structure on the basis of hourly load and weather conditions by applying a polynomial regression model. It has presented a model to predict the load on the distributed generation side using SVM and the fruit fly immune (FFI) algorithm. In [50], the author has proposed an IoT-based deep neural network for high precision. In addition, factors like temperature, humidity, and weather conditions [12] are taken into consideration. The author in [51] has presented a hybrid learning model to forecast the intensity of solar radiation [52]. The dynamic behavior of data is analyzed using a genetic algorithm (GA), back propagation (BP), and neural network. This model outperforms both in STLF and long-term load forecasting. To harvest the solar energy, it is necessary to optimally forecast its generation. For this purpose, the author in [53] has proposed a regression technique called least absolute shrinkage and selection operator (LASSO), which enhances forecast accuracy. ARIMA, wavelet neural network (WNN), and improved empirical mode decomposition (EMD) used to forecast load and FFI optimization is done. Its simulation results outperform those of existing methods when compared. The ANN model forecasts hourly energy consumption, and its model is trained using Levenberg–Marquardt (LM) and BP techniques [54]. Different parameters, like temperature, hourly/weekly energy usage, and dry bulb data, are taken as input. The accuracy of the model is tested on the basis of RMSE. An AI-based hybrid model [55] is proposed to predict the 24 hr load of polish grid and it is validated on offline data of Poland. EMD-based ensemble model using deep learning approach is used to forecast load, and it is tested on the Australian energy grid. In this paper, the data is broken up into intrinsic mode functions (IMFs), and each IMF is used to improve accuracy. The author in [56] presented the fully automated machine learning structure for forecasting the load. In [57], a hybrid incremental learning technique is used that combines discrete WT, EMD, and a random vector functional link network (RVFLN). The simulation result is evaluated on Australian energy data, and this model outperforms eight benchmark models. In [58], load forecasting is done using an extreme learning machine model (ELM), and the proposed study is validated by half hour resolution data of Australia. Its result outperforms existing methods such as RBF-ELM and mixed ELM. In [59], the author has applied hybrid of particle swarm optimization approach (PSO) and ELM where tanh function is used as activation function and avoids unwanted hidden nodes and overtraining. This proposed approach is better than RBFNN, according to the obtained results. In [60], to forecast the hotel building demand, i.e., highly irregular, the online modifier forecaster is proposed. This paper uses a clustering-based hybrid model, i.e., a combination of SVR and wavelet decomposition techniques. It results in higher accuracy than traditional methods. In [61], STLF is done using deep learning approach and tested on energy consumption of year 2014 of China cities. The simulation results show a significant impact of parameters like temperature and other weather conditions on energy usage. It also highlights the better prediction accuracy of the deep learning model in comparison to the random forecast and gradient boosting models. To improve efficiency and relative prediction accuracy, this study [62] will feed the output of the forecaster module to the optimization module. It improves the prediction accuracy at the cost of calculation complexity. To improve forecasting accuracy using dynamic mode decomposition and an extreme value constraint approach, the author [63] has presented the STLF model. The authors [64] presented STLF for distribution feeders. To improve both the stability and prediction accuracy of the model, the author has developed an integrated model of VMD, LSTM, and Bayesian techniques. The author [65] developed a hybrid approach to forecast the energy generation from solar panel-microgrid. This model used GA, PSO, and neuro-fuzzy approaches is tested on real-time power generation data. This prediction module forms the historical load profile by analyzing the stochastic load pattern of consumer demand and then forecasting the future load demand.

Due to ease in the implementation of electric load forecasting, ANN is mostly used as a machine learning approach. Number of layers, number of neurons, and learning rate are the most promising parameters to define the performance of ANN. Mainly, learning algorithms such as gradient descent, BP algorithm, etc. suffer from premature convergence and overfitting. In order to overcome this disadvantage, hybrid forecasting approaches have been discussed. Overfitting is reduced by data augmentation, feature selection, and creating ensemble load predictions. Hybrid techniques have enhanced model capabilities, but the problems of slow convergence and large computational time still persist compared to nonhybrid techniques.

All the aforementioned techniques produce satisfactory results for small data size only and their performance is highly dependent on knowledge and experience. Using a clustering approach on smaller data size leads to the problem of overfitting and creates a high generalization error. In a practical scenario, the data size is invariable, and there is no technique that can handle large data. This proposed study outperforms in comparison to present ANN methods and linear regression models. Table 1 presents a summary of related work.

Conclusions are figure out from above relevant literature survey: (1) no forecasting method is perfect in all respects; however, it depends on the application; (2) suffers from overfitting problem where model performs better in training but not in forecasting; and (3) compromise between forecasting accuracy and convergence rate. In this respect, a novel hybrid approach is proposed, which is integrated with three processes: (i) hybrid architecture composed of RBF, convolution, and pooling in a fully connected two-layer network; (ii) fuzzy clustering algorithm after data preprocessing; and (iii) FLANN algorithm-based optimization technique.

3. Detailed Hybrid Framework

The proposed model comprises 3 layers: fuzzy clustering, RBF-CNN regressor, and composite layer. Initially, the novel method aggregates the input data into multiple clusters using the fuzzy clustering method. Each cluster shares its position with other clusters in the neighborhood for ensemble prediction at the second layer. These clusters are generated as per the fuzzy rule, and this given clustering method needs very variable sets represented as membership function sets. These function sets depends on variables required for the problem [44], also on required cluster which split up input dataset. As per the proposed clustering method, the inputs are assigned to a second layer, where for each cluster the RBF-CNN regressor is applied. At the first layer, a corresponding cluster (fuzzy rule) creates the dataset, which is received by the RBF-CNN regressor. The training of these regressors requires kernel numbers, centres, and widths. These are assumed by the RBFNN training procedure. RBF kernels with optimized values convert the input data into higher-dimensional space, and CNN is trained with the implementation of this converted input data at the second stage. This process performs a detailed analysis of the relation between a kernel element and its neighbor, compared to RBFNN. With this above-mentioned 2 stage training procedure, RBF-CNN works as one neural network that comprises an RBF, a convolutional layer, an averaging pooling layer, and two fully connected layers. At the third layer, the average of ensemble predictions of the RBF-CNN regressor is done; this is the final prediction stage of the proposed model and corresponds to activated clusters in the clustering layer. A proposed model structure with three clusters that corresponds to the final load forecast is presented in Figure 1. Here, input $x\ (i)$ actuate fuzzy rules 1, 2, 3, . . ., k and RBFCNNs to provide an independent prediction. Figure 2 shows the basic flowchart diagram of the hybrid forecasting fuzzy-RBFNN model.

4. Training of Model Structure

4.1. Fuzzy Clustering Approach. To model the fuzzy-RBF-CNN, the input dataset is grouped according to variables that carry significant information. Hence, the highly correlated one defines the shape and quantity of clusters in the first layer. Most important input variables to model the load forecast are temperature, hours/months/weekend, and current load status. The training process of the first layer hybrid model contains variables with fuzzy membership functions. For continuous input variables, the Gaussian membership function is preferred, whereas for definite variables, the trapezoidal membership functions. Each variable is classified into fuzzy sets having the same behavior and semantic description, which are denoted by the membership function. The "hour" variable is represented with multifuzzy sets for daytime hours, and for nighttime hours, it is represented with a single-fuzzy set. Variables not involved in clustering are initialized with a membership value equal to 1.

During training cycle, x_i input vector is chosen through dataset and forms fuzzy rules. The membership output M (x_i) is determined per variable i using the x_i input vector. Furthermore, the activation functions for each variable are calculated, which are denoted as $S\ (i,\ k)$

$$S(i,k) = \sqrt[n]{\Pi Y_M(x_i)}, \tag{1}$$

where $Y_M\ (x_i)$: output of variable M contained in fuzzy rule k. n: most important input variable in numbers on which dataset is grouped.

For activation of variables less the existing rules, next fuzzy rule is

$$i_s = \operatorname{argmin} i\{fi, \operatorname{argmax}\{S\}(i,k)\}. \tag{2}$$

Table 1: Relevant work summary report in terms of techniques, aim, drawbacks, and remarks.

Techniques	Aim	Drawback	Remarks
Load forecasting based on weather information for bulk power system	Improvement in forecasting accuracy	Suitable for bulk power system only	Incorporating exogenous variables, the performance of bulk and distributed power systems improves
Residential forecasting using DRNN	Enhance the user's comfort level by reliable electricity availability	Model complexity increases	Residential energy forecasting is possible by sharing the load data of consumers to energy regulation commission
LSTM-RNN for residential forecasting	Improved accuracy	Increased in accuracy only for meter level forecast	Improvement in accuracy not in convergence rate
IoT-based load forecasting	Improved operation of power system with accuracy	Large complex framework	Impact on convergence rate
Forecasting based on big data approach	Accuracy improved for scalable models	Complex structure with less convergence	High complexity with improved accuracy
Week ahead forecasting using deep model with denoising auto encoders	Improved accuracy	Model performance is affected with reduced data size	Convergence rate is affected, but accuracy improved with large data size
Artificial intelligence-based load forecasting	Reduction in MSE with improved accuracy	Accuracy with high convergence rate	Sigmoid function reduced convergence rate
Intelligent hybrid model for load forecasting	Day-ahead load forecasting	Effective management of grid operation	Reliability is improved with high model complexity

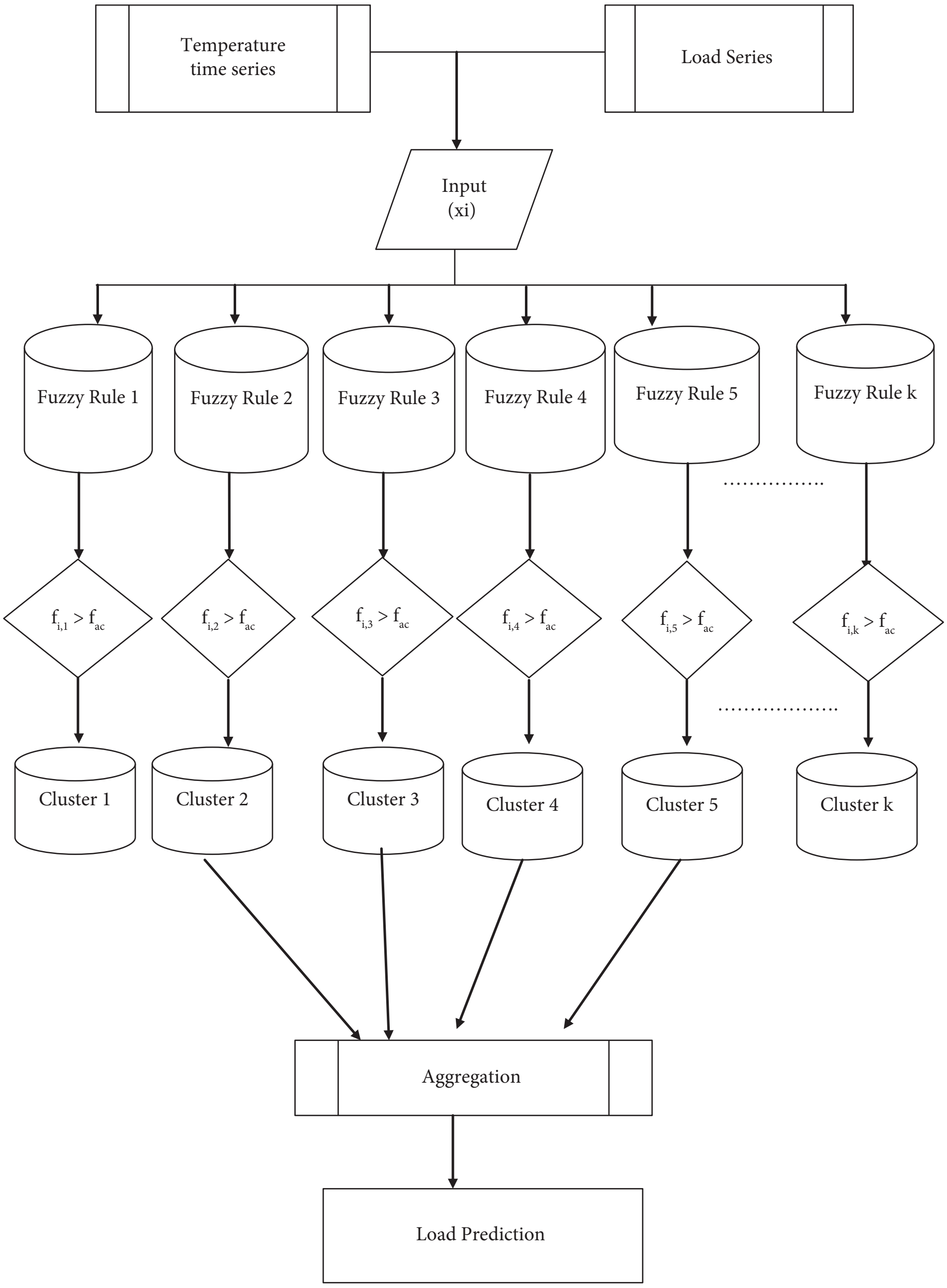


Figure 1: Fuzzy architecture for the proposed model.

Until the minimum activation $fi, \text{argmax}\{S\}\{(i,k)\}$ stops increasing, the training of the first layer of fuzzy-RBF-CNN continues till it reaches to the threshold value denoted by S_{TH}. The significance behind defining this threshold value is to reduce complexity of design framework by limiting number of clusters. The value of S_{TH} is 0.73, which is determined using the trial and error method.

The algorithm behind this approach is defined in steps, as shown below in Table 2.

4.2. Steps for RBFNN Training. In the case of RBF-CNN, kernels are associated with CNN while the training process is executed in dual phases: Initially, autonomous RBFNN is trained and different sets of RBFs are created through the cross-validation method. Secondly, input data is transformed into 3-dimensional arrays to train a CNN. By applying the permutation approach [45], the input variables that are least significant are removed from the set at the initial stage of the RBF-CNN process. In this process, the

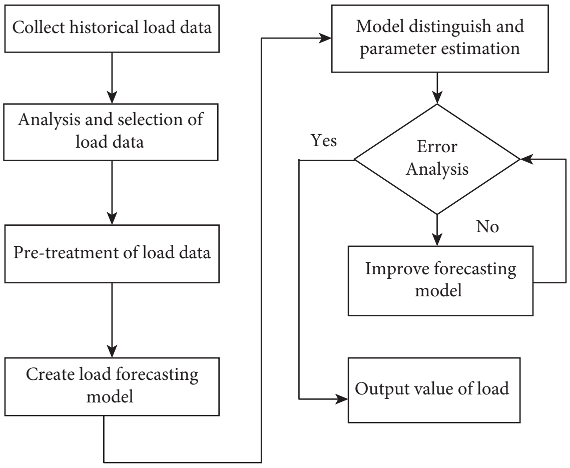

Collect historical load data
Analysis and selection of load data
Pre-treatment of load data
Create load forecasting model
Model distinguish and parameter estimation
Error Analysis
Yes
No
Improve forecasting model
Output value of load

Figure 2: Basic flowchart diagram of the hybrid forecasting fuzzy-RBFNN model.

Table 2: Fuzzy clustering algorithm.

(1) Load the membership function for each variable M
(2) Load input data set
(3) Select input variable x_i randomly
(4) Determine output variable $M\ (x_i)$
(5) Chose membership function of maximum output
(6) Develop first layer rule
(7) Calculate $S\ (i,\ k)$ for each variable using equation (1)
(8) Determine $S\max\ (i)$ for each $x\ (i)$
(9) Determine $x\ (i')$ for minima of $S\max\ (i)$
(10) Determine output variable $M\ (x_i)$
(11) Chose membership function for each variable having maximum output
(12) Develop next rule
(13) From equation (1), calculate $S\ (i,\ k)$ for each variable
(14) Min $S\max\ (i) < S_{\mathrm{TH}}$

baseline is defined, and then a randomly selected variable is permuted using the random forest algorithm [46].

Basically, RBFNN comprises a 2 layer neural network with a hidden layer consisting of nonlinear RBFs and linear output. Its hidden layer parameters are controlled by RBF numbers (control the model whether over or under fitting), RBF centers (plan the model characteristics), and radii of RBFs (compute activation function for RBF). The type of RBF is given by the following equation:

$$S_j^{2,k} = \frac{1}{e^{\sqrt{\sum\left(\left\|x\,(i) - c_{i,j}^{2,k}\right\|/\left(r_{i,j}^{2,k}\right)^2\right)}}}, \tag{3}$$

where k: cluster. i: input variable. j: indices for RBF, $j \in 1, J_m$ $S_j^{2,k}, c_{i,j}^{2,k}$, $(r_{i,j}^{2,k})^2$: activation, centers, and radii of RBFs corresponding the cluster k.

Applying k cluster with permutation approach and forming dataset, equation (3) presents the RBFNN hidden layer output, which is trained with the k-means algorithm to locate RBF centers and latter to evaluate RBF radii, FLANN algorithm is used. The FLANN approach is a higher order derivative optimization technique algorithm [47], which uses a single-layer feed forward network and is extensively used on account of its low computational complexity. Moreover, it overcomes nonlinearity in outputs due to the functional expansion feature.

In the RBFNN training process, the objective function is the sum of squared error, which is determined using validation and testing sets. The RBF weights and biases are upgraded and compared with the forecasted value per iteration. Then, the mean of squared error on validation and testing are computed and continues till error value is minimized. At that optimal point, the radii value is determined. With the hyperparameter optimization approach, namely "coarse-to-fine" [48], the above procedure is carried out for different RBFs.

Furthermore, a three-foldcross-validation is performed for each tested RBF that comprises validation and testing of sets. This testing deals with a continuous portion of the set, whereas the validation set is selected randomly. While averaging the cross-validation results, the optimal value of RBF's hidden layer is obtained. This is the process of obtaining the corresponding RBFNNs.

4.3. Steps of CNN Training. RBFNN deals with the "curse of dimensionality" issue because of the summation of activation for each element, and to mitigate this issue, activations are computed individually for each RBF element using the following equations:

$$S_j^{2,k} = \frac{1}{e^{\sqrt{\sum\left(\left\|x\,(i) - c_{i,j}^{2,k}\right\|/\left(r_{i,j}^{2,k}\right)^2\right)}}}, \quad \mathrm{j} \in 1, J_k. \tag{4}$$

After performing the RBFNN training procedure, each input vector is converted to 3-dimensional arrays. Then, training, validation, and testing are employed in equation (4), and three arrays of four-dimensional sets are obtained as $3 \times E \times J_k \times D_{\text{train},k}$, $3 \times E \times J_k \times D_{\text{val}›k}$ and $3 \times E \times J_k \times D_{\text{test},k}$, where E: input variables. J_k: most appropriate value of RBF number. $D_{\text{train},k}$, $D_{\text{val},k}$, and $D_{\text{test},k}$: sizes of training, validation, and testing sets, respectively.

In image processing applications, these datasets are used to establish CNN for deep neural network. CNN comprises input layer, output layer, and a number of hidden layers. This hidden layer involves convolution layers accompanied by a pooling layer with two coupled layers. The convolution layer takes out the features from images utilizing kernels and creates filters [19]. In the convolution layer, different images are obtained by applying weights to each filter. It is a deep neural network with feed forward propagation techniques. As compared with multilayer perceptrons, it provides the best accuracy in all nonlinear problems, such as load forecasting.

In CNN training, the FLANN optimization algorithm is implemented, CNN parameters are updated, and mean square errors are computed to evaluate the performance. Convolution layer consists of 32 filters with a kernel of size 2 × 4 pixels. It is followed by a pooling layer that sums up the filters with a stride of 2 pixels, while the coupled layers have 2048 and 512 neurons, respectively.

5. Case Studies

This proposed model is evaluated as per the customer consumption rate provided by the energy market. Iteratively, it runs on a daily basis as per recorded data, predicts the hourly loads for the current day, and then predicts the load for the next seven days ahead. While following data selection approach explained in [49], the input variables are historical load [50], data of temperature forecast [51], and calendar data (month/hour/year) along with special days indication. For the fuzzy clustering method, most significant variables are average value of previous day load, maximum temperature, and special days index with hour/month.

To compute the performance of the proposed hybrid model, two case studies are examined. The first study is for the Hellenic interconnected power system, where time series load data [53] covers the period from 1 January 2015 to 30 June 2019, and temperature prediction data is acquired from the SKIRON meteorological model [54]. This presented model was trained with recorded data for the first four years and tested with recorded data during 2018. Secondly, it is applied to the Cretan power system, which is an isolated system and highly loaded during summer. The Hellenic power system considers period from 1 Jan 2017 to 31 Dec 2019. During the summer of 2019, the peak load was 660 MW and the minimum load was 135 MW. Then, in 2017, the yearly peak load was 665 MW, while in 2018, it was 610 MW. Also, the fuzzy membership functions of the first layer are applied to these two case studies. In the case of the fuzzy clustering approach, the significant input variables in the first layer are the average value of day-ahead loads, the maximum temperature on a daily basis, the maximum temperature of the forecasted day, and the "hour," "month," and special day index of forecasting time. For "hour/month," three trapezoidal membership functions are applied, whereas for special days index, two trapezoidal membership functions are employed. The average value of day-ahead loads is modeled by employing 3 Gaussian membership functions, and the maximum temperature predicted on a daily basis is divided into 5 fuzzy sets presented by Gaussian membership functions. The performance of the proposed model is evaluated using the mean absolute error (MAE), mean absolute percentage error (MAPE), and the root mean squared error (RMSE).

In the Hellenic power system, the clusters formed are 76 for first layer while in the Cretan power system it is 68. The historical loads are modeled by applying 3 Gaussian membership functions. Variables "such as hour/month/special day index" are modeled using a fuzzy membership function. Maximum temperature is split into five fuzzy sets and is designed with a definite membership function of 1.

Clusters form subsets, which contain input samples that activate the fuzzy rule. Then, the training procedure of RBF-CNNs for each fuzzy rule is repeated. Further, ensemble predictions are produced for RBF-CNNs that indicate fuzzy initialized by the activation value more than the threshold $S_{\text{activation}} = 0$.

6. Outcome of Proposed Framework

6.1. Standards. To certify the work of the presented hybrid model, two advances and a traditional forecasting model are developed. The shape of the first two models consists of 24 regressors with similar structures. Each regressor is trained to obtain prediction data for one hour a day. For the first standard model, regressors are developed using the MLR approach [55], while for the second model, SVMs [66] are employed. This standard model is marked as 24 hr MLR and 24 h-SVM. For validating the performance of the proposed fuzzy-RBF-CNN, the RBF-CNN regressors are being replaced by SVMs. This method is marked as ML-SVM, having 3 layers. SVM is a supervised learning algorithm that does very well in the classification of data into different datasets. MLR offers generic and extendable configurations for clustering, classification, regression, etc. Fuzzy-RBFNN is eminent from other techniques in terms of universal approximation and higher learning speed. These three techniques are widely used to address other research problems. The membership functions corresponding to all variables of the fuzzy-RBF-CNN first layer are designed once and applied to all presented case studies without fine tuning. Fuzzification of the input variables leads to the creation of fuzzy rules using the linguistic representations of the corresponding membership functions. Each fuzzy rule defines a data cluster of the fuzzy-RBF-CNN first layer to which an RBF-CNN regressor will be connected. The fuzzy rules are constructed using an iterative training procedure.

Using the same technique as discussed in this proposed work, the input data is clustered in the first layer. Different SVMs are trained for each cluster to develop different forecasts, and the mean of the developed ensemble predictions gives rise to the final prediction value. The forecast values applied to the proposed model obtained from RBFNN are determined.

For every activated cluster, these values first produce the mean of three RBFNN outputs and then all activated cluster predictions. This model is marked as fuzzy-RBFNN. Furthermore, the execution of the persistence method is evaluated and then compared with the proposed model and the two aforesaid models [56].

6.2. Hellenic Interconnected Network. During forecasting, the MAPEs of fuzzy-RBF-CNN and fuzzy-RBFNN have identical values in comparison to other benchmarks, which indicate less robustness. In Tables 3 and 4, the MAPE and RMSE of the proposed model are shown. Initially, load time series data of a nonlinear and complex nature are demonstrated and compared for the accuracy obtained from 24 h-MLR and 24 h-SVM. The performance of 24 h-SVM and ML-SVM indicates less improvement by the proposed model. The proposed method achieves improvements in the range of 5% to 20% [56]. However, in comparison to the least square method (generally used for RBFNN second layer) activations performed by CNN on RBF gives around a 9% improvement in outcome.

Table 3: MAPEs for the proposed model and standard models.

Day ahead	Persistence method	MLR	SVM	ML-SVM	LSTM	Fuzzy-RBFNN	Fuzzy-RBF-CNN
1	5.90	3.45	3.00	1.93	1.80	1.85	1.57
2	7.72	3.78	3.25	2.19	1.90	1.95	1.63
3	6.33	3.81	3.30	2.29	1.85	1.95	1.65
4	6.70	3.94	3.50	2.45	2.20	1.97	1.63
5	6.70	4.08	3.61	2.50	2.19	2.00	1.68
6	6.01	4.11	3.68	2.54	2.17	2.01	1.69
7	7.85	4.14	3.70	2.60	2.20	2.05	1.72

Table 4: RMSEs for the proposed model and standard models.

Day ahead	Persistence method	MLR	SVM	ML-SVM	Fuzzy-RBFNN	Fuzzy-RBF-CNN
1	1308	710	593	583	490	460
2	1771	780	662	627	511	470
3	1902	801	689	660	512	471
4	1978	852	750	731	515	475
5	1908	885	782	782	520	479
6	1839	901	798	788	525	480
7	1770	888	802	800	534	487

In this case study, comparing the execution of ML-SVM with that of 24 h-SVM, the structure developed from the fuzzy clustering approach shows remarkable improvement for the short horizon, whereas for the longer horizon (more than 4 days ahead), both ML-SVM and 24 h-SVM show similar performance. For this study, higher forecast accuracy is attained by RBF-CNN regressors.

Table 5 shows the calculated MAPEs for fuzzy-RBF-CNN and ML-SVM structures obtained while calculating forecast values for working and nonworking days. For a complete prediction scope including both working and nonworking days, this proposed model shows better performance than the standard model. On nonworking days, the MAPE of Fuzzy-RBF-CNN is 2.11% in the case of three and four days-ahead horizon.

Figure 3 illustrates the accuracy of the aforementioned models for successive days in March 2018. Finally, shows the finest performances procured from the fuzzy-RBFNN and fuzzy-RBF-CNN models, whereas the standard model shows analogous performance.

6.3. Cretan Power Network. Tables 6 and 7 show the MAPE and RMSE for the week ahead horizon of the fuzzy-RBF-CNN and standard models. The working of the proposed model is closer to ML-SVM. In this case, when predicting one day ahead, the MAPE and RMSE of ML-SVM have lesser difference in comparison to the proposed model. In case of more than 2 days ahead horizon, this proposed model outperforms all the standard models. On comparison of the MAPEs of 24 h-SVM with the MAPEs of fuzzy-RBFNN and ML-SVM, the proposed fuzzy clustering approach is more successful for longer horizons, as it provides 39% and 34% mean improvements in ML-SVM and fuzzy-RBF-CNN, respectively, when compared with 24 h-SVM. The application of CNN to the output of the RBFNN hidden layer gives remarkable robustness to RBFNNs and performs better than SVM. It is shown that the unsatisfactory performance of 24 h-MLR proves more complexity in the previous case study. Furthermore, "persistence method" is better than 24 h-MLR and 24 h-SVM in the case of a longer horizon. Table 8 shows the MAPE values for the ML-SVM and proposed model that are obtained with a similar evaluation approach. Table 9 shows the performance results from training, testing, and validation of the fuzzy-RBFNN model in terms of RMSE, MSE, and MAE.

For nonworking days, the proposed model (Fuzzy-RBF-CNN) performs better than ML-SVM and results in a 14% improvement on average. While on working days, ML-SVM performs better than the proposed model in the case of current-day prediction while having similar performance for day-ahead forecasting. Figure 4 illustrates prediction of both presented models for a nonworking day (15/08/2019) and a working day (16/12/2019). It shows a performance improvement of fuzzy-RBF-CNN when compared with the standard model for both days. Specifically, during morning hours, the standard model error outreach 9%, whereas using the proposed fuzzy-RBF-CNN model for the same morning hours, this error comes out to be lower than 1%. Figure 5 shows the convergence characteristics for four models (i.e., ML-SVM, LSTM, Fuzzy-RBFNN, and Fuzzy-RBF-CNN) and represents MSE values for each model. This graph shows that our proposed model shows harmony in both accuracy and convergence rate when compared with other models. Results for training, testing, and validation are presented in Figure 6.

Table 5: Performance evaluation of the presented model and ML-SVM in the case of both normal and special days using MAPE values.

Day ahead	ML-SVM		Fuzzy-RBF-CNN	
	Working days	Nonworking days	Working days	Nonworking days
1	2.82	2.20	2.48	2.10
2	3.15	3.40	2.70	1.05
3	3.25	3.51	2.71	1.11
4	3.48	3.58	2.69	1.11
5	3.58	3.60	2.74	1.15
6	3.60	3.62	2.75	1.13
7	3.65	3.70	2.78	1.15

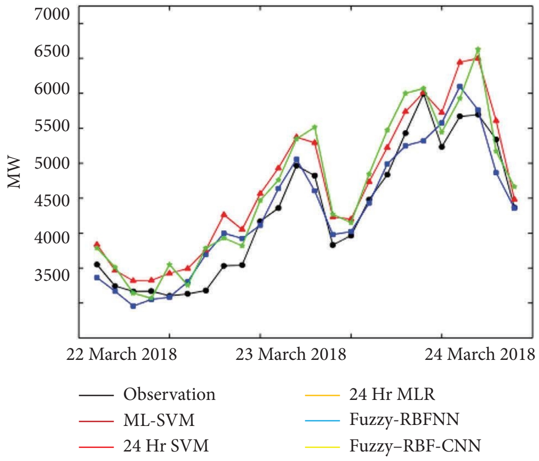


Figure 3: Comparison of the proposed model with standard models in a two-day period of March 2018.

Table 6: MAPEs for the proposed model and standard models.

Days ahead	Persistence	24 h-MLR	24 h-SVM	ML-SVM	LSTM	Fuzzy-RBFNN	Fuzzy-RBF-CNN
1	6.15	3.95	3.58	2.85	2.20	3.28	1.90
2	5.51	4.25	3.89	2.38	2.25	3.68	3.35
3	6.87	6.95	5.85	4.66	3.75	3.89	3.52
4	6.51	6.05	5.01	4.82	4.05	3.08	2.70
5	7.08	6.45	5.25	4.95	4.37	4.18	3.89
6	8.68	6.28	7.17	6.03	5.47	4.28	2.90
7	8.78	7.45	7.69	6.18	5.98	4.38	4.01

Table 7: RMSEs for the proposed model and standard models.

Days ahead	Persistence	24 h-MLR	24 h-SVM	ML-SVM	Fuzzy-RBFNN	Fuzzy-RBF-CNN
1	102.69	84.55	60.34	44.35	64.12	56.40
2	125.11	99.48	65.52	50.48	67.93	60.35
3	126.30	101.88	78.72	74.11	70.93	64.01
4	134.28	100.72	99.55	76.85	73.02	66.45
5	140.56	115.51	132.49	58.38	74.32	67.65
6	144.20	120.17	178.45	59.11	75.01	68.08
7	119.02	139.20	222.56	69.31	75.94	69.12

Table 8: Performance evaluation of the presented model and ML-SVM in both working and nonworking days using MAPE values.

Days ahead	ML-SVM		Fuzzy-RBF-CNN	
	Working days	Nonworking days	Working days	Nonworking days
1	2.82	2.15	3.92	1.63
2	2.39	2.75	2.40	2.12
3	2.65	2.91	2.63	2.42
4	2.80	2.95	2.81	2.40
5	2.94	3.23	2.92	2.62
6	5.01	4.38	4.95	2.68
7	5.15	4.52	5.02	4.87

Table 9: Results from training, testing, and validation of the fuzzy-RBFNN model.

Fuzzy-RBFNN model	Performance		
	R^2	MSE (%)	MAE (%)
All data	0.89	2.20	10.28
Training	0.93	1.39	8.20
Testing	0.87	3.38	14.0
Validation	0.86	1.22	8.30

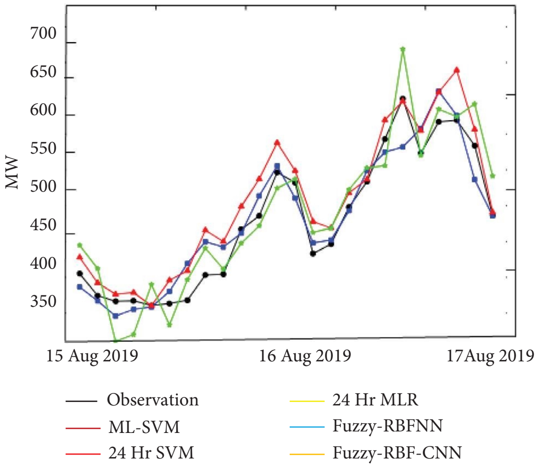


Figure 4: Comparison of the proposed model with standard models in a two-day period of August 2019.

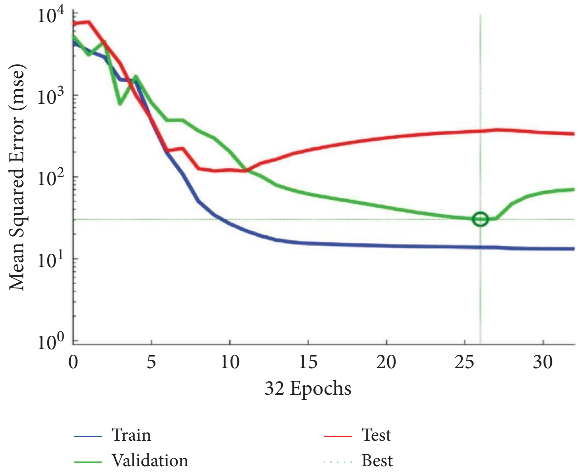


Figure 6: Results from training, testing, and validation.

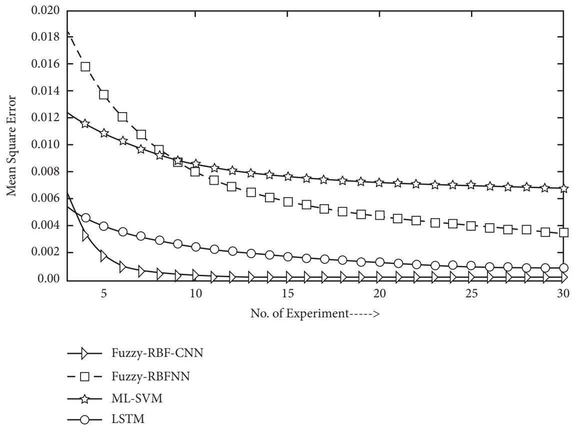


Figure 5: Convergence characteristics with MSE values.

7. Conclusion and Future Scope

Accurate LF is important for applications in different operations of the power grid and in decision-making. More accurate load forecasting mitigates the energy cost, enhances power system security, develops the optimal power plan, and therefore provides socioeconomic benefits for power grid management. The individual forecasting models fail to achieve desirable performance due to some limitations: (1) being unable to conduct forecasting with highly varying data like electric loads; (2) needing a large amount of historical data for forecasting; (3) having low accuracy due to the absence of a data preprocessing feature; and (4) overfitting and a low convergence rate. A hybrid model or combined model reduces the negative influences that are inherent in each individual model. It also takes the most advantages of individual models and is less sensitive to the certain factors that make individual model to give unsatisfactory performance. It is clear that for load forecasting, the hybrid model is highly fruitful than the individual model. So, this study develops the hybrid forecasting model and completely utilizes the benefits of individual models with enhanced performance. In this paper, a fuzzy-based prediction framework integrated with a deep learning network has been presented for STLF. This hybrid approach can capture hidden characteristics of load pattern and gain the accuracy in results of load forecasting. This complete framework is integrated with three processes: (i) hybrid architecture composed by RBF, convolution, and pooling in a fully connected two-layer network; (ii) fuzzy clustering algorithm that splits the input variables into orthogonal expansions after data preprocessing; and (iii) FLANN algorithm-based optimization technique. The main idea behind this study is to generate ensemble predictions from multiple local regressors, and this regression variable activates the forecast process using the data clustering method to assign the input to different clusters. This proposed model is designed to predict the one-week ahead load demand, and its performance is tested on two power networks, i.e., the Hellenic interconnected and Cretan power networks. This method is verified by comparing it with four benchmark models like 24 hr-MLR, 24 hr-SVM, ML-SVM, and Fuzzy-RBFNN, in terms of forecasting accuracy. On the basis of the obtained results and complete analysis, the following conclusions are being drawn: firstly, in comparison to the LSTM method (generally for RBFNN second layer), activations performed by CNN on RBF give around a 9% improvement in forecasting accuracy. It indicates that higher forecast accuracy is attained by RBF-CNN regressors. Secondly, the application of CNN on the RBFNN hidden layer gives high robustness. Third, the proposed model (Fuzzy-RBF-CNN) performs better than ML-SVM and results in a 14% improvement on average. Fourth, in comparison of the MAPEs of 24 h-SVM and fuzzy-RBFNN, the fuzzy clustering approach is more successful, as it provides 39% and 34% better performance with reference to 24 h-SVM. Thus, it shows the effectiveness of the fuzzy clustering method and the improvement in RBFNN response by CNN. This deep learning hybrid technique offers the limitation that it will not perform properly for complex hierarchical data structures. For future work, weather conditions can also be fed as an input to the hybrid model to improve the research findings. Also, researchers can forecast the power loads for weather insensitive customers by searching the choice of input frames to mitigate negative impact of weather characteristics.

Data Availability

The data obtained in results are calculated values after applying computational algorithm and the steps of algorithms are given in the manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Self-funded paper.

References

[1] L. Ekonomou, C. Christodoulou, and V. Mladenov, “A short-term load forecasting method using artificial neural networks and wavelet analysis,” *Int. J. Power Syst*, vol. 1, pp. 64–68, 2016.

[2] A. Veit, C. Goebel, R. Tidke, C. Doblander, and H. A. Jacobsen, “Household electricity demand forecasting: benchmarking state-of-the-art methods,” in *Proceedings of the 5th international conference on Future energy systems*, pp. 233-234, ACM, Cambridge, United Kingdom, June 2014.

[3] H. Cho, Y. Goude, X. Brossat, and Q. Yao, “Modeling and forecasting daily electricity load curves: a hybrid approach,” *Journal of the American Statistical Association*, vol. 108, pp. 7–21, 2013.

[4] F. Javed, N. Arshad, F. Wallin, I. Vassileva, and E. Dahlquist, “Forecasting for demand response in smart grids: an analysis on use of anthropologic and structural data and short term multiple loads forecasting,” *Applied Energy*, vol. 96, pp. 150–160, 2012.

[5] Y. Iwafune, Y. Yagita, T. Ikegami, and K. Ogimoto, “Short-term forecasting of residential building load for distributed energy management,” in *Proceedings of the Energy Conference (ENERGYCON), 2014 IEEE International*, pp. 1197–1204, IEEE, Cavtat, Croatia, May 2014.

[6] Short Term Electricity Load Forecasting on Varying Levels of Aggregation, https://arxiv.org/abs/1404.0058, 2018.

[7] C. Gerwig, “Short term load forecasting for residential buildings—an extensive literature review,” in *Intelligent Decision Technologies*, pp. 181–193, Springer, Cham, Salmon Tower Building NY, USA, 2015.

[8] H. Hippert, C. Pedreira, R. Souza, and R. C. Souza, “Neural networks for short-term load forecasting: a review and evaluation.” *IEEE Transactions on Power Systems*, vol. 16, pp. 44–55, 2001.

[9] K. Metaxiotis, A. Kagiannas, D. Askounis, and J. Psarras, “Artificial intelligence in short term electric load forecasting: a state-of-the-art survey for the researcher.” *Energy Conversion and Management*, vol. 44, pp. 1525–1534, 2003.

[10] S. Tzafestas and E. Tzafestas, “ElpidaTzafestas. Computational intelligence techniques for short-term electric load forecasting.” *Journal of Intelligent and Robotic Systems*, vol. 31, no. 1/3, pp. 7–68, 2001.

[11] M. Ghofrani, M. Ghayekhloo, A. Arabali, and A. Ghayekhloo, “A hybrid short-term load forecasting with a new input selection framework.” *Energy*, vol. 81, pp. 777–786, 2015.

[12] M. Rizwan, M. Jamil, and D. P. Kothari, “Generalized neural network approach for global solar energy estimation in India,” *IEEE Transactions on Sustainable Energy*, vol. 3, no. 3, pp. 576–584, 2012.

[13] M. Chui and S. Francisco, “Artificial intelligence the next digital Frontier?” *McKinsey and Company Global Institute*, vol. 47, 2017.

[14] C. Oh, T. Lee, Y. Kim, SoH. Park, and B. Suh, “Us vs. Them: understanding artificial intelligence technophobia over the Google DeepMind challenge match,” in *Proceedings of the 2017 CHI conference on human factors in computing systems*, pp. 2523–2534, ACM, Denver, Colorado, USA, May 2017.

[15] M. Skilton and F. Hovsepian, “Example case studies of impact of artificial intelligence on jobs and productivity,” in *The 4th Industrial Revolution*, pp. 269–291, Palgrave Macmillan, Cham, London, 2018.

[16] O. Valgaev, K. Friedrich, and S. Harmut, “Low-voltage power demand forecasting using k-nearest neighbors approach,” in *Proceedings of the Innovative Smart Grid Technologies-Asia (ISGT-Asia)*, pp. 1019–1024, IEEE, Melbourne, VIC, Australiapp, November 2016.

[17] O. Valgaev and K. Friederich, “Building Power Demand Forecasting Using K-Nearest Neighbors Model-Initial approach,” in *Proceedings of the Power and Energy Engineering Conference (APPEEC), 2016 IEEE PES Asia-Pacific*, pp. 1055–1060, IEEE, Xi’an, China, Octobe 2016.

[18] T. Soubdhan, J. Ndong, H. Ould-Baba, and M. T. Do, “A robust forecasting framework based on the Kalman filtering approach with a twofold parameter tuning procedure: application to solar and photovoltaic prediction.” *Solar Energy*, vol. 131, pp. 246–259, 2016.

[19] I. Goodfellow, B. Yoshua, and C. Aaron, *Deep learning*, MIT press, Cambridge, MA, USA, 2016.

[20] J. G. Jetcheva, M. Majidpour, and W.-P. Chen, "Neural network model ensembles for building-level electricity load forecasts." *Energy and Buildings*, vol. 84, pp. 214–223, 2014.

[21] Y. Liu, W. Liu, M. A. Obaid, and I. A. Abbas, "Exponential stability of Markovian jumping Cohen–Grossberg neural networks with mixed mode-dependenttime-delays," *Neurocomputing*, vol. 177, pp. 409–415, 2016.

[22] M. Chaouch, "Clustering-based improvement of non-parametric functional time series forecasting: application to intra-dayhousehold-level load curves," *IEEE Transactions on Smart Grid*, vol. 5, pp. 411–419, 2014.

[23] D. Niu and S. Dai, "A short-term load forecasting model with a modified particle swarm optimization algorithm and least squares support vector machine based on the denoising method of empirical mode decomposition and grey relational analysis." *Energies*, vol. 10, p. 408, 2017.

[24] R. C. Deo, X. Wen, and F. Qi, "A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset." *Applied Energy*, vol. 168, pp. 568–593, 2016.

[25] S. Liu, G. Wei, Y. Song, and Y. Liu, "Extended Kalman filtering for stochastic nonlinear systems with randomly occurring cyber attacks," *Neurocomputing*, vol. 207, pp. 708–716, 2016.

[26] C. Wen, Y. Cai, Y. Liu, and C. Wen, "A reduced-order approach to filtering for systems with linear equality constraints," *Neurocomputing*, vol. 193, pp. 219–226, 2016.

[27] N Zeng, H. Zhang, W. Liu, J. Liang, and F. E. Alsaadi, "A switching delayed PSO optimized extreme learning machine for short-term load forecasting," *Neurocomputing*, vol. 240, pp. 175–182, 2017.

[28] J.-B. Fiot and F. Dinuzzo, "Electricity demand forecasting by multi-task learning," *IEEE Transactions on Smart Grid*, vol. 9, pp. 544–551, 2018.

[29] V. Thouvenot, A. Pichavant, Y. Goude, A. Antoniadis, and J.-M. Poggi, "Electricity forecasting using multi-stage estimators of nonlinear additive models," *IEEE Transactions on Power Systems*, vol. 31, pp. 3665–3673, 2016.

[30] L. Ghelardoni, A. Ghio, and D. Anguita, "Energy load forecasting using empirical mode decomposition and support vector regression." *IEEE Transactions on Smart Grid*, vol. 4, pp. 549–556, 2013.

[31] H Kebriaei, B. N. Araabi, and A. Rahimi-Kian, "Short-term load forecasting with a new nonsymmetric penalty function," *IEEE Transactions on Power Systems*, vol. 26, pp. 1817–1825, 2011.

[32] H Mao, X. J. Zeng, G. Leng, Y. JieZhai, and J. A. Keane, "Short-term and midterm load forecasting using a bilevel optimization model." *IEEE Transactions on Power Systems*, vol. 24, pp. 1080–1090, 2009.

[33] H. Hahn, S. Meyer-Nieberg, and S. Pickl, "Electric load forecasting methods: tools for decision making," *European Journal of Operational Research*, vol. 199, pp. 902–907, 2009.

[34] S. N. Fallah, R. Deo, M. Shojafar, M. Conti, and S Shamshirband, "Computational intelligence approaches for energy load forecasting in smart energy management grids: state of the art, future challenges, and research directions," *Energies*, vol. 11, p. 596, 2018.

[35] L Hernandez, C Baladron, J. M. Aguiar et al., "A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings," *IEEE Communications Surveys and Tutorials*, vol. 16, pp. 1460–1495, 2014.

[36] K. Amarasinghe, D. L. Marino, and M. Manic, "Deep neural networks for energy load forecasting," in *Proceedings of the IEEE 26th International Symposium on Industrial Electronics (ISIE)*, pp. 1483–1488, Edinburgh, UK, June 2017.

[37] J. Bedi and D. Toshniwal, "Deep learning framework to forecast electricity demand." *Applied Energy*, vol. 238, pp. 1312–1326, 2019.

[38] F. He, J. Zhou, Z.-kai Feng, G. Liu, and Y. Yang, "A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm." *Applied Energy*, vol. 237, pp. 103–116, 2019.

[39] S. Wang, X. Wang, S. Wang, and D. Wang, "Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting," *International Journal of Electrical Power and Energy Systems*, vol. 109, pp. 470–479, 2019.

[40] X. Dong, L. Qian, and L. Huang, "A cnn based bagging learning approach to short-term load forecasting in smart grid," in *Proceedings of the 2017 IEEE SmartWorld, ubiquitous intelligence and computing, advanced and trusted computed, scalable computing and communications, cloud and big data computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pp. 1–6, IEEE, San Francisco, CA, USA, August 2017.

[41] L. Li, K. Ota, and M. Dong, "Everything is image: cnn-basedshort-term electrical load forecasting for smart grid," in *Proceedings of the 2017 14th International Symposium on Pervasive Systems, algorithms and networks and 2017 11th international conference on frontier of computer science and technology and 2017 third international symposium of creative computing (ISPAN-FCST-ISCC)*, pp. 344–351, IEEE, Exeter, UK, June 2017.

[42] X. Dong, L. Qian, and L. Huang, "Short-term load forecasting in smart grid: a combined cnn and k-means clustering approach," in *Proceedings of the 2017 IEEE international conference on big data and smart computing (BigComp)*, pp. 119–125, IEEE, February 2017.

[43] P.-H. Kuo and C.-J. Huang, "A high precision artificial neural networks model for short-term energy load forecasting." *Energies*, vol. 11, p. 213, 2018.

[44] G. Sideratos and N. D. Hatziargyriou, "An advanced statistical method for wind power forecasting." *IEEE Transactions on Power Systems*, vol. 22, pp. 258–265, 2007.

[45] C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis, "Conditional variable importance for random forests." *BMC Bioinformatics*, vol. 9, p. 307, 2008.

[46] G. Sideratos, A. Ikonomopoulos, and N. Hatziargyriou, "A committee of machine learning techniques for load forecasting in a smart grid environment," *Int. J. Energy Power*, vol. 4, pp. 98–108, 2015.

[47] S. K. Mishra, G. Panda, M. Sukadev, and S. Ajit Kumar, "Exponential functional link artificial neural networks for denoising of image corrupted by gaussian noise," in *Proceedings of the 2009 International Conference on Advanced Computer Control*, pp. 355–359, IEEE, Singapore, January 2009.

[48] H.-X. Tang and H. Wei, "A coarse-to-fine method for shape recognition," *Journal of Computer Science and Technology*, vol. 22, pp. 330–334, 2007.

[49] P. Jiang, F. Liu, and Y. Song, "A hybrid forecasting model based on date-framework strategy and improved feature selection technology for short-term load forecasting," *Energy*, vol. 119, pp. 694–709, 2017.

[50] L. Xiao, W. Shao, T. Liang, and C. Wang, "A combined model based on multiple seasonal patterns and modified firefly

algorithm for electrical load forecasting," *Applied Energy*, vol. 167, pp. 135–153, 2016.

[51] Z. Hu, Y. Bao, T. Xiong, and R. Chiong, "Hybrid filter–wrapper feature selection for short-term load forecasting." *Engineering Applications of Artificial Intelligence*, vol. 40, pp. 17–27, 2015.

[52] M. Jamil and A. S. Anees, "Optimal sizing and location of SPV (solar photovoltaic) based MLDG (multiple location distributed generator) in distribution system for loss reduction, voltage profile improvement with economical benefits," *Energy*, vol. 103, pp. 231–239, 2016.

[53] Son, Y. Geon, B. Chan Oh, M. Amoasi Acquah, and S. Yul Kim, "Optimal facility combination set of integrated energy system based on consensus point between independent system operator and independent power producer," *Energy*, vol. 266, p. 126422, 2023.

[54] Castorina, Giuseppe, A. Semprebello et al., "Performance of the WRF model for the forecasting of the V-shaped storm recorded on 11–12 November 2019 in the eastern sicily," *Atmosphere*, vol. 14, no. 2, p. 390, 2023.

[55] T. Hong, Pu Wang, and H. Lee Willis, "A naïve multiple linear regression benchmark for short term load forecasting," in *Proceedings of the 2011 IEEE Power and Energy Society General Meeting*, pp. 1–6, IEEE, Detroit, MI, USA, July 2011.

[56] S. Dutta, Y. Li, A. Venkataraman et al., "Load and renewable energy forecasting for a microgrid using persistence technique." *Energy Procedia*, vol. 143, pp. 617–622, 2017.

[57] A. S. Khwaja, X. Zhang, A. Anpalagan, and B. Venkatesh, "Boosted neural networks for improved short-term electric load forecasting." *Electric Power Systems Research*, vol. 143, pp. 431–437, 2017.

[58] G. Sideratos, A. Ikonomopoulos, and N. D. Hatziargyriou, "A novel fuzzy-based ensemble model for load forecasting using hybrid deep neural networks," *Electric Power Systems Research*, vol. 178, Article ID 106025, 2020.

[59] J. Li, D. Deng, J. Zhao et al., "A novel hybrid short-term load forecasting method of smart grid using MLR and LSTM neural network," *IEEE transactions on industrial informatics*, vol. 17, 2020.

[60] G. Hafeez, K. S. Alimgeer, and I. Khan, "Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid." *Applied Energy*, vol. 269, p. 114915, 2020.

[61] M. Jamil and S. Mittal, "Hourly load shifting approach for demand side management in smart grid using grasshopper optimisation algorithm," *IET Generation, Transmission and Distribution*, vol. 14, no. 5, pp. 808–815, 2019.

[62] M. Jamil, A. Kalam, A. Ansari, and M. Rizwan, "Generalized neural network and wavelet transform based approach for fault location estimation of a transmission line," *Applied Soft Computing*, vol. 19, pp. 322–332, 2014.

[63] M. Rizwan, M. Jamil, S. Kirmani, and D. Kothari, "Fuzzy logic based modeling and estimation of global solar energy using meteorological parameters." *Energy*, vol. 70, pp. 685–691, 2014.

[64] S. H. Rafi, S. Hasan, S. R. Deeba, and E. Hossain, "A short-term load forecasting method using integrated CNN and LSTM network." *IEEE Access*, vol. 9, pp. 32436–32448, 2021.

[65] H. Eskandari, M. Imani, and M. P. Moghaddam, "Convolutional and recurrent neural network based model for short-term load forecasting," *Electric Power Systems Research*, vol. 195, p. 107173, 2021.

[66] E. Ceperic, V. Ceperic, and A. Baric, "A strategy for short-term load forecasting by support vector regression machines," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4356–4364, 2013.

Kochi Chapter

Indian Geotechnical Conference
IGC 2022
15th – 17th December, 2022, Kochi

Application of Machine Learning in Prediction of Load Settlement Behavior of Piles Based on CPT Data

Mansi Aggarwal[1] and Ashok K. Gupta[2]

[1] Assistant Professor, Department of Civil Engineering, Meerut Institute of Engineering & Technology, Meerut – 250005. India, mansiaggarwal258@gmail.com

[2] Professor, Department of Civil Engineering, Delhi Technological University, Delhi - 110042, India, akgupta@dtu.ac.in

Abstract. Machine Learning can be successfully utilized in geotechnical designing applications, where vulnerability is a portion of nature, to create a vigorous predictive models foundation for designing parameters/behaviours. Formerly, geotechnical plan parameters are not continuously straightforwardly measured from a research facility and in-situ tests, or maybe frequently assessed from observational or numerical relationships that are created from regression fitting to a dataset. ML models were created to train a nearby dataset. The developing volume of data databases presents openings for progressed information examination methods from machine learning inquire about. Applied applications of ML are exceptionally distinctive from hypothetical or observational studies. In arrange to cure this circumstance, examined feasible applications of ML and created a proposition for a seven-step preparation that can direct viable applications of ML in design. In this paper, an ML model is created for predicting pile behaviours based on the results of cone penetration test (CPT) data. Roughly 500 data sets, gotten from the published literature, are utilized to create the ML model. The paper compares the predictions obtained by the ML with those given by a number of conventional methods and it is watched that the ML model essentially outperforms the conventional strategies.

Keywords: Pile Behaviour, Load Settlement. Machine Learning applications, seven-step model; modelling.

1 Introduction

Over the final few decades, we have seen a blast in the data era related to all perspectives of life counting all designing disciplines. There has been an increment in dynamic data collection to be utilized for fathoming basic building issues such as framework administration [8]. One striking case of information collection is the National Bridge Stock within the US. In most data collection cases, data has been accumulated without knowing how it'll be analyzed or utilized, and to date, no major commonsense benefit has been picked up from these information collection endeavors. As of late, an unused set of strategies for information extraction from information has emerged from machine learning (ML), which may be a department of counterfeit insights (AI). The initial objective of ML methods was the mechanized era of information for its joining in master frameworks. This era was anticipated to lighten the information procurement bottleneck frequently related to the construction of expert systems.

Whereas there have been demonstrations of information obtained by single ML methods (e.g., [4]), there has not been a critical commonsense advance in utilizing single ML strategies as standard devices by engineers due basically for two reasons. To begin with, practical issues are frequently as well complex to be taken care of by a single strategy and moment, the errand of applying ML strategies in building hone is much more complex than portrayed in those early considers; it isn't essentially a matter of taking a program and applying it to information. To overcome the impediments of existing learning strategies with respect to the primary reason, ML analysts hypothesized that the arrangement of differences and complexity in learning circumstances requires the utilization of numerous ML methods. Such multi strategy learning [1] would empower the assortment of information available for learning to be taken into consideration.

In this paper, artificial neural systems (ANNs) are utilized to anticipate the behavior of piles based on 56 individual pile load tests. These tests were carried out on locales joining different soil sorts, a few commonly received pile types, and a extend of geotechnical conditions counting layered soil profiles. The planned ANN demonstrate is in a position to anticipate the whole load-settlement behavior of concrete, steel, and composite heaps, either bored or driven, and to account more precisely for the changeability of soil properties along the shaft of the pile, the implanted length of the pile is sub-divided into 5 fragments of break even with thickness, with each related with a normal of qc and CPT sleeve friction, fs, over that portion. The points of the paper are to (i) create an ANN show for precisely predicting the load-settlement behavior of single, axially-loaded piles over a wide run of connected loads, pile characteristics and establishment strategies, and soil and ground conditions (ii) to investigate the relative significance of the components influencing pile behavior by carrying out affectability examinations; (iii) compare the execution of the ideal ANN show against a few of the foremost commonly utilized conventional strategies; and (iv) propose an arrangement of ANN-based load-settlement charts for anticipating pile behavior, to encourage the demonstrate being embraced in practice.

2 Application of ML in Civil Engineering

ML programs in the civil building included testing distinctive existing instruments on basic issues, continuously, more troublesome issues were tended to [6], and recently, the arrangement of few complex practical issues have been investigated [3], and architectural design [9]. In numerous early considers, as well as numerous modern, a single ML method has been employed. By and expansive, the determination of these strategies was based on accessibility and not essentially applicability of the ML method to the target tissue. Frequently, the issue representation utilized was a rearrangement driven by the restriction of the accessible ML procedure. There have been special cases to this practice. In a few cases, modern procedures or alterations of existing procedures were created to extend the appropriateness of ML procedures for architectural design in FABEL [9], or for observing water treatment plants. In other cases, a few strategies and imaginative information representations were utilized to address diverse varieties of learning issues (e.g., modeling material stress-strain relations [6]. Whereas tending to progressively complex issues, the need to integrate a few ML methods for understanding them was recognized and an introductory hypothetical foundation for such integration was created. A few ensuing frameworks that managed huge issues utilized numerous strategies. These frameworks too consolidated unused or altogether adjusted ML instruments. The part allocated to ML strategies in respectful building applications changed significantly. There has been considering on information extraction considers understanding total issues in which learning played a major part and thinks about that utilized learning as a portion of their operation (e.g., steel bridge plan [1], thruway truck stack checking [2], transmission line towers plan, and building plan [9]). In expansion, there have been thinking about coordinated data modeling for making estimation models and considers modeling pointed at moving forward the understanding of a wonder. The last-mentioned considers utilized different ML techniques. There are two issues that put work on ML in respectful building into point of view. To begin with considers to date ML applications in civil design have investigated a little number of ML procedures, most strikingly administered concept learning with few exemptions utilizing unsupervised learning (e.g., Bridger) or other procedures. Usually, it differentiates from the potential that numerous other ML methods [7]. In this way, the utilization of ML in civil engineering is as it were in its earliest stages. Second, numerous past thinks contained small or no precise testing and have had small or no follow-up work. This recommends that numerous of these considers were preparatory and did not develop. It moreover cautions us to fundamentally survey the conclusions of these things.

3 Development of Neural Network Model

Arrangements to numerous issues take a few steps driving from issue investigation to solution deployment. A few of these steps may be executed in parallel or indeed in turn around arrange and the method may repeat some time recently an effective and worthy arrangement is obtained. The taking after subsections details seven steps that methodically address the basic issues involved in building ML applications. These steps together could be a proposed method anticipated to lead to the improvement of effective ML applications.

Practical involvement in understanding issues utilizing ML methods and information around the properties of these procedures can reveal characteristics of issues and their mapping to appropriate ML strategies. Such mapping can be utilized to choose and apply ML procedures in a schedule design. We as of now specified a few considers coordinated at making such a mapping but in most cases, a clear determination and application will not suffice. Issues will be disentangled to coordinate the capabilities of ML methods (e.g., in Bridger as well as in most other thinks about), and arrangement strategies will be adjusted (e.g., in Bridger) or recently created and their utilize may hence be named as innovative or creative.

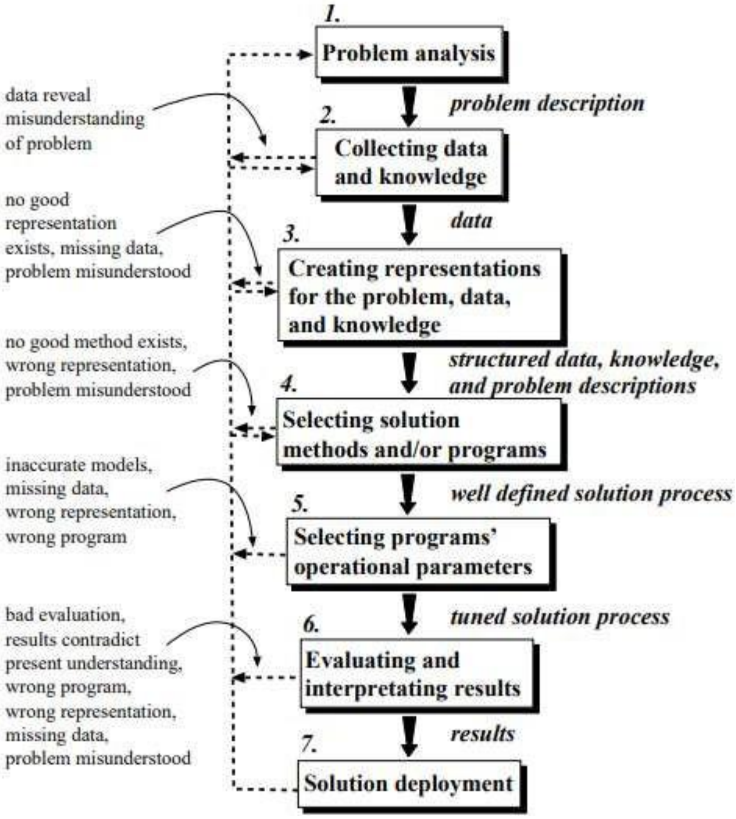

1.
Problem analysis
problem description
2.
Collecting data and knowledge
data
3.
Creating representations for the problem, data, and knowledge
structured data, knowledge, and problem descriptions
4.
Selecting solution methods and/or programs
well defined solution process
5.
Selecting programs' operational parameters
tuned solution process
6.
Evaluating and interpretating results
results
7.
Solution deployment
data reveal misunderstanding of problem
no good representation exists, missing data, problem misunderstood
no good method exists, wrong representation, problem misunderstood
inaccurate models, missing data, wrong representation, wrong program
bad evaluation, results contradict present understanding, wrong program, wrong representation, missing data, problem misunderstood

Fig. 1. A model of engineering problem solving

3.1 Problem analysis

Piles are a vital shape of deep foundation, utilized to transfer axial building loads through low-strength soil layers or bodies of water into more appropriate bearing strata. Pile foundations moreover are more beneficial than shallow foundations, indeed in the case of generally shallow load-bearing soil strata, for both financial and construction-related reasons. The assessment of the load-settlement execution of one pile is one of the foremost perspectives of the plan of piled foundations moreover; the behavior of a pile is impacted by a few components such as the mechanical non-linear behavior of the soil, the characteristics of the pile itself, still as its strategy of installation. To appraise the load-bearing capacity and settlement of piles, one or more of a few Pile Loading Tests (PLT) and Pile Dynamic Analysis (PDA) tests might indeed be performed, depending on the significance of a venture, since the high cost and time required for conducting such tests, it's a normal hone for engineers to estimate the load-bearing capacity of piles utilizing in-situ tests, just like the Cone Penetration Test (CPT), Standard Penetration Test (SPT), dilatometer test, and weight meter test, so to utilize reasonable proportion esteem during the arranging handle to accomplish a steady establishment. In any case, such a strategy is time-consuming, conjointly the costs are frequently troublesome to justify for standard or little ventures, while other strategies have lower accuracy. As a result, a few approaches are created to foresee the axial bearing capacity of piles or to boost the expected precision. The character of those strategies included a few simplifications, presumptions, or observational approaches with relevance to the soil stratigraphy, soil–pile structure interactions, and thus the dissemination of soil resistance alongside the pile.

In such considers, the test comes about was utilized as a complementary component to improve the forecast exactness. In later a long time, a far-reaching improvement in the utilization of data innovation in the civil building has cleared the way for a few promising applications, particularly the utilization of machine learning (ML) approaches to resolve viable designing issues. Additionally, distinctive ML methods are utilized, for case, the intelligent developmental approach, artificial neural network (ANN) and support vector machine (SVM) in tackling numerous real-world issues.

3.2 Collecting data and knowledge

The database was created from published literature sources that contain an add-up to 499 cases from 56 individual pile load tests. This consideration appears cross-validation, as clarified by is carried out to partition the information into three sets training, testing, and validation. Where the training set is utilized to acclimate the association weights, though the testing set is connected to check the translation of the show at different stages of preparing and to choose when to anticipate preparing to dodge overfitting [11]. The validation set is worked to assess the interpretation of the trained network within the arranged medium. On the complete, 90% of the data (450 cases) is worked for training and 10% (49 cases) are worked for validation. The preparing information is further broken up into 88% (395 cases) for the preparing set and 12% (55 cases) for the testing set. Since it's required that the information worked for training, testing and validation depict the same populace, Moreover, since the test set is worked to decide when to stop training, it has to be an agent of the training set and should in this way so also contain all of the designs.

3.3 Creating representations for the problem, data, and knowledge

To get precise forecasts of pile behavior (counting settlement and capacity), an understanding of the variables influencing pile behavior is required. Since pile behavior depends on soil quality and compressibility, and so the CPT is one among the first commonly utilized tests in hone for evaluating such soil characteristics, the CPT comes about (qc, fs) along the inserted length of the pile are utilized in this ponder. To depict more precisely the inconstancy of soil properties along the shaft of the pile, the implanted length of the heap is part into five portions of break even with thickness, with each related with a normal of qc and fs over that portion. The average of qc and fs (q¯c ; f s) for each subdivision, j, is calculated as below:

$$\overline{q_{cj}} = \frac{\sum q_{ci} Z_i}{\sum Z_i} \tag{1}$$

$$\overline{\mathrm{f}_{\mathrm{sj}}} = \frac{\sum \mathrm{f}_{\mathrm{si}}\, \mathrm{Z}_{\mathrm{i}}}{\sum \mathrm{Z}_{\mathrm{i}}} \tag{2}$$

Where qci and f si are the CPT estimations inside each portion and Zi is the soil layer thickness of layer i of fragment j. Subsequently, the different components which are displayed to the ANN within the frame of demonstrating input factors are (1) sort of test (kept up a stack or consistent rate entrance), (2) sort of pile (steel, concrete, and composite), (3) sort of installation (driven or bored), (4) conclusion of the pile (open or closed), (5) pivotal inflexibility of the pile (EA), (6) cross-sectional region of the conclusion of the pile (Atip), (7) border of the pile in contact with the soil (O), (8) length of the pile (L), (9) implanted length of the pile (Lembed), (10–19) the found the middle value of CPT comes about along the implanted length of the pile (qc1 , f s1 , qc2 , f s2 , qc3 , f s3 , qc4 , f s4 , qc5 , f s5), (20) cone tip resistance at the conclusion of the pile (qc tip), and (21) the connected stack (P). Pile settlement (sm) is the single yield variable.

3.4 Selecting solution methods and/or ML programs

An artificial neural network (ANN) or neuron arrangement may be a computing calculation. It ought to reenact the behavior of natural frameworks from "neurons". ANNs are a computational show propelled by the creature's central apprehensive framework. It is able of learning as well as design acknowledgment. These are displayed as interconnected "neuron" frameworks, which can calculate the value of the input. A neural network may be a coordinated chart. In organic similarity, it comprises nodes that speak to neurons associated with curves. Compares to dendrites and neural connections. Each circular segment has relegated a weight on each node. It applies the esteem gotten from the node as input and characterizes an activation function along the input circular segment, tuned by the bend weights. A Neuron network may be a machine learning algorithm formed on a demonstration of a human neuron. The human brain contains millions of neurons. It sends and forms signals within the outline of electrical and chemical signals. These neurons are associated with extraordinary structures known as synapses. Synapses empower neurons. From a huge number of recreated neuron neural network shapes. Artificial neural systems are data-preparing procedures. The human brain works as a way to handle data. ANN incorporates a number of related handling units that work together to handle data. They moreover create significant comes about. We cannot apply neuron systems to classify. It may moreover be connected to the relapse of nonstop target traits. A Neuron network could be a major application for information mining utilized in segments. For case, design acknowledgment, such as economic and scientific. After carefully preparing it, you'll be able to utilize it for information classification of an expansive sum of information. The neural network can contain three layers:

1. Input Layer - The action of the input unit speaks to the crude data that can be provided to the organization.
2. Hidden Layers - Decide the movement of each covered-up unit. The movement of the input unit and the weight of the association between the input unit and the covered-up unit. There may too be one or more covered-up layers.

3. Output Layer - The behavior of the output unit depends on the action of the covered-up unit and the weight between the covered-up unit and the yield unit.

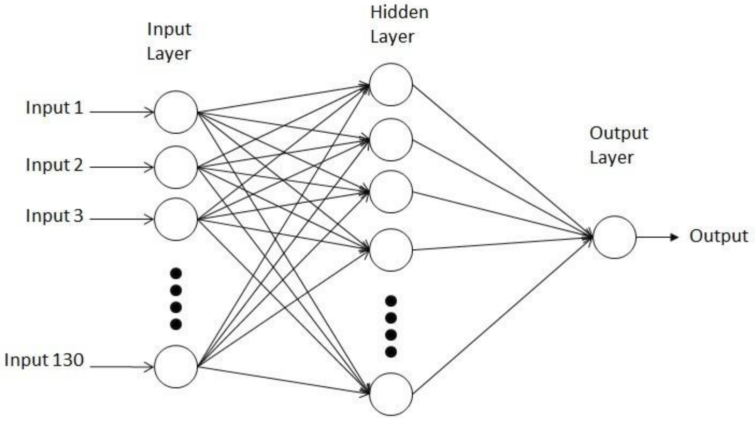


Fig. 2. The ANN structures

3.5 Selecting program operational parameters or options

Deciding the network surrounding is one of the foremost critical and sensitive errands in ANN demonstrate elaboration. It requires the choice of the ideal number of hidden layers and the number of nodes in each of these. There's no bound-together supposition for the choice of an ideal ANN surrounding. The number of nodes within the input and output layers is limited by the number of demonstrated inputs and outputs. An entirety of 21 input factors is carried in this consideration, and the output layer includes a single node characterizing the measured value of the settlement. In this consideration, model coordination of different hidden layers is inspected. In course of action to choose the optimum network figure, ANNs with one, two, and three hidden layers with different numbers of nodes within the hidden layers, for the multi-hidden layer models Rectified Linear Unit (ReLU) work is connected for the hidden layers.

Training. Training, or learning, is the operation of optimizing the association weights. Its objective is to recognize a universal arrangement to what's, by and large, a broadly non-linear optimization case. The technique most customarily utilized for finding the ideal weight combination of feed-forward neural systems is the back-propagation calculation, which is grounded on first-order angle plummet

Stopping criteria. Stopping criteria choose whether the demonstration has been ideally or sub-optimally prepared (Maier and Dandy, 2000). Various approaches can be worked to decide when to stop preparing. As said, to begin with, the cross-validation approach is worked in this work, as it's accounted that adequate information is reachable to deliver preparing, testing, and validation sets and it's the foremost valuable instrument to guarantee over-fitting does not happen [12]. The preparing set is connected to alter the association weights, though the testing set catalysts the capability of the show to generalize and, applying this set, the execution of the show is checked at various stages amid the preparing handle, and preparing is stopped when the testing set mistake starts to extend.

Model validation. Once demonstrate preparation has been effectively satisfied, the translation of the prepared show ought to be approved against information that has not been utilized within the learning preparation. The deliberate of the demonstration approval stage is to guarantee that the demonstration has the capability to generalize inside the limits set by the preparing information in a well-conditioned mold, instead of fair having memorized the input- output associations that are held within the preparing information

3.6 Evaluating, and interpreting results

As expressed already, in this consideration ANN models have been created with three hidden layers. In arrange to decide the ideal arrange geometry, ANNs are prepared with three hidden layers with diverse numbers of nodes within the hidden layers, it can be seen that the finest result is gotten by the three hidden layers show consolidating all input parameters, and 150-100-50 hubs within the three covered up layers separately. It is watched that demonstrate performs well. To ensure that the ANN demonstration is suitable, it is essential to look at its vigor over the total extent of the input and output information [11] characterized a strong ANN demonstration as one which shows smooth capacities with regard to the input and output factors and does not show behavior which cannot be clarified by a physical understanding of the framework being modeled.

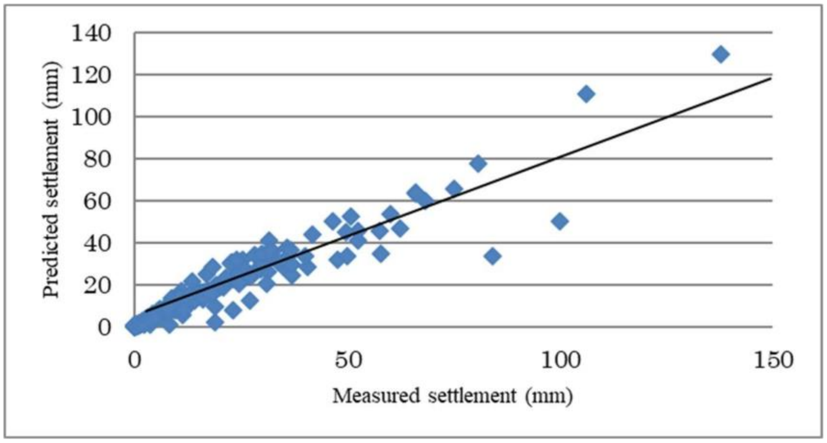


Fig. 2. Measured versus predicted settlements for ANN models with 3 hidden layers.

The plots of the measured versus anticipated settlement for the training set appear in Fig. 2. The comes about demonstrate that the demonstrate performs well, with r = 0.9, and MAE = 2.00 RMSE = 3.5 mm for the approval set. r = 0.93, and MAE = 1.2 RMSE = 4.01 mm for the preparing set and r = 0.86, and MAE = 1.8 RMSE = 3.87 mm for the testing set. Fig. 3 compares the anticipated load-settlement bends with the estimations gotten from the two pile stack tests. The comes about demonstrates that the show performs well for both the concrete pile, with r = 0.956 and RMSE = 4.39 mm, and the steel heap, with r = 0.98 and RMSE = 3.5 mm.

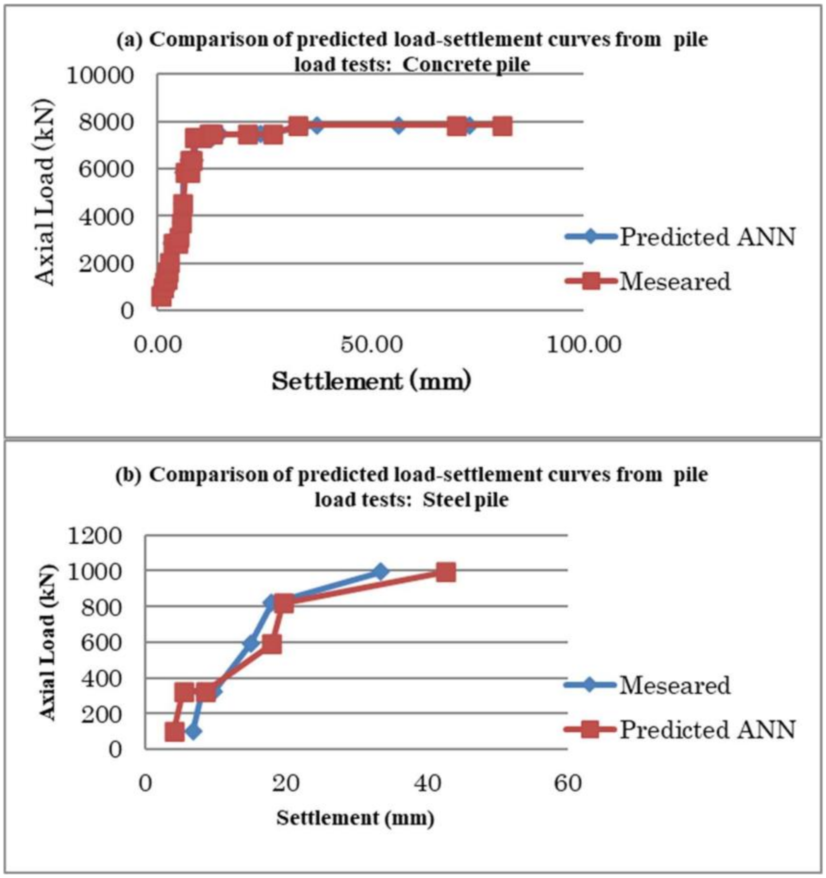


Fig. 3. Comparison of predicted load-settlement curves from two additional pile load tests: (a) A&M1-Concrete pile; (b) TWNTP4-Steel pile

3.7 Solution deployment

Linear regression examination is utilized to anticipate the value of one variable based on the value of another variable. The variable to foresee is named the subordinate variable. Factors utilized to anticipate the values of other factors are called independent factors. This frame of examination gauges the coefficients of a straight condition that contains one or more autonomous factors that best anticipate the esteem of the subordinate variable. Direct relapse fits a straight line or locale that minimizes the error between the predicted and real output values. There's a simple linear regression calculator that employments the least-squares method to find the most excellent line for a set of combined information. At that point assess the esteem of X (subordinate variable) from Y (free variable).

3.8 Conclusion

A back-propagation neural network has been utilized to consider the possibility of ANNs to foresee the load-settlement characteristics of piles. A database bearing 499 case records of field measures of heap settlements was worked to create and confirm the show. The comes about indicate that back-propagation neural systems have the capability to foresee the behavior of heaps with a respectable degree of exactness for settlements. The ANN approach incorporates an advanced advantage over customary approaches in that, once the demonstration is conditioned, it can be utilized as a correct and speedy instrument for assessing the behavior of piles. From the ideal demonstration, a few, stack- settlement charts for concrete bored piles of different lengths and distances across, presented in soil with an extended CPT value have been advertised back with a pile plan. In arrange to ease the spread and progressing headway of the ANN demonstrated.

It is frequently seen that the ANN strategy performs significantly superior to the conventional strategies by utilizing linear regression analysis in common, with ordinary strategies, the ANNs calculation, like numerous machine learning calculations, includes an auxiliary advantage that once the show is set up; it may be utilized as a correct, quick numerical instrument for assessing the bearing capacity of piles. Thus, the execution of comparative numerical apparatuses is basic in establishment designing. In like manner, overhauling the forecast precision is one viewpoint of the current work, for occurrence, utilizing Machine Learning calculations to anticipate the bearing capacity of piles.

References

1. Borner, K.: Modules for Design Support. Technical Report FABEL-report NO. 35. GMD, Sankt Augustin, Germany (1995).
2. Fan, C., Xiao, F., Zhao, Y.: A short-term building cooling load prediction method using deep learning algorithms, vol. 195, pp. 222-223, Appl. Energy (2017).
3. Debnath, P., Dey, A. K.: Prediction of Bearing Capacity of Geogrid-Reinforced Stone Columns Using Support Vector Regression, vol. 15, pp. 04017147(1-15). International Journal of Geomechanics (2018).
4. Ghorbani, B., Sadrossadat, E., Bazaz, J. B., Oskooei, P. R.: Numerical ANFIS-Based Formulation for Prediction of the Ultimate Axial Load Bearing Capacity of Piles through CPT Data, vol. 36, pp. 2057–2076, Geotechnical and Geological Engineering (2018).
5. Rashid, K. M, Louis, J.: Times-series data augmentation and deep learning for construction equipment activity recognition. vol. 42, pp. 100944. Adv. Eng. Inf (2019).

6. Kordjazi, A., Nejad, F. P. and Jaska, M. B.: Prediction of Ultimate Axial Load-Carrying Capacity of Piles Using A Support Vector Machine Based On CPT Data. vol. 55, pp. 91–102. Computers And Geotechnics (2014).
7. Li, M., Shen, Y. and Ren, Q.: A new distributed time series evolution prediction model for dam deformation based on constituent elements, vol. 39, pp. 41–52. Adv. Eng. Inf (2019).
8. Maier, H. R., Dandy, G. C.: Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications, vol. 15, pp. 101–24. Environmental Modelling and Software (2000).
9. Nath, N. D., Chaspari, T. Behzadan, A. H.: Automated ergonomic risk monitoring using body-mounted sensors and machine learning, vol. 38, pp. 514–526. Adv. Eng. Inf (2018)
10. Ren, Q., Li, M. and Zhang, M.: Prediction of ultimate axial capacity of square concrete-filled steel tubular short columns using a hybrid intelligent algorithm, vol. 14, pp. 2802–13. Appl. Sci (2019).
11. Shahin, M. A.: Load-settlement modelling of axially loaded steel driven piles using CPT-based recurrent neural networks, vol. 54, pp. 515–22. Soils and Foundations (2014).
12. Shahin, M. A., Maier, H. R. and Jaska, M. B.: Predicting settlements of shallow foundations using artificial neural networks, vol. 128, pp. 785–93. Journal of Geotech Geoenviron Engineering (2002).
13. Fang, W., Zhong, B. and Zhao, N.: A deep learning-based approach for mitigating falls from height with computer vision Convolutional neural network, vol. 39, pp. 170–177. Adv. Eng. Inf (2019).

2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)

Argument Mining using BERT and Self-Attention based Embeddings

Pranjal Srivastava
Department of Computer Science and Engineering
Delhi Technological University
New Delhi-110042, India
pranjaloct22@gmail.com

Pranav Bhatnagar
Department of Computer Science and Engineering
Delhi Technological University
New Delhi-110042, India
pranavbhatnagar2000@gmail.com

Anurag Goel
Department of Computer Science and Engineering
Delhi Technological University
New Delhi-110042, India
anurag@dtu.ac.in

Abstract— **Argument mining automatically identifies and extracts the structure of inference and reasoning conveyed in natural language arguments. To the best of our knowledge, most of the state-of-the-art works in this field have focused on using tree-like structures and linguistic modeling. But, these approaches are not able to model more complex structures which are often found in online forums and real world argumentation structures. In this paper, a novel methodology for argument mining is proposed which employs attention-based embeddings for link prediction to model the causational hierarchies in typical argument structures prevalent in online discourse.**

Keywords—Argument Mining, Transformer, Self-Attention, BERT

Introduction

Human communication is fundamentally composed of debate and argument. With online forums increasingly serving as the primary medium for discourse and discussion, the importance of automated data processing is increasing rapidly. There are various data science techniques which proved to be successful in these natural language processing tasks. But, still there is a lot of scope of research in identifying the more complex structural relationships between concepts.

The theory of argumentation and the use of logical reasoning to justify claims and conclusions is an extensively studied field, but using data science techniques to automate the process is relatively new. A prevalent practice in the initial work in argument mining is to represent the argument structure using one or more trees or tree-like structures. This provided ease of computation as various techniques existed for tress and tree-related parsing, but arguments in the real world rarely follow the ideal system that these methods imposed.

In recent times, there have been many methods to explore argument mining in the wild using argument structures that are not required to be tree-based. Architectures like Recurrent Neural Networks, Convoluted Neural Networks, Long Short-Term Memory, and Attention-based mechanisms have allowed us to leverage contextual information in making informed machine decisions. Most recently, transformer-based architectures have given state-of-the-art performance in various Natural Language Processing (NLP) related tasks. It uses attention to boost the speed of tasks. We attempt to use the same in argument mining.

The recent trend in NLP is leveraging Transfer Learning on huge pre-trained models for better performance. Transfer Learning is a technique that was instrumental in the advancements in the domain of computer vision. It was popularized in NLP in 2018 when Google released the transformer model. Since then, transfer learning in natural language processing has aided in solving several tasks with state-of-the-art performance.

We use the Cornell eRulemaking Corpus: Consumer Debt Collection Practices (CDCP) [16], a collection of argument annotations on comments from an eRule-making discussion forum, where the argumentative structures do not necessarily form trees. We use a language representation model called Bidirectional Encoder Representations from Transformers (BERT). We also use the transformer encoder layer and generate embeddings from the encodings that capture the hierarchical relations between argument components.

Related Work

A. Argument Mining

Argument mining has been a problem that has attracted a lot of research interest. Moens et al. [1] attempted to identify features like n-grams, keywords, parts of speech, etc., to classify argumentative text in a collection of legal texts. Levy et al. [2] proposed a three-step approach for context-dependent claim detection. In [3], the authors presented an end-to-end approach to model the arguments and their relations in corpora. In this work, the authors proposed a pipeline of three steps: Argument detection, Argument proposition classification, and Detection of the argumentation structure. They utilized manual Context Free Grammar rules to predict a tree like relation structure between the arguments.

Stab and Gurevych [4] defined three major subroutines for effective argument mining:

1. Component Identification: Separation of argumentative spans of text from non-argumentative text for a given corpus.
2. Component Classification: Identification of the various types of argument components.
3. Argument Relation Prediction: Linking of the different parts of arguments and identification of logical dependencies.

ISBN: 978-X-XXXX-XXXX-X/22/$31.00 ©2022 IEEE

The preliminary work in Argument Mining revolved around modeling the argument structures as trees [4, 5, 6]. These works allowed using maximum spanning tree-like parsing methods for dependency mining, enhancing the computation speed and ease. But, this failed to consider the more complex and divergent graph structures that could be found in the discourse, available on resources like online forums and discussion threads. Niculae et al. [7] proposed the first non-tree argument mining approach with a factor graph model utilizing structured Support Vector Machines (SVMs) and Bidirectional Long Short Term Memory (LSTM). Galassi et al. [8] explored the LSTMs and residual network connections to focus on link prediction between argument components. Morio et al. [9] proposed another approach that utilized Task-Specific Parameterization (TSP) to encode the sequences of propositions and a Proposition-Level Biaffine Attention (PLBA) to predict non-tree arguments with boosted edge prediction performance.

B. BERT

Vaswani et al. [10] first proposed the transformer architecture, with its self-attention mechanism, to counter the memory and processing cost of Recurrent Neural Networks (RNNs). Both of these architectures deal with sequential data to model global dependencies. Transformers can be trained in a highly parallelizable mode compared to the innate sequential nature of RNNs.

Devlin et al. [11] introduced Bidirectional Encoder Representations from Transformers (BERT) in a breakthrough paper. This model uses pre-train deep bidirectional representations from the unlabeled text. The pre-trained model is fine-tuned with one additional layer to create cutting-edge models for a diverse range of problems.

Reimers et al. [12] utilized BERT and ELMo (Peters et al. [13]) in an open-domain argument search to classify and cluster topic-dependent arguments, with excellent results on the UKP Sentential Argument Mining Corpus and the IBM Debater - Evidence Sentences dataset. Chakrabarty et al. [14] developed a novel approach of two fine-tuning steps on BERT, the first on a distant-labeled dataset and then on the labeled persuasive forum dataset. Their approach obtained significant improvements in comparison to other state-of-the-art techniques. Ting Chen [15] proposed an approach based on BERT and Proposition-Level Biaffine Attention (PLBA) that achieved good results.

PREREQUISITES

A. Problem Formulation

Our inputs will be the annotated text of a user remark, with each annotation indicating an argument. Each component of the argument corresponds to a specific span specified in the annotated text. Thus, the outputs of our model will include the argument proposition type associated with each span and the outgoing edges connecting the span to other components of the argument. Fig. 1 depicts this modeling of the problem.

B. Embeddings

The high-dimensional vectors can be translated into the relatively low-dimensional space known as Embeddings. Thus, it becomes easier to apply the models on the embeddings. In embeddings, the semantically similar inputs are placed together to capture the semantics of the input. Embeddings can be learned and reused across the models. The major inspiration comes from word embeddings used extensively in NLP tasks. These embeddings have been shown to capture semantic information about words and model relationships easily. We have an embedding vector and Context vector that we use to depict the elements and learn the underlying feature representations in our space.

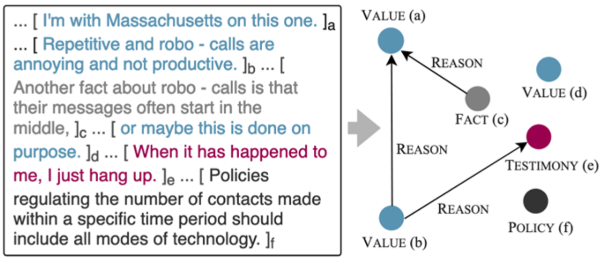

... [I'm with Massachusetts on this one.]a
... [Repetitive and robo - calls are annoying and not productive.]b ... [
Another fact about robo - calls is that their messages often start in the middle,]c ... [or maybe this is done on purpose.]d ... [When it has happened to me, I just hang up.]e ... [Policies regulating the number of contacts made within a specific time period should include all modes of technology.]f
VALUE (a)
REASON
VALUE (d)
FACT (c)
REASON
TESTIMONY (e)
REASON
POLICY (f)
VALUE (b)

Fig. 1. Problem Formulation

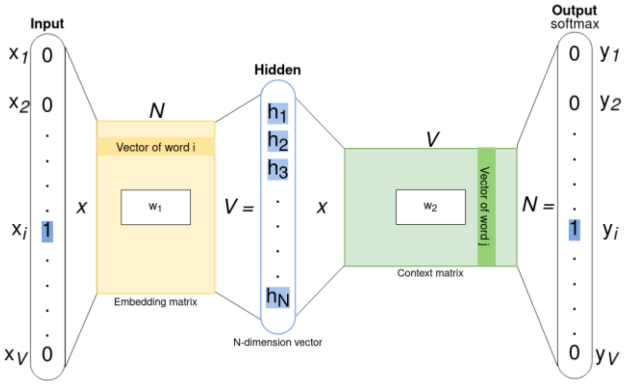


Fig. 2. Embeddings

C. Transformer

The transformer architecture and its self-attention mechanism were originally proposed in response to the growing computational and memory requirements of state-of-the-art Recurrent Neural Networks (RNN). By using just multi-head self-attention, transformers allow for significantly more parallel training, as opposed to the inherent sequential nature of RNN.

In our approach, we use BERT to get a sentence level embeddings for our argument components. We use these to create a sentence level representation that helps with classification as argument component type as well as downstream task of edge detection. Then, we use these encodings to create further encodings that capture context of the sentence in the argument. We use a Transformer Encoder Layer in the architecture to do so. The transformer has the following architecture as depicted in Fig. 4. The left layer is the encoding layer and the right layer is the decoding layer. We use the encoding layer in our approach.

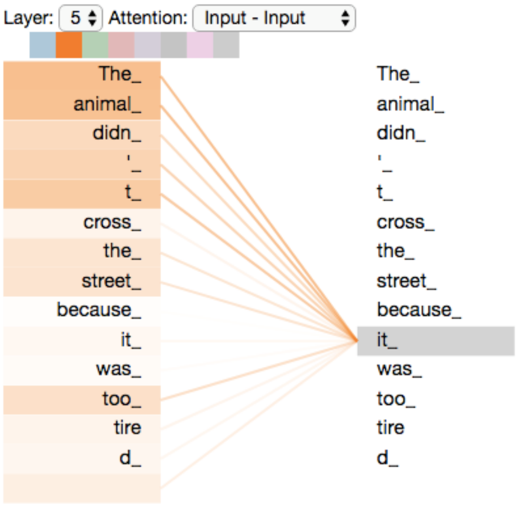


Fig. 3. Attention to context

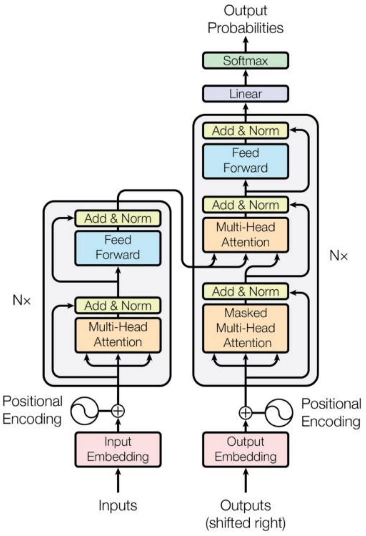


Fig. 4. Transformer

D. BERT

In our approach to argument mining, we look rely on power of the contextual word embeddings derived from BERT. In our case we tested the 'bert-base-uncased' model from the Huggingface transformers library. BERT is essentially made up of 12 transformer encoder layers stacked on top of each other. The primary advantage of using a Transformer-based model with attention over a traditional RNN is that the attention mechanism enables the model to simultaneously see all of the words and choose which ones are most important for the given task, whereas RNN typically see words sequentially. This allows us to take advantage of computation parallelism, allowing for a model that is more efficient. BERT is pre-trained on a English corpus containing 3.3 billion words and utilising Masked Language Modelling (MLM) and Next Sentence Prediction (NSP) tasks.

1. Masked Language Modelling (MLM) – This model predicts a masked word in the sentence. About 15% of the words in a given input sentence are randomly masked for this task.
2. Next Sentence Prediction (NSP) - This model predicts whether two randomly chosen masked sentences naturally follow each other. These tasks have the unique ability to look at all the words in the sentence at once.

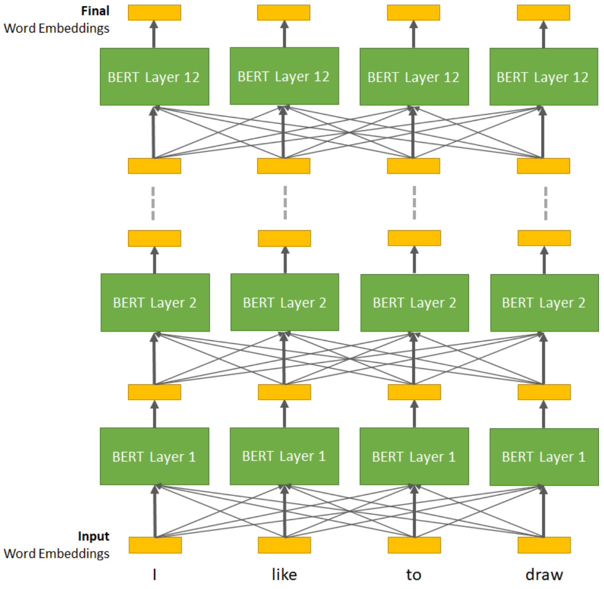


Fig. 5. BERT Architecture

PROPOSED APPROACH

The proposed approach is divided into two phases:

Phase I: Classification of argument components

Phase II: Edge detection on encodings

I. Classification of Argument Components

In this phase, the argument spans are classified into types of argument components using BERT. The pre-trained BERT model is fine-tuned on the CDCP corpus using a supervised learning goal. Then, transfer learning is applied to the resultant pre-trained BERT model.

Since BERT produces a collection of embeddings, the classifier uses the first element of these embeddings, the special token [CLS] used to pool sentence-level semantic features. When applying BERT to a classification task, we always use the final embedding for the [CLS] token as the input to our classifier and ignore the individual token embeddings. We use the [CLS] token not just for classification but also for producing sentence-level embeddings by pooling the [CLS] output of the hidden layers for the next task of edge detection.

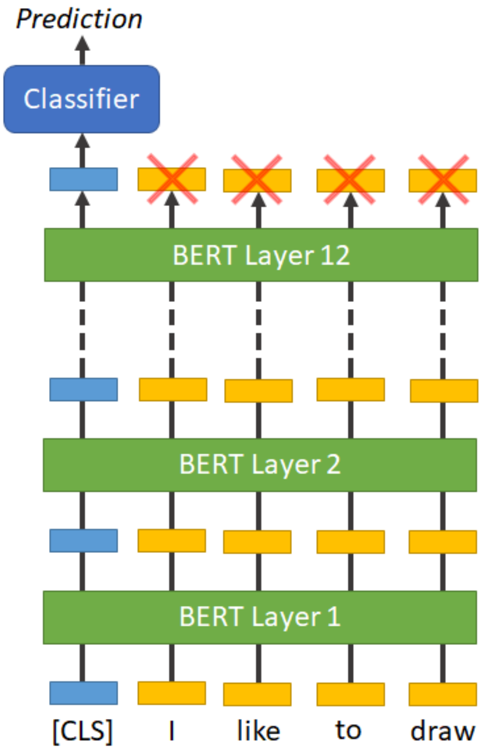


Fig. 6. Classification

II. Edge Detection on Encodings

We take a weighted average of the last n layers of the hidden layer outputs to make the embeddings. This is given as:

$$e_i = \sum_{j=1}^{n} w_j * E_j$$

where e_i is the sentence encoding for i^{th} sentence, w_j the weight for the j^{th} layer and E_j = [Emb_1 ... Emb_n] is the embedding of layer j of hidden outputs.

We pass these embeddings through a transformer encoder layer that allows a context based embedding for sentences based on the surrounding argument components. These embeddings z_i are then projected to a lower dimensional space through two different projection matrices, C and P for projection into Conclusion and Projection spaces.

$$Conclusion\_emb_i = C \cdot e\_transformer_i$$
$$Premise\_emb_i = P \cdot e\_transformer_i$$

We model the following relation in the model after normalizing the projections.

$$Conclusion\_emb_i \cdot Premise\_emb_j = \begin{cases} 1, if j \Rightarrow i \\ 0, if j \nRightarrow i \end{cases}$$

where $j \Rightarrow i$ signifies that *sentence* i is the conclusion and *sentence* j is the premise.

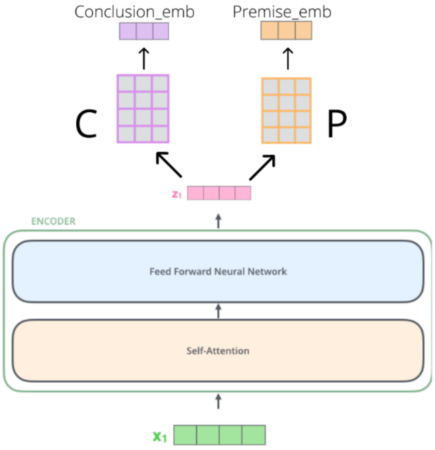


Fig. 7. Edge Detection

Experimentation and Results

I. Dataset Used

For this work, we use the Cornell eRulemaking Corpus (CDCP) proposed by Park and Cardie [16], which models argument relations as links in a directed graph. It comprises about 731 user comments from an online eRulemaking forum. It has over 4500 propositions, with about 88,000 words. The propositions in the corpus are of five types:

i) Value (45%): Values are propositions that contain value judgments without specific claims or suggestions.

ii) Policy (17%): Policy is a proposition that directs a specific direction of action.

iii) Reference (1%): Reference is citing some source to support a position.

iv) Fact (16%): Fact is an assertion that can be verified with objective evidence.

v) Testimony (21%): Testimony is an objective proposition stemming from the author's experiential knowledge.

Predicting the argument structure reduces to imbalanced link prediction on the dataset, with about 3% of pages being linked.

TABLE I. CDCP Classification Statistics

Type	# in Training Set	# in Test Set
Propositions	3698	1233
VALUE	1633	544
POLICY	611	204
REFERENCE	24	8
FACT	592	197
TESTIMONY	838	280

II. Experimental Setup

The dataset is divided into 75% and 25% as training and testing sets, respectively. Table 1 depicts the relevant split for our application. We trained the BERT model for type classification on spans of argumentative propositions. As is the norm with transfer learning, the pre-trained model has trained for task-specific optimization over three epochs using the Adam optimizer with weight decay [17] with an initial learning rate of 5×10^{-3}.

For edge detection, the last $n = 4$ layers from the BERT classification model are considered for weighted accumulation. The weights were sampled from an arithmetic progression for best results. It passes through a transformer encoder layer of dimension 768×768 which corresponds to the output dimensions of BERT, followed by projection into a 100-feature embedding space for Premise and Conclusion embeddings, respectively. We train the link prediction model for 100 epochs with AdamW optimizer with 0.4 dropout probability and early stopping for regularization.

III. Results Analysis

We evaluate the F1 score for classification and link prediction for convenient comparative analysis. F1 score is the harmonic mean of Precision and Recall and is a suitable metric considering the imbalanced nature of precisely the link-prediction task. The results are shown in Table II. The results depict that the proposed approach outperforms all the benchmarks except the approach that utilized Task-Specific Parameterization and Proposition-Level Bi-Affine Attention (TSP+PLBA) [9]. The proposed approach gives a comparable performance with [9].

TABLE II. RESULTS

Model	Edge Prediction	Type Prediction	Average
Deep Basic: PG [8]	0.22	0.63	0.43
Deep Residual : LG [8]	0.29	0.65	0.47
RNN : Basic [7]	0.14	0.73	0.44
SVM : Strict [7]	0.27	0.73	0.50
TSP+PLBA [9]	0.34	0.79	0.56
BERT+MLP/PLBA [15]	0.15	0.86	0.51
Ours Approach	0.25	0.81	0.52

CONCLUSION

This paper presented an attention-based approach to model causational hierarchies in typical argument structures in online discourse. The proposed approach uses BERT to get a collection of embeddings, which are then passed through a transformer encoder layer to discover edges between them. The experimental results show that the proposed approach outperforms most of the baselines and most contemporary approaches.

REFERENCES

[1] Moens, M.-F., Boiy, E., Palau, R.M., Reed, C.: Automatic Detection of arguments in legal texts. In: Proceedings of the 11th International Conference on Artificial Intelligence and Law, pp. 225–230 (2007)

[2] Levy, R., Bilu, Y., Hershcovich, D., Aharoni, E., Slonim, N.: Context dependent claim detection. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pp. 1489–1500 (2014)

[3] Palau, R.M., Moens, M.-F.: Argumentation mining: the Detection, classification and structure of arguments in text. In: Proceedings of the 12th International Conference on Artificial Intelligence and Law, pp. 98–107 (2009)

[4] Stab, C., Gurevych, I.: Annotating argument components and relations in persuasive essays. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pp. 1501–1510 (2014)

[5] Peldszus, A., Stede, M.: Joint prediction in mst-style discourse parsing for argumentation mining. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 938–948 (2015)

[6] Stab, C., Gurevych, I.: Parsing argumentation structures in persuasive essays. Computational Linguistics 43(3), 619–659 (2017)

[7] Niculae, V., Park, J., Cardie, C.: Argument mining with structured svms and rnns. arXiv preprint arXiv:1704.06869 (2017)

[8] Galassi, A., Lippi, M., Torroni, P.: Argumentative link prediction using residual networks and multi-objective learning. In: Proceedings of the 5th Workshop on Argument Mining, pp. 1–10 (2018)

[9] Morio, G., Ozaki, H., Morishita, T., Koreeda, Y., Yanai, K.: Towards better non-tree argument mining: Proposition-level biaffine parsing with task-specific parameterization. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 3259–3266 (2020)

[10] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems 30 (2017)

[11] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)

[12] Reimers, N., Schiller, B., Beck, T., Daxenberger, J., Stab, C., Gurevych, I.: Classification and clustering of arguments with contextualized word embeddings. arXiv preprint arXiv:1906.09821 (2019)

[13] Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations (2018)

[14] Chakrabarty, T., Hidey, C., Muresan, S., McKeown, K., Hwang, A.: Ampersand: Argument mining for persuasive online discussions. arXiv preprint arXiv:2004.14677 (2020)

[15] Chen, Ting, "BERT Argues: How Attention Informs Argument Mining" (2021). *Honors Theses.* 1589 https://scholarship.richmond.edu/honors-theses/1589

[16] Park, J., Cardie, C.: A corpus of erulemaking user comments for measuring evaluability of arguments. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) (2018)

[17] Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)

Artificial Intelligence driven Intrusion Detection Framework for the Internet of Medical Things

Prashant Giridhar Shambharkar
Delhi Technological University
Nikhil Sharma (✉ nikhilsharma1694@gmail.com)
Delhi Technological University

Research Article

Keywords: Artificial Intelligence (AI), Cyber Attacks, Deep Learning (DL), Healthcare, Internet of Medical Things (IoMT), Intrusion Detection System (IDS), Machine Learning (ML), Security

Posted Date: March 14th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2634004/v1

Abstract

The fusion of the internet of things (IoT) in the healthcare discipline has appreciably improved the medical treatment and operations activities of patients. Using the Internet of Medical Things (IoMT) technology, a doctor may treat more patients and save lives by employing real-time patient monitoring (RPM) and outlying diagnostics. Despite the many advantages, cyber-attacks on linked healthcare equipment can jeopardize privacy and even endanger the patient's health. However, it is a significant problem to offer high-safety attributes that ensure the secrecy and accuracy of patient health data. Any modification to the data might impact how the patients are treated, resulting in human fatalities under emergency circumstances. To assure patients' data safety and privacy in the network, and to meet the enormous requirement for IoMT devices with efficient healthcare services for the huge population, a secured robust model is necessary. Artificial Intelligence (AI) based approaches like Machine Learning (ML), and Deep Learning (DL) have the potential to be useful methodology for intrusion detection because of the high dynamicity and enormous dimensionality of the data used in such systems. In this paper, three DL models have been proposed to build an intrusion detection system (IDS) for IoMT network. With a 100% accuracy rate, our proposed AI models exceed the current existing methodology in detecting network intrusions by utilizing the patient’s biometric data features with network traffic flow. Furthermore, a thorough examination of employing several ML and DL approaches has been discussed for detecting intrusion in the IoMT network.

1. Introduction

The latest advancement and innovations in developing IoT systems have revolutionized the industry as internet-connected devices are increasing rapidly. According to Gartner, by 2025, there will be 27 billion IoT devices, which will be twice as many as the number of internet-connected intelligent devices today [1]. It revolutionized the healthcare industry by integrating intelligence accomplishment into medical devices. Nowadays, Healthcare monitoring systems have been designed using IoT as they have low-cost sensors and consume low power. In recent years, these sensors have been used to expedite the patient’s remote monitoring. It also lessens the requirement for doctors being physically presented every time. The fusion of IoT into medical devices is known as the Internet of Medical Things (IoMT). Nearly 30% of the IoT devices market has been covered by IoMT [2]. The modern development and trends in the field of wireless communication and IoT have widely supported an immense range of healthcare applications like patients’ health monitoring, early detection and diagnosis of various diseases, remote healthcare services, and healthcare emergencies. Adapting practical and secure techniques for life-threatening (critical) emergency cases can lessen the dependency and need on caregivers and cheapen medical service costs. The development of intelligent decisiveness systems can facilitate early cures and possibly saves people’s life. Wearable devices are employed to examine the patient health of the community continuously. So, the caregivers can impart communication regarding monitoring and diagnosis services to the people belonging to smart societies. It may create a severe problem if these systems are affected by any security threats because this gives rise to a contravention of patients’ health data, and medical and privacy issues, which may also result in the death of those patients in extreme cases [3]. According to the Cynerio report, most of the IoT-connected healthcare equipment may clasp critical susceptibility [4]. The unexpected disclosure of these gadgets in public may give rise to uncertainty risks and aid the attackers in accessing the patient’s records. The adversaries may also use the known susceptibility and Advanced Persistent Threats (APT) to deal with the victim device. This might endanger people's lives. Therefore, security must contemplate as a top concern in an IoMT-based healthcare monitoring system. Using different methodologies and techniques, the detection and prevention of attacks can be implemented. Other attacks, namely Denial of Services (DoS), malicious network traffic injection, and the Man in the Middle (MITM), are accomplished to compromise and attack the IoMT networks. Various techniques like Intrusion Detection Systems (IDS), vulnerability management, log monitoring, end device monitoring, prevention systems, and threat intelligence are used for detection and mitigation purposes. Among these systems, IDS is a modern technology used to recognize various network attacks and security issues in IoMT.

In IDS, signature-based rules, security policies, or network traffic anomalies are executed to recognize security attacks in IoT-based networks. As the attackers uses modern hacking methods and updated attacking strategies, these traditional policies were inadequate to secure the system. For example, the attacker can easily evade security policies by performing network reconnaissance, and configuring firewalls and routers by reverse engineering the network devices. Researchers are investigating ML, and DL approaches as they gained a vital place in offering solutions for developing efficacious attack detection systems. Exposure to processing and computing capabilities permits researchers to use ML and DL approaches as they give better results in precisely forecasting attack events.

1.1 Motivation

Mainly, the research was carried out on traditional network-based datasets containing network flow traffic samples only. In [5], an intelligent IDS framework has been developed using ML, and DL techniques for identifying cyber-attacks in traditional networks. However, these approaches are inappropriate in the IoMT as, in today’s scenario, different IoT-enabled health sensors are linked to the internet. Even for evaluating attack detection, the traditional network datasets are unsatisfactory in IoMT. To recognize IoMT attacks in intelligent healthcare applications, most research emphasizes evaluating the network traffic [6–7]. In healthcare applications, patient biometric data is essential to understand their health conditions better. There is a connection between unexpected falls in the attack impacting the network and patients sensing information which affects the availability, confidentiality, and integrity of healthcare records. Therefore, in this paper, dataset [8] containing both patient biometrics and network traffic samples has been considered to forecast the attack incident and evaluate the two diverse data types when the attack incident takes place. The security violation in the network or IoMT devices may cause intervention in obstructed communication, loss of information related to patient’s health, and disease diagnosis. As the attackers are so expert and skilled, the security risks to IoMT devices are increasing daily. As far as our knowledge, we contemplate this research as the first effort to use DL-based embedding approach to recognize intrusion in the IoMT environment that has bypassed the current state of the art performance. The major contribution of this work is explained as follows:

- Demonstrated various preprocessing and feature engineering techniques that includes dropping irrelevant columns with the help of a covariance matrix as these features may not have a significant effect on the detecting process; To provide enough data availability for both the normal and attack samples, the input dataset is balanced using data augmentation approach; Categorized the dataset features into categorical and continuous columns, and then applied StandardScalar along with Power transformer for continuous columns, and ordinal encoders for categorical columns.

- Proposed three deep learning models such as LinSVM (Linear_Support Vector Machine), ConvSVM (Convolutional_ Support Vector Machine), and CatEmbedding (Categorical Embedding) models to identify IoMT attacks using a dataset that includes network traffic flow data and biometric features of patients.
- Analyzed the intrusion detection performance of divers ML and DL methods exhaustively to enhance the IoMT IDS's ability to identify attacks in the network.
- The proposed models have been contrasted with the existing techniques explored on the same dataset with an accuracy of 100%.

The rest of the paper is systematically categorized as follows: Section 2 highlights the literature review, which covers the capability of ML and DL models on different network flow dataset; Section 3 elucidates about the dataset background; Section 4 explains the architecture and workflow of the proposed methodology for intrusions detection in the IoMT ecosystem; Section 5 illustrates about the experimental work and in-depth performance evaluation of proposed models with the existing work followed by the conclusion in Section 6.

2. Literature Review

Many researchers have proposed several methods for designing smart health monitoring systems. Fotouhi et al. [3] developed a healthcare monitoring framework comprising three integrants, i.e., a gateway, an Access Points (APS), and a coordinator. The coordinator is defined as the node that is attached to the body of the patients for gathering information related to patient’s health using sensors. In the room’s walls, static nodes, also called access points, are located by the sensors that use communication protocols (like 6LoWPAN, BLE, or ZigBee). The information gathered by the APs gets forwarded to the gateway. Then using the internet, this information gets transferred to the cloud server. Without proper testing and explanation, some techniques have been proposed in this system for securing the data. Clifton et al. [9] explained the ML technique’s role in health monitoring systems. These techniques have been used for controlling and managing false alerts while revealing serious health issues. The data used in their experiment is a combination of the patient’s clinical observations to provide quick alerts in an envisaged emergency. This work has been conducted at Oxford University Hospital.

Rani et al. [10] introduced a cloud-based healthcare platform that uses an SVM (Support Vector Machine) approach to forecast patients’ situations and envisioned diseases. No unauthorized users are permitted in the system. Blockchain-based healthcare system framework has been designed by Chakraborty et al. [11], which is helpful in overcoming the problems associated with the traditional healthcare system related to the security issues of the records created during the treatment of the patient. The framework has been set up for supervising the treatment process all over the time from beginning to end.

Alabdulatif et al. [12] developed a cloud-based smart prediction framework. This system was based on Fully Homomorphic Encryption (FHE) approach. This system comprises three blocks, i.e., the smart community resident, cloud storage, and smart prediction model. In the first block, the data is gathered and dispatched to the cloud storage repository system. The encrypted data gets amassed in the second block. The third block has been used to detect anomalous changes like attacks without decrypting data. A secured anticipating approach based on Holt's linear trend method has been developed that is used to predict anomalous changes in the vital sign of patients, which helps to detect different chronic diseases. The author also introduced a novel parallel technique of Holt's method for improving the effectiveness of the FHE model. Tao et al. [13] introduced a SecureData scheme that delivers both privacy and security to the patient's private records. FPGA (Field Programmable Gate Array) platform is used for the optimization of the KATAN secret algorithm, which is implemented for secure communication.

A security system has been suggested by Zhang et al. [14] that applied RF (Random Forest) technique to detect anomaly traffic on KDD 1999 dataset. With a 1% false positive rate, this technique achieves 95% accuracy as an anomaly detector. This dataset is employed to test anomaly detection algorithms. It is a generic Knowledge discovery and data mining dataset. Since 1999, this dataset has been used in many competitions. Rao et al. [15] employed the Indexed Partial Distance Search k-Nearest Neighbor (IPDS-KNN) technique that is employed to assess a diverse variety of attacks. It achieved 99.6% of accuracy performance. Shapoorifard et al. [16] use the k-Nearest Neighbor (KNN) technique, which achieves 85.2% of accuracy. The author mainly emphasizes on decreasing the False Alarm Rate (FAR). To forecast various attack simulations in Deep Brain Stimulators (DBSs), Rathore et al. [17] developed a DL algorithm that efficiently identifies the pattern of attacks and alerts a patient regarding that.

To overcome the attack detection problem in IoMT, Yaacoub et al. [18] discussed different types of ML-based privacy and security solutions. But, according to the author, there is still a need to introduce an effective IDS system to detect attacks. To analyze attacks in the smart hospital, an ensemble classifier IDS has been developed by T. Saba [19]. The Decision Tree (DT) technique attained 93.2% of accuracy performance in categorizing the cyber-attacks in the KDDcup-99 dataset. This dataset was created in the traditional network without adding IoT device traffic. Kumar et al. [20] performed the experiment in three stages. In the first stage, the author introduced an ensemble of the RF, naïve Bayes (NB), and DT. In the second stage, to categorize both regular and attack network records, XGBoost was applied. In the third stage, to categorize the attacks in the IoMT environment, the developed model was then applied to the ToN-IoT dataset, which attained 96.35% of the accuracy. The industrial IoT network setup has been used to create this dataset using Modbus weather sensors. In the IoMT environment, these sensors are not commonly employed. Therefore, the data presented above could not be appropriate for identifying network attacks.

Radoglou et al. [21] developed an Intrusion detection and prevention system (IDPS) for the identification and prevention of various cyberattacks against communication protocols like Modbus/TCP and HTTP, which are broadly used by e-healthcare services. EHR uses HTTP, whereas IoMT uses Modbus/TCP protocol. The proposed IDPS can retrain ML techniques and test itself using an active learning approach. The CIC-IDS2017 dataset was employed in this experiment to analyze the functioning of ML techniques on the

HTTP network dataset. DT classifier achieved an accuracy of 96.44% in categorizing network attacks. In comparison, RF attained 94.45% of the accuracy on the Modbus dataset.

Zachos et al. [22] introduced a systematic and potent Anomaly-based IDS (AIDS) for the IoMT environment. To devise a unique feature set, the three features, i.e., gateways, IoT device features, and network traffic features, were combined together. To enhance the functionality of attack detection, various ML techniques have been applied to identify deviations in the gathered malicious and data events in the network. For evaluation in IoT devices, memory consumption level attributes, and CPU were taken into consideration. The TON_IoT Telemetry dataset has been used in this experiment. According to the result reported by the author, KNN, RF, and DT are the most appropriate ML technique that is employed for the central detection integrant of the introduced system.

A mobile agent-based IDS has been introduced by Thamilarasu et al. [23] to identify both network and device-based attacks in the IoMT environment. The simulation-generated datasets were tested using ML and regression techniques. Using the DT technique in the evaluation process, the device and network-level intrusion detection achieve an accuracy of 97.93% and 99.8%, respectively.

Binbusayyis et al. [24] inspected and showed a detailed comparison of different techniques like KNN, SVM, ANN (Artificial Neural Network), NB, and DT. The Bot-IoT dataset was used in the experiment to compare the working performance of ML methods. This dataset comprises various attack categories like Denial of service (DoS), theft attacks, and Distributed Denial of Service (DDoS) attacks. Spoofing attacks and MITM attacks are IoMT attacks that are not covered in this dataset. On the tested dataset, DT attained an accuracy of 100%, and other ML techniques like SVM, NB, and KNN achieved an accuracy of 99%.

As per the study, ML techniques are used to identify attacks in IoMT. But most of the datasets were created without considering the IoMT environment and attacks. The result presented by the authors in their research were outstanding, as in many contributions, the ML techniques achieved an accuracy of 95%. For the IoMT study, many input features like IoT device memory, network traffic, CPU features, or metric features were considered. But features like patient biometric data were not used or mentioned by any researchers in their work to identify cyber-attacks in the IoMT. To classify or identify the attack in the IoMT ecosystem, many researchers explored DL techniques.

For feature selection, Saheed et al. [25] used Particle swarm optimization (PSO) and applied ML/DL-based techniques to identify cyber-attacks in the network. Researchers used the NSL-KDD dataset to analyze the functionality of the suggest technique. The introduced model attained 99.76% of the accuracy performance. This dataset was not created by keeping the IoT environment in mind and should not be used to assess attack identification in IoMT.

Awotunde et al. [26] developed a swarm neural network (SNN)-based method that detects intruders while transmitting the data and permits accurate and efficient assessment of medical data at the network edge. For the experiment, the author used NF-ToN-IoT dataset, which is the amalgamation of network data, operating systems, and telemetry. The author used a deep autoencoder (DAE) to decrease the dimensions of features. To recognize the network attacks, the author used a deep feed-forward neural network (DFFNN) in the IoT environment. The DAE-DFFNN model achieved an accuracy of 89%, which is superior to ML techniques such as DT and SVM claimed by the researcher.

For identifying malware in the IoMT ecosystem, Khan et al. [27] introduced SDN (Software Defined Network) enabled CNN (Convolutional Neural Network) and LSTM (Long short-term memory) hybrid DL model, which attained an accuracy of 99%. Howbeit, this framework was not used as IDS to determine network attacks in IoMT ecosystem. Nandy et al. [28] developed intelligent agent-based SNN for detecting intruders in IoMT. The experiment was conducted using the proposed approach on the ToN-IoT dataset, which attained 99.5% of the accuracy. To identify the network attack in the IoT environment, Manimurugan et al. [29] introduced a DL-based deep belief network (DBN) algorithm that achieved an accuracy of 96%. The experiment was conducted using the proposed model on the CICIDS dataset. This dataset generation did not concentrate on IoMT network attacks.

The above study of DL approaches indicates that these techniques were not highly introduced to identify IoMT network attacks. Most of the authors only explored the network traffic dataset in their experimental work to identify the attacks in the network, as discussed in Table 1. None of the aforementioned works contemplate the combined features of patients biometric with the network flow data.

Table 1
ML and DL Methods to identify an attack in the network.

Year	Authors	Methodology	Dataset	Description	Accuracy	Limitations
2008	Zhang et al. [14]	RF technique	KDD 1999 dataset	In this paper, the proposed security framework was used to detect anomaly traffic on KDD 1999 dataset.	95%	This dataset was developed on a conventional network without adding IoT device traffic. The dataset is not taken into account when looking for IoMT network attacks.
2017	Rao et al. [15]	IPDS-KNN	NSL-KDD Dataset	The proposed technique was used to test diverse types of attacks in the network.	99.6%	The dataset is relevant for Network traffic data only, and not applicable for IoMT.
2017	Shapoorifard et al. [16]	KNN	NSL-KDD Dataset	The introduced approach enhanced the working performance of the IDS and mainly emphasized on decreasing the FAR.	85.2%	The suggested approach achieves low accuracy. The dataset is relevant for Network traffic data only, and not applicable for IoMT.
2018	Su et al. [30]	Lightweight CNN	IoTPOT	The proposed methodology was employed to recognize DDoS cyber-attacks in IoT networks.	94%	This dataset is a combination of IoT threats. This dataset does not use to identify attacks in the IoMT environment.

Year	Authors	Methodology	Dataset	Description	Accuracy	Limitations
2018	Nguyen et al. [31]	CNN classifier	IoT Botnet	The author introduced the model that combines the CNN classifier and PSI graph for Linux IoT botnet identifications in this paper.	92%	This dataset contained flow-based features and was only used for malware identification in IoT networks.
2019	Rathore et al. [17]	Recurrent Neural Network (RNN)	Dataset obtained from Physionet	To forecast various attack simulations in Deep brain stimulators (DBSs), the author introduced a DL model that efficiently identifies the pattern of attacks and alerts a patient regarding that.	Different low Loss Value	There was no discussion of accuracy in the paper, and the simulated attacks were not actual.
2020	T. Saba [19]	Ensemble classifier	KDDcup-99	In this paper, an ensemble classifier IDS was introduced to analyze attacks in the smart hospital.	93.2%	This dataset was created in the traditional network without adding IoT device traffic; The dataset is not considered for identifying attacks in the IoMT network.
2020	Manimurugan et al. [29]	DL-based DBN algorithm	CICIDS dataset	The author introduced a DL-based DBN approach to recognize the network attack in the IoT environment.	96%	The dataset is relevant for Network traffic data only and not applicable to IoMT.

Year	Authors	Methodology	Dataset	Description	Accuracy	Limitations
2020	Hussain et al. [32]	LR (Logistic Regression), KNN, NB, RF	CICIDS2017, IoT-23, CTU-13	Presented the idea of a Universal feature set. Different ML models were considered for classifying the attacks.	89%	Low accuracy; No Hyper-parameter tuning was carried out during the experimental process; Techniques took high prediction time; For Universal Classification Process, the dataset needed to be combined.
2020	Farhan et al. [33]	DNN (Deep Neural Network)	CSE-CIC-IDS 2018	To classify the attack in the network, the author proposed the DNN model.	90%	The proposed method achieved low precision and recall values, i.e., 0.65 & 0.59, respectively; No other techniques were considered for classifying the attack; The dataset is not considered for identifying attacks in the IoMT network.

Year	Authors	Methodology	Dataset	Description	Accuracy	Limitations
2020	Sarhan et al. [34]	Extra Tree Classifier	UNSW-BN 15, ToN-IoT, CSE-CIC-IDS2018, BoT-IoT,	In this paper, four datasets were considered to show NetFlow features. Using the n Probe Tool, the datasets were generalized into universal feature datasets. The proposed approach achieved a weighted average of 70.81%, a Prediction Time of 14.67 µs, and an F1-Score of 0.79.	70.81%	Prediction time is high; For classification, no other techniques have been used; High FAR rate.
2021	Kumar et al. [20]	Ensemble Classifier like RF, NB, and DT	ToN-IoT dataset	The suggested models were applied to categorize the attacks in the IoMT environment.	96.35%	Modbus weather sensors are employed to create a dataset that is generally not used for the IoMT environment. It contains only network traffic data. Therefore, the dataset could not be appropriate for identifying network attacks in IoMT; The false Acceptance Rate (FAR) is high.

Year	Authors	Methodology	Dataset	Description	Accuracy	Limitations
2021	Radoglou et al. [21]	Intrusion Detection and Prevention System (IDPS), DT, RF	CIC-IDS2017 dataset	The IDPS has been proposed to identify and prevent various cyberattacks against communication protocols like Modbus/TCP and HTTP, which are broadly used by e-healthcare services.	96.44%	CIC-IDS2017 is the HTTP network dataset that bothered with network traffic aspects. This dataset is not appropriate for identifying attacks in the IoMT network.
2021	Saheed et al. [25]	PSO-RF	NSL-KDD dataset	The author used PSO and ML/DL-based techniques to identify malicious attacks in the network.	99.76%	The dataset is relevant for Network traffic data only and not applicable to IoMT.
2021	Awotunde et al. [26]	DAE-DFFNN	NF-ToN-IoT dataset	The author presented an SNN-based method that detects intruders while transmitting the data and permits accurate and efficient assessment of medical data at the network edge.	89%	The proposed model achieves low accuracy.
2021	Khan et al. [27]	SDN enabled LSTM and CNN hybrid DL model	—	The author proposed SDN enabled hybrid DL model to detect malicious attacks in the network.	99%	This work only explored the identification of Malware attacks.
2021	Nandy et al. [28]	Intelligent agent-based SNN	ToN-IoT dataset	The author proposed an intelligent agent-based SNN for detecting intruders in IoMT.	99.5%	The dataset is relevant for Network traffic data only and not applicable to IoMT.

Year	Authors	Methodology	Dataset	Description	Accuracy	Limitations
2022	Binbusayyis et al. [24]	KNN, SVM, ANN, NB, and DT	Bot-IoT dataset	This paper showed a detailed comparison of different techniques like KNN, SVM, ANN, NB, and DT for identifying attacks in the network.	100%	The dataset is relevant for Network traffic data only.

3. Dataset Description

In [8], the author developed a testbed using a health monitoring sensor board. It gathered the information using various healthcare sensor devices affixed to the patient body. With the help of a USB port, the panels get connected to the window-based system, which uses C++-based software for collecting the sensed data. The testbed is the combination of components such as a sensor device that help in gathering information related to patient health, SDN controller, and the network gateway for envisaging the network flow traffic. The information generated using sensor and network traffic in the testbed gets utilized to identify the attacks and anomalies in the data.

While generating the dataset, three attacks, i.e., spoofing attacks, MITM attacks, and data injection, were considered. In the MITM attack, the intruder positioned himself as a router and received the message first. After altering or spoofing the message, reroute that file to the server. This results in a contravention of data integrity and confidentiality of the personal information in the network. In a spoofing attack, a program or a person pretends to be someone else to gain the victim’s confidence so that the intruder can successfully ingress the system, purloin information or money, and spread the virus. It results in a contravention of patient data protection and confidentiality. In a data alteration attack, the intruder manipulates or alters the information received at the intruder system from the gateway. The alteration could be made at random or in accordance with a rule. Then the data is forwarded to the server. The modification made by the intruder in the data causes a severe problem for the patient’s health as, depending on the alteration message, false diagnostics will be given to the individual.

The ARGUS Tool [35] creates this combined dataset of network traffic flow, and bio-metric data features of patients. To attain the network traffic flow data, their traffic flow metrics and associated records were collected, along with the biometrics features contain diastolic blood pressure, temperature, respiration rate, systolic blood pressure, pulse rate, ECG ST segment information, heart rate, and peripheral oxygen saturation data. Table 2 shows different features of the dataset. All-inclusive, the dataset includes 44 features, out of which 35 features belong to the network traffic flow category. The output of the dataset is categorized as regular traffic or attack traffic data. “0” represents the normal traffic, whilst “1” illustrates attack traffic data.

Table 2
Description of ML Features in the dataset

Metric	Type	Features Description	Metric Datatypes
Flgs	Flow Metric	Flags in the network	Int
SIntPkt	Flow Metric	Source inter-packet count	Float
DIntPkt	Flow Metric	Destination inter-packet count	Float
SrcBytes	Flow Metric	Source bytes	Float
DstBytes	Flow Metric	Destination bytes	Float
SrcLoad	Flow Metric	Source load	Float
DstLoad	Flow Metric	Destination load	Float
SreGap	Flow Metric	Source missing bytes	Float
DstGap	Flow Metric	Destination missing bytes	Float
sMinPktSz	Flow Metric	Minimum packet size of traffic sent from the source	Float
dMinPktSz	Flow Metric	Minimum packet size of traffic sent from the Destination	Float
sMaxPktSz	Flow Metric	Maximum packet size of traffic sent from the source	Float
dMaxPktSz	Flow Metric	Maximum packet size of traffic sent from the Destination	Float
SIntPktAct	Flow Metric	Source active inter-packet arrival time	Float
DIntPktAct	Flow Metric	Destination active inter-packet arrival time	Float
SrcMac	Flow Metric	Source Mac Address	Int
DtMac	Flow Metric	Destination Mac Address	Int
Loss	Flow Metric	Dropped packets or Retransmitted	Float

Metric	Type	Features Description	Metric Datatypes
pLoss	Flow Metric	Percentage of dropped packets or retransmitted	Float
SrcJitter	Flow Metric	Source jitter	Float
DstJitter	Flow Metric	Destination jitter	Float
Trans	Flow Metric	Cumulative packets count	Float
Dur	Flow Metric	Total Duration	Float
TotPkts	Flow Metric	Total number of packets count	Float
TotBytes	Flow Metric	Total number of packets bytes	Float
pSrcLoss	Flow Metric	Percentage of source dropped packets count or retransmitted packets	Float
pDstLoss	Flow Metric	Percentage of destination dropped packets count or retransmitted packets	Float
Rate	Flow Metric	Number of packets count per second	Float
Load	Flow Metric	Load	Float
Heart_Rate	Biometric	Heart rate	Float
SpO2	Biometric	Peripheral oxygen saturation	Float
Temp	Biometric	Temperature	Float
DIA	Biometric	Diastolic blood pressure	Float
Resp_Rate	Biometric	Respiration rate	Float
SYS	Biometric	Systolic blood pressure	Float
Pulse_Rate	Biometric	Pulse rate	Float
ST	Biometric	ECG ST segment	Float

4. Proposed Methodology

The performance of attack identification utilizing patient biometrics and network flow dataset has been improved in this paper by employing various methodologies to find attacks in the IoMT environment. Before explaining our proposed procedure, we first outline the architecture of IoMT IDS, as shown in Fig. 1. This architecture maintains security performance and predicts attacks using ML and DL methods. The IoMT IDS architecture combines IoT gateway, security operators, network traffic collector, IDS, patient sensor devices, and ML or DL data processing techniques, which are used to monitor different types of attacks. The patient's device sensor comprises ECG devices, a pulse rate detector, a respiration rate tracking machine, temperature sensors, a heart rate tracking machine, and many more. The information gathered using these sensor devices gets transferred to the remote servers using IoT network protocols like Message Queuing Telemetry Transport (MQTT) and Advanced Message Queuing Protocol (AMQP). Using wired or wireless communication, the IoT gateways gathered the sensor's information and transferred that information to remote areas. To identify intrusion in the IoMT ecosystem, the proposed techniques comprise analytical accomplishments and data processing. To reduce false positives and alerts the user regarding the attacker, continuous analysis and monitoring are needed at the IDS level. The physical premises monitoring, and remote monitoring of patients are the primary application of the IoMT environment, which helps to cure and save patients' health in the hospitals.

Figure 2 depicts the operational structure of the IDS framework to identify attacks using ML and DL techniques in the IoMT ecosystem. Rather than analyzing network flow traffic and identifying intrusion in the system, we used IoT sensing data that belongs to the patient's specific IoMT ecosystem. This IoT-based sensing data helps to detect patient biometrics anomalies and enhances the performance of attack identification whenever the intruder tries to attack the IoMT environment. With the help of both events' timestamps, i.e., patient biometric data events and network traffic events, the patient's biometric and network traffic data get combined. The final dataset presents a minority of features belonging to patient biometric data and the majority of the network traffic features.

4.1 Data Preprocessing

Data processing is a preliminary step so that the ML or DL models would provide better results. In this paper, popular techniques like standard scalar and simple imputer were used to preprocess the data. Standard scalar is a well-known method used for data standardization and primarily a preprocessing procedure performed before many ML or DL models to normalize the functional range of the input datasets. It is employed to scale the value distribution in such a matter that the observed value's mean is 0, and its standard deviation is 1. Eq. (1) shows the feature scaling with StandardScaler [36]. A power transformer [37], along with a StandardScaler, is utilized to scale the data in the numeric columns. Power transformers are a family of monotonic and parametric transformations which convert skewed features into normal distributions using logarithmic transform to create data that is more gaussian-like. It handles the modeling problem related to non-constant variance or any other type of condition where normality is required.

$$X_{FeatureScaled} = \frac{X_{FeatureCurrentValue} - mean\left(X_{Feature}\right)}{StandardDeviation\left(X_{Feature}\right)}$$

1

In Eq. (1), $X_{FeatureScaled}$ shows the scaled feature value after the successful execution of the scaling process; Without any alterations, $X_{FeatureCurrentValue}$ displays the current feature value; $X_{Feature}$ represents all of the dataset's feature columns; and the mean for each feature depicted by mean ().

4.2 Feature Engineering

It is an ML approach that uses data to develop new variables that aren't considered in the training set. It helps to modify datasets, including mutation, combination, addition, and deletion operations, in order to upgrade the training of the ML models, and attain ameliorate accuracy and performance. Here, the Feature selection approach has been applied to minimize the input variable by employing the pertinent features and eliminating the noise from the data [38]. In this paper, we used a co-variation matrix heatmap. Some columns were correlated with each other throughout the dataset, so there is no need to consider those columns. Therefore, we are dropping four columns, namely Dir, SrcAddr, DstAddr, and DtMac, by taking advice from the domain experts. After that, Ordinal Encoder [39] is used to encode the categorical columns, i.e., Flgs and SrcMac. Each label is converted into an integer value through ordinal encoding, and the encoded data shows the order of labels.

The dataset was portioned using distributions of 80% for the training dataset and 20% for the testing dataset, respectively, to compute the working performance of the ML and DL models accurately. The K-Fold cross-validation with ten folds has been used on the training dataset sample to demonstrate the range of working performance between the folds. It may be challenging to identify the attacks using this model as the dataset needed to be more balanced, with standard samples making up roughly 88% of the data. Hence, to balance the dataset in the training phase, we employed oversampling method, i.e., SMOTE (Synthetic Minority Oversampling Technique) [40]. The role of this technique is to balance the dataset by producing new synthetic samples for the minority class. In the SMOTE class, the fit_resample method is used from the imblearn library [41] that fits the SMOTE algorithm to the input data X and y, and returns the resampled data. Here, X is a 2D array-like object representing the features of the input data, and y is a 1D array-like object, representing the target or label of the input data.

The amount of resampled data depends on the parameters passed to the SMOTE object. By default, SMOTE creates synthetic samples for the minority state class until the class distribution is balanced. Therefore, there will be an equivalent number of data samples from the minority and majority classes. However, we can specify the desired amount of oversampling by setting the sampling_strategy parameter. For example, if sampling_strategy is set with a value of 0.5, then SMOTE will oversample the minority class by a factor of 0.5. Hence, there will be 50% more samples in the minority class than in the majority class. Table 3 represents the dataset before and after resampling.

Table 3
Dataset before and after Resampling

IoMT Dataset	Dataset before SMOTE Resampling		Dataset after SMOTE Resampling	
	Normal Sample Value	***Attack Sample Value***	***Normal Sample Value***	***Attack Sample Value***
WUSTL-EHMS-2020	14272	2046	14272	14277

4.3 ML Models & Hyperparameters Tuning

In this section, seven ML models have been explored to identify the attack in the IoMT environment.

4.3.1 SVM

SVM is a linear model used for addressing both classification and regression, which work well for the real-world problem and has the capability to supervise linear and non-linear related issues. This technique made a line or hyperplane that distinguished the data into classes [42]. The initial training and testing of the SVM classifier using the dataset sample is explained in Algorithm 1. In order to ensure that $y_i(w.x+b)\geq 1$ is always met, the objective function must be reduced. The Eq. (2) represents the hyperplane function where w represents Weights, b refers to bias.

$$H\left(x\right)=\begin{cases}+1, ifw.x+b\geq 1\\ -1, ifw.x+b\leq 1\end{cases}$$

2

Algorithm 1: SVM Classifier

```
1: Input: Training Features and Classes
2: Output: Hyper-tuned SVM Classifier
3: Begin
4. Preprocessing:
4.1 Performed feature engineering
i. Performed co-variance matrix to identify irrelevant attributes
ii. dropped irrelevant attributes from dataset
iii. applied SMOTE for resampling
iv. applied StandardScalar for continuous attributes
v. applied power transformer to scale the data in the numeric columns
vi. applied ordinal encoder for categorical attributes
5. Training:
i. Initialized SVM Classifier
ii. Set hyperparameter as gamma='auto'
iii. Fit the classifier with labeled classes, and scaled training features, and procure the classifier.
6. Testing:
i. Once all the features have been processed, predict the class of each testing feature.
ii. Use Performance Evaluation Metrics, namely Accuracy, Precision, Recall, and F1-Score, to compute the classifier.
7: End
```

4.3.2 DT

DT is a supervised ML technique that constantly divides the data based on a given parameter. The two entities may be used to explain the DT viz decision leaves and nodes. It is employed for addressing regression and classification issues [43]. The main objective of this algorithm is to develop a model that predicts the intent variable value by learning simple decision rules derived from the data attributes. The initial training and testing of the DT classifier using the dataset sample is explained in Algorithm 2. In our analysis, we set the maximum depth value to 8.

Algorithm 2: DT Classifier

```
1: Input: Training Features and Classes
2: Output: Hyper-tuned DT Classifier
3: Begin
4. Preprocessing:
4.1 Performed feature engineering
i. Performed co-variance matrix to identify irrelevant attributes
ii. dropped irrelevant attributes from dataset
iii. applied SMOTE for resampling
iv. applied StandardScalar for continuous attributes
v. applied power transformer to scale the data in the numeric columns
vi. applied ordinal encoder for categorical attributes
5. Training:
i. Initialized DT Classifier
ii. Set hyperparameter as max_depth = 8
iii. Fit the classifier with labeled classes, and scaled training features, and procure the classifier.
6. Testing:
i. Once all the features have been processed, predict the class of each testing feature.
ii. Use Performance Evaluation Metrics, namely Accuracy, Precision, Recall, and F1-Score, to compute the classifier.
7: End
```

4.3.3 LR

LR is a supervised ML approach that is utilized for addressing classification and prediction issues. It is simple to elucidate and implement, and well organized to train the data. It takes less time to unknown records classification [44]. It is used to forecast or compute the probability that the event occurs should be binary target variables and provides the probabilistic values that range between 0 and 1. The initial training and testing of the LR classifier using the dataset sample is explained in Algorithm 3. Eq. (3) represents the LR model. In our analysis, we set the value of the random state and maximum iterations to 0 and 100, respectively.

$$q = \frac{\mathrm{e}^{(}\mathrm{b0} + \mathrm{b1}{*}\mathrm{p})}{(1 + \mathrm{e}^{(}\mathrm{b0} + \mathrm{b1}{*}\mathrm{p})}$$

3

Where *p* is input values that are merged linearly using coefficient or weight values, *b0* is the intercept or bias value, *b1* represents a single input term *(p)*, and *q* is the predicted output.

Algorithm 3: LR Classifier

```
1: Input: Training Features and Classes
2: Output: Hyper-tuned LR Classifier
3: Begin
4. Preprocessing:
4.1 Performed feature engineering
i. Performed co-variance matrix to identify irrelevant attributes
ii. dropped irrelevant attributes from dataset
iii. applied SMOTE for resampling
iv. applied StandardScalar for continuous attributes
v. applied power transformer to scale the data in the numeric columns
vi. applied ordinal encoder for categorical attributes
5. Training:
i. Initialized LR Classifier
ii. Set hyperparameter as random_state = 0, max_iter = 100
iii. Fit the classifier with labeled classes, and scaled training features and procured the classifier.
6. Testing:
i. Once all the features have been processed, predict the class of each testing feature.
ii. Use Performance Evaluation Metrics, namely Accuracy, Precision, Recall, and F1-Score, to compute the classifier.
7: End
```

4.3.4 KNN

It is a supervised ML approach that is utilized for addressing classification and regression issues. It aims to forecast the proper class samples by measuring the distance amid all test data and training points. After that, it picks the K-Points, i.e., nearest to the test data. This technique computes the probability of the test data sample belonging to K training data classes and selects the class with the excessive probability [45]. In our analysis, we choose two neighbours, set the leaf size to 100, and assign the average values from nearby neighbours as the final projected values. The initial training and testing of the KNN classifier using the dataset sample is explained in Algorithm 4. The closest neighbours are computed using the Euclidean distance (E), as shown below in Eq. (4), where number of neighbours denoted by k and the data points in the i^{th} dimension are a_i and b_i.

$$E = \sqrt{\sum_{i=1}^{k} (a_i - b_i)^2}$$

4

Algorithm 4: KNN Classifier

```
1: Input: Training Features and Classes
2: Output: Hyper-tuned KNN Classifier
3: Begin
4. Preprocessing:
4.1 Performed feature engineering
i. Performed co-variance matrix to identify irrelevant attributes
ii. dropped irrelevant attributes from dataset
iii. applied SMOTE for resampling
iv. applied StandardScalar for continuous attributes
v. applied power transformer to scale the data in the numeric columns
vi. applied ordinal encoder for categorical attributes
5. Training:
i. Initialized KNN Classifier
ii. Set hyperparameter as n_neighbors = 2, leaf_size = 100
iii. Fit the classifier with labeled classes, and scaled training features and procured the classifier.
6. Testing:
i. Once all the features have been processed, predict the class of each testing feature.
ii. Use Performance Evaluation Metrics, namely Accuracy, Precision, Recall, and F1-Score, to compute the classifier.
7: End
```

4.3.5 Gradient Boosting Classifier (Grad_Boost)

It is a collection of ML models that fuse multiple weak learning techniques, usually DT, to build a dedicated predictive model. It may be applied to address both regression and classification issues. Regularization techniques are employed to minimize the consequences of overfitting and to prevent deterioration by confirming that the fitting operation is limited [46]. It is a good technique that deals with unbalanced datasets. The initial training and testing of the Grad_Boost classifier using the dataset sample is explained in Algorithm 5. In our analysis, we set the value of the subsample, estimators, maximum depth, and maximum leaf nodes to 0.5, 1000, 8, and 1000 respectively. The result of H(x) for input x defined in Eq. (5).

$$H\left(x\right)=\sum_{i=1}^{n}\alpha_{i}h_{i}\left(x\right)$$

5

Where, n shows number of leaves, and the value predicted for the region h_i is represented by a_i.

Algorithm 5: Grad_Boost Classifier

```
1: Input: Training Features and Classes
2: Output: Hyper-tuned Grad_Boost Classifier
3: Begin
4. Preprocessing:
4.1 Performed feature engineering
i. Performed co-variance matrix to identify irrelevant attributes
ii. dropped irrelevant attributes from dataset
iii. applied SMOTE for resampling
iv. applied StandardScalar for continuous attributes
v. applied power transformer to scale the data in the numeric columns
vi. applied ordinal encoder for categorical attributes
5. Training:
i. Initialized Grad_Boost Classifier
ii. Set hyperparameter as subsample = 0.5, estimators = 1000, max_depth = 8, max_leaf_nodes = 1000
iii. Fit the classifier with labeled classes, and scaled training features and procured the classifier.
6. Testing:
i. Once all the features have been processed, predict the class of each testing feature.
ii. Use Performance Evaluation Metrics, namely Accuracy, Precision, Recall, and F1-Score, to compute the classifier.
7: End
```

4.3.6 RF

It is a supervised ML approach that is applied to address regression and classification analysis. It is based on the concept of ensemble learning, which integrates many classifiers to address challenging issues, upgrade model performance, and eliminate the problem of overfitting. With high dimensionality, this technique can manage big datasets [47]. Rather than depending on the DT, RF uses prediction from all of the trees and forecasts the final outcome result based on which predictions received the votes. The initial training and testing of the RF classifier using the dataset sample is explained in Algorithm 6. In our analysis, we set the value of estimators, random state, and maximum leaf nodes to 1000, 1, and 1000, respectively.

Algorithm 6: RF Classifier

```
1: Input: Training Features and Classes
2: Output: Hyper-tuned RF Classifier
3: Begin
4. Preprocessing:
    4.1 Performed feature engineering
        i. Performed co-variance matrix to identify irrelevant attributes
        ii. dropped irrelevant attributes from dataset
        iii. applied SMOTE for resampling
        iv. applied StandardScalar for continuous attributes
        v. applied power transformer to scale the data in the numeric columns
        vi. applied ordinal encoder for categorical attributes
5. Training:
    i. Initialized RF Classifier
    ii. Set hyperparameter as n_estimators = 1000, random_state = 1, max_leaf_nodes = 1000
    iii. Fit the classifier with labeled classes, and scaled training features and procured the classifier.
6. Testing:
    i. Once all the features have been processed, predict the class of each testing feature.
    ii. Use Performance Evaluation Metrics, namely Accuracy, Precision, Recall, and F1-Score, to compute the classifier.
7: End
```

4.3.7 XGBoost

It is a well-known supervised ML approach that is enforced to address classification and regression issues. It makes an effort to precisely forecast a target variable by fusing multiple weak learning techniques. It is fundamentally similar to the Grad_Boost classifier, but the residual trees are constructed differently in XGBoost. The variables that are utilized as the roots and nodes of the residual trees are chosen by computing similarity scores between the previous nodes and the leaves [44]. The initial training and testing of the XGBoost classifier using the dataset sample is explained in Algorithm 7. In our analysis, we set the value of the estimators and maximum depth to 1000 and 8, respectively.

Algorithm 7: XGBoost Classifier

```
1: Input: Training Features and Classes
2: Output: Hyper-tuned XGBoost Classifier
3: Begin
4. Preprocessing:
4.1 Performed feature engineering
i. Performed co-variance matrix to identify irrelevant attributes
ii. dropped irrelevant attributes from dataset
iii. applied SMOTE for resampling
iv. applied StandardScalar for continuous attributes
v. applied power transformer to scale the data in the numeric columns
vi. applied ordinal encoder for categorical attributes
5. Training:
i. Initialized XGBoost Classifier
ii. Set hyperparameter as $n\_estimators = 1000, max\_depth = 8$
iii. Fit the classifier with labeled classes, scaled training features, and procured the classifier.
6. Testing:
i. Once all the features have been processed, predict the class of each testing feature.
ii. Use Performance Evaluation Metrics, namely Accuracy, Precision, Recall, and F1-Score, to compute the classifier.
7: End
```

4.4 DL Models

In this section, three deep learning models were discussed to recognize attack in the IoMT network. In all DL models, an Adam Optimizer [48] has been used to train the dataset.

4.4.1 LinSVM (Linear Support Vector Machine)

In this model, we initialize the weight using Xavier Uniform [49], which helps to prevent layer activation outputs from vanishing or exploding during the forward process through a DNN. Both categorical and numerical features were passed to block 1, containing fully connected layers (FC), in which the input gets augmented by a weight matrix in a FC, and a bias vector gets added to it. The output of block 1 is fed to block 2, containing the FC and Tanh activation functions. The classification between two classes is mostly accomplished using the tanh activation function. It has a range of between − 1 to 1. It makes learning easier for the next layer by centering the data and generating mean near to 0. The output of block 2 transferred to block 3, containing the same components. Then, the output from Block 3 is fed to SVM Block containing the FC layer & sigmoid activation function in the end for the binary classification. Figure 3 depicts the working flow of LinearSVM for detecting IoMT attack.

4.4.2 ConvSVM (Convolutional Support Vector Machine)

ConvSVM is combination of a convolutional layer and SVM. In this model, Categorical and numerical features were passed to block 1, containing a 1D convolutional layer followed by Batch normalization (BN) and a Tanh activation function. In this block, sliding convolutional filters are applied to 1-D input via a 1-D convolutional layer. Every value in the input vector gets multiplied by the kernel value in this operation. The purpose of using BN with a 1-D convolutional layer is to ameliorate the model functionality by expediting the training process, employing greater learning rates, and eliminating internal covariate shifts. It also consumes less memory. Tanh activation function makes learning easier for the next layer by centering the data and generating a mean near to 0. The output from Block 1 fed into block 2, containing the same components as the earlier block. After that, the output of Block 2 was sent to block 3, containing a 1D convolutional layer and BN. Then, we used the dropout layer to drop noise or data value that is purposefully removed from a neural network in order to accelerate the processing of the model, and prohibit a model from overfitting. To save from vanishing gradient, we added a residual branch and concatenated it with an output of block 3, which was then passed to an SVM block containing a FC and the sigmoid activation function in the end for the binary classification. Figure 4 depicts the working flow of ConvSVM for detecting IoMT attack

4.4.3 CatEmbedding (Categorical Embedding)

Categorical entity embedding extracts the embedding layers of categorical variables from a neural network model and apply numeric vectors to represent the properties of the categorical values. It is usually used on categorical variables with high cardinalities. In this model, categorical features are separately handled by column embeddings, whereas numerical features are separately handled by the BN, then these features are concatenated together and fed to decider Block 1 containing FC layer, BN, ReLU (Rectified Linear Unit) activation function, and drop-out layer. In FC layers, the input gets multiplied by a weight matrix in a FC layer, and a bias vector gets added to it; BN helps to ameliorate the model functionality by expediting the training process, employing greater learning rates, and eliminating internal covariate shifts; ReLU activation function fixes the vanishing gradients problem and does not allow activation of all the neurons at same interval of time. Then we used the dropout layer to drop noise or data value that is purposefully removed from a neural network in order to accelerate the processing of the model and restrain the model from overfitting. The output of block 1 further sent to block 2, containing the same components. Then, the output of block 2 passed to the SVM block containing the FC layer and sigmoid activation function in the end for the binary classification. Figure 5 depicts the working flow of Categorical Embedding for detecting IoMT attack

5. Experiment And Result Discussion

In this paper, seven ML models, like SVM, DT, LR, KNN, Grad_Boost, RF, and XGBoost, and three DL models, namely LinSVM, ConvSVM, and Cat_Embeddings, have been assessed using performance metrics, including accuracy, precision, recall, and f1-score. Further, this section highlights the working performance of the proposed model compared to the Existing Work.

5.1 Hardware and Software Requirements

The experiment work has been implemented on Windows 10 (64-bit Operating System) with the configuration of 16 GB RAM, Intel(R) Core (TM) i5-7200U CPU @ 2.50GHz. For training and testing the ML and DL models, the python libraries like PyTorch (Commonly utilized for developing Deep neural networks) [50] and Scikit-learn [51] were used.

5.2 Performance Evaluation Metrics

The statistical parameters are applied to determine the working performance of the proposed models. To assess the performance of IoMT attack classification, we used four well-known metrics, which are explained as follows: -

1. True Positive (TP): In this case, both the forecasted outcome and the true outcome of the data point are valid. It is appropriately categorized as an attack sample.
2. True Negative (TN): In this case, both the forecasted outcome and the true outcome of the data point are false. Therefore, it is appropriately categorized as a standard sample.
3. False Positive (FP): In this case, the forecasted outcome of the data point is precise, and the true outcome of the data point is false. It designated a standard data set as an attack sample.
4. False Negative (FN): In this case, the forecasted outcome of the data point is false, whereas the true outcome of the data point is valid. Therefore, despite being an attack sample, it was restricted as a normal sample.

Using the above metrics, the working performance of the ML and DL models gets assessed in terms of Accuracy, Precision, Recall, and F-1 Score, which is explained as follows:

Accuracy is elucidated as the ratio of the correct traffic categorization to the overall prediction of the test data for network traffic. Accuracy is measured using Eq. (6).

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

6

Precision is elucidated as the ratio of the correct categorization of attack traffic to the total of the correct and incorrect categorization of attack traffic samples in the dataset. The model functions well if the accuracy is high. Precision is measured using Eq. (7).

$$Precision = \frac{TP}{TP + FP}$$

7

The recall is elucidated as the ratio of the correctly classified attack network traffic to the total of precisely classified attack network traffic and incorrectly classified attack network traffic in the dataset. Recall is measured using Eq. (8).

$$Recall = \frac{TP}{TP + FP}$$

8

F-Score amalgamates the precision and recall of a classifier into a single statistic by calculating their harmonic mean. F1-Score is measured using Eq. (9).

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

9

The ML models, including SVM, DT, LR, KNN, Grad Boost, RF, and XGBoost, are used in our experimental work to identify the IoMT attacks. Figure 6 depicts the performance of the ML model. Out of seven models, five models, namely LR, DT, RF, Grad Boost, and XGBoost, achieved an accuracy of 100%, whereas the other two models, SVM and KNN, accomplished an accuracy of 99.9% and 99.8%, respectively.

Table 4 represents the working performance metrics of ML Models to identify the IoMT attacks. Out of seven models, five ML models namely LR, DT, RF, Grad_Boost, and XGBoost obtained the best precision, recall, and F-1 Score with a value of 1.0. Whereas the other two model i.e., KNN and SVM also achieved good score value as depicted in the table below.

Table 4
The working performance metrics of ML Models

ML Techniques	Accuracy	Precision	Recall	F-1 Score
Grad Boost	1.0	1.0	1.0	1.0
RF	1.0	1.0	1.0	1.0
XGB	1.0	1.0	1.0	1.0
LR	1.0	1.0	1.0	1.0
KNN	0.9981	0.9974	0.9989	0.9982
DT	1.0	1.0	1.0	1.0
SVM	0.9998	0.9997	1.0	0.9998

In Fig. 7, the loss is measured on the vertical axis (Y), while the number of epochs is displayed on the horizontal axis (X). The training loss is calculated for the number of epoch counts ranging from 1 to 200 for all three proposed DL models namely LinSVM, ConvSVM, and CatEmbedding. It is observed from the above result that, LinSVM and ConvSVM converge to 0 after 50 epochs.

In Fig. 8, the validation loss is measured on the vertical axis (Y), while the number of epochs is displayed on the horizontal axis (X). The validation loss is calculated for the number of epoch counts ranging from 1 to 200 for all three DL models. It is observed from the above result that after 50 epochs, ConvSVM converges to 0 rapidly as compared to other two models.

The DL models, including LinSVM, ConvSVM, and Cat_Embeddings are used in our experimental work to identify the IoMT attacks. Figure 9 depicts the performance of the DL model. Out of three models, two DL models namely ConvSVM, and Cat_Embeddings achieved an accuracy of 100%, whereas the other models, LinSVM achieved an accuracy of 99.94%.

Table 5 depicts the working performance metrics of DL Models to identify the IoMT attacks. Out of three DL models, two DL models, namely ConvSVM and Cat_Embeddings, obtained the precision, recall, and F-1 Score with a value of 1.0. Whereas the precision, recall, and F-1 score value of LinSVM is 1.0, 0.9989, and 0.9994, respectively.

Table 5
Performance Metrics of DL Models

DL Techniques	Accuracy	Precision	Recall	F-1 Score
LinSVM	0.9994	1.0	0.9994	0.9989
ConvSVM	1.0	1.0	1.0	1.0
CatEmb	1.0	1.0	1.0	1.0

5.3 Performance evaluation of Proposed Approach with Existing Work

In this paper, seven ML Models, including SVM, DT, LR, KNN, Grad_Boost, RF, and XGBoost, along with three deep learning models, namely LinSVM, ConvSVM, and Cat_Embedding model, have been implemented, which further compared with the existing work on the same dataset as shown in Table 6. Our proposed models achieved the highest accuracy as compared to the other existing work.

In [8], Hady et al. designed an upgraded healthcare testbed for monitoring the patients and also used to gather biometrics data of the patient along with network flow metrics. In this work, author built the dataset that combines the feature of both biometric data of the patients and network flow metrics. Then applied, different ML approaches, including ANN, SVM, RF, and KNN and attained an accuracy of 88.75%, 92.44%, 92.06%, and 92.27%, respectively. ANN achieved the lowest accuracy as compared to other models. In [52], Dina et al. proposed Feed Forward neural network model (FFNN) with the Focal Loss method to detect intrusion in the network and achieved 93.26% of accuracy. The author applied the focal loss function to remove the data imbalance challenge. In [53], Gupta et al. proposed a tree classifier to recognize intrusion in the network and attained an accuracy of 93%. In [54], Chaganti et al. developed PSO-DNN (Particle Swarm Optimization Deep Neural Network) model to recognize intrusion in the network and achieved 96% of accuracy. In [55], Kilincer et al. proposed XGBoost and LR model with 10-fold cross validation on four different datasets namely ICU dataset, ECU-IoHT, WUSTL-EHMS, and TON-IoT. For WUSTL-EHMS dataset, the proposed model attained 96.2% of accuracy. Our proposed models achieved 100% of the accuracy which is better than the existing work performance.

Table 6
Performance evaluation of Proposed Models with Existing Work

Dataset	Year	Author	ML/DL Techniques	Accuracy	Pros	Cons
WUSTL-EHMS-2020	2020	Hady et al. [8]	KNN	90%	Merged the features of network flow traffic and patients' biometric features data.	Performance might be enhanced.
	2022	Dina et al. [52]	FFNN-Focal Loss	93.26%	In this paper, the author used focal loss technique to remove the data imbalance challenge.	The working performance of the model can be upgraded by applying proper feature engineering. Because in this work, the author only used Focal Loss method to mitigate the data imbalance issue, which is a traditional approach.
	2022	Gupta et al. [53]	Tree Classifier	94.23%	Enhanced the performance as compared to [8, 50].	Unrealistic datasets were produced via data augmentation. The percentage of attack traffic in the network is relatively low.
	2022	Chaganti et al. [54]	PSO-DNN	96%	Enhanced the working performance of the models to identify attack in the network as compared to [8, 50, 51].	Performance can be further enhanced using feature engineering.

Dataset	Year	Author	ML/DL Techniques	Accuracy	Pros	Cons
	2022	Kilincer et al. [55]	XGBoost	96.2%	In this paper, the author used two ML technique including XGBoost and LR based on recursive feature elimination, and then applied that on four different datasets. RFE is utilized in this work so that the working functionality of chosen features can be easily managed, and can achieved high accuracy as compared to filter methods.	In this work, the author only considered limited data features from the dataset. The performance can be improved using deep learning models.
	Proposed Approach		**ML Models- Grad_Boost, RF, XGBoost, LR, DT.** **DL Models- ConvSVM, CatEmbedding.**	**100%**	**Proposed ML/DL models which attained the highest accuracy as compared to the existing work.**	

6. Conclusion

IoMT services are in more demand than ever before as the number of users has been increasing rapidly. This calls for a robust security model to thwart any unwanted activities. A detection method is also required in the security architecture for such distributive networks to determine whether an intrusion has affected the network's data. Therefore, to overcome the impact causes by the attacker, we proposed three DL models, namely LinSVM, ConvSVM, and CatEmbedding, to design and implement a secure IDS for IoMT network. With a 100% accuracy rate, our developed models exceed the state-of-the-art in detecting network intrusions by utilizing the combined features of both network traffic flow metric and the patient’s biometric data. At last, an in-depth evaluation of several ML models has been explored for intrusion detection in IoMT ecosystem. Out of seven proposed ML models, five models, including LR, DT, RF, GradBoost, and XGBoost, also attained 100% of accuracy, and the other two models, SVM and KNN, accomplished an accuracy of 99.9% and 99.8%, respectively. For future work, we will consider more adversarial attacks with the network flow and patients’ biometric features to detect intruders in the IoMT network.

Declarations

Ethical Approval

This study uses publicly available data or data from published sources; therefore, no subject testing or data collection procedure was taken into account for this particular study.

Competing interests

The authors declare that they have no conflict of interest.

Authors’ contributions

All authors contributed to the study's conception, validation, and design. Material preparation, data analysis, and experimental work were performed by PGS and NS. NS wrote the final draft of the manuscript and prepared figures and tables based on the script. All authors read and approved the final manuscript.

Funding Not Applicable

Availability of data and materials

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request. For more details on the dataset, please refer to the paper [8].

References

1. State of IOT 2022: Number of connected IOT devices growing 18% to 14.4 billion globally. IoT Analytics. (2022, June 14). Retrieved February 20, 2023, from https://iot-analytics.com/number-connected-iot-devices/
2. Internet of medical things (IOMT): Innovative Future for Healthcare Industry. Cogniteq. (n.d.). Retrieved February 20, 2023, from http://www.cogniteq.com/blog/internet-medical-things-iomt-innovative-future-healthcare-industry
3. Fotouhi, H., Causevic, A., Lundqvist, K., &amp; Bjorkman, M. (2016). Communication and security in Health Monitoring Systems – a review. 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC). https://doi.org/10.1109/compsac.2016.8
4. Newman, L. H. (2022, March 8). Critical bugs expose hundreds of thousands of medical devices and atms. Wired. Retrieved February 20, 2023, from https://www.wired.com/story/access7-iot-vulnerabilities-medical-devices-atms/
5. Ravi, V., Alazab, M., Selvaganapathy, S., &amp; Chaganti, R. (2022). A multi-view attention-based deep learning framework for malware detection in Smart Healthcare Systems. Computer Communications, 195, 73–81. https://doi.org/10.1016/j.comcom.2022.08.015
6. Radoglou-Grammatikis, P., Sarigiannidis, P., Efstathopoulos, G., Lagkas, T., Fragulis, G., &amp; Sarigiannidis, A. (2021). A self-learning approach for detecting intrusions in Healthcare Systems. ICC 2021 - IEEE International Conference on Communications. https://doi.org/10.1109/icc42927.2021.9500354

7. Ghubaish, A., Salman, T., Zolanvari, M., Unal, D., Al-Ali, A., &amp; Jain, R. (2021). Recent advances in the internet-of-medical-things (IOMT) systems security. IEEE Internet of Things Journal, 8(11), 8707–8718. https://doi.org/10.1109/jiot.2020.3045653.
8. Hady, A. A., Ghubaish, A., Salman, T., Unal, D., &amp; Jain, R. (2020). Intrusion detection system for healthcare systems using medical and network data: A comparison study. IEEE Access, 8, 106576–106584. https://doi.org/10.1109/access.2020.3000421
9. Clifton, L., Clifton, D. A., Pimentel, M. A., Watkinson, P. J., &amp; Tarassenko, L. (2014). Predictive monitoring of mobile patients by combining clinical observations with data from wearable sensors. IEEE Journal of Biomedical and Health Informatics, 18(3), 722–730. https://doi.org/10.1109/jbhi.2013.2293059
10. Rani, A. A., &amp; Baburaj, E. (2019). Secure and intelligent architecture for cloud-based healthcare applications in Wireless Body Sensor Networks. International Journal of Biomedical Engineering and Technology, 29(2), 186. https://doi.org/10.1504/ijbet.2019.097305
11. Chakraborty, S., Aich, S., &amp; Kim, H.-C. (2019). A secure healthcare system design framework using Blockchain technology. 2019 21st International Conference on Advanced Communication Technology (ICACT). https://doi.org/10.23919/icact.2019.8701983
12. Alabdulatif, A., Khalil, I., Forkan, A. R., &amp; Atiquzzaman, M. (2019). Real-time secure health surveillance for Smarter Health Communities. IEEE Communications Magazine, 57(1), 122–129. https://doi.org/10.1109/mcom.2017.1700547
13. Tao, H., Bhuiyan, M. Z., Abdalla, A. N., Hassan, M. M., Zain, J. M., &amp; Hayajneh, T. (2019). Secured data collection with hardware-based ciphers for IOT-based healthcare. IEEE Internet of Things Journal, 6(1), 410–420. https://doi.org/10.1109/jiot.2018.2854714
14. Jiong Zhang, Zulkernine, M., &amp; Haque, A. (2008). Random-forests-based network intrusion detection systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 38(5), 649–659. https://doi.org/10.1109/tsmcc.2008.923876
15. Rao, B. B., &amp; Swathi, K. (2017). Fast knn classifiers for network Intrusion Detection System. Indian Journal of Science and Technology, 10(14), 1–10. https://doi.org/10.17485/ijst/2017/v10i14/93690
16. Shapoorifard, H., &amp; Shamsinejad, P. (2017). Intrusion detection using a novel hybrid method incorporating an improved KNN. International Journal of Computer Applications, 173(1), 5–9. https://doi.org/10.5120/ijca2017914340
17. Rathore, H., Al-Ali, A. K., Mohamed, A., Du, X., &amp; Guizani, M. (2019). A novel deep learning strategy for classifying different attack patterns for deep brain implants. IEEE Access, 7, 24154–24164. https://doi.org/10.1109/access.2019.2899558
18. Yaacoub, J.-P. A., Noura, M., Noura, H. N., Salman, O., Yaacoub, E., Couturier, R., &amp; Chehab, A. (2020). Securing internet of medical things systems: Limitations, issues and recommendations. Future Generation Computer Systems, 105, 581–606. https://doi.org/10.1016/j.future.2019.12.028
19. Saba, T. (2020). Intrusion detection in Smart City Hospitals using ensemble classifiers. 2020 13th International Conference on Developments in ESystems Engineering (DeSE). https://doi.org/10.1109/dese51703.2020.9450247

20. Kumar, P., Gupta, G. P., &amp; Tripathi, R. (2021). An ensemble learning and fog-cloud architecture-driven cyber-attack detection framework for IOMT networks. Computer Communications, 166, 110–124. https://doi.org/10.1016/j.comcom.2020.12.003
21. Radoglou-Grammatikis, P., Sarigiannidis, P., Efstathopoulos, G., Lagkas, T., Fragulis, G., &amp; Sarigiannidis, A. (2021). A self-learning approach for detecting intrusions in Healthcare Systems. ICC 2021 - IEEE International Conference on Communications. https://doi.org/10.1109/icc42927.2021.9500354
22. Zachos, G., Essop, I., Mantas, G., Porfyrakis, K., Ribeiro, J. C., &amp; Rodriguez, J. (2021). An anomaly-based intrusion detection system for internet of medical things networks. Electronics, 10(21), 2562. https://doi.org/10.3390/electronics10212562
23. Thamilarasu, G., Odesile, A., &amp; Hoang, A. (2020). An intrusion detection system for internet of medical things. IEEE Access, 8, 181560–181576. https://doi.org/10.1109/access.2020.3026260
24. Binbusayyis, A., Alaskar, H., Vaiyapuri, T., &amp; Dinesh, M. (2022). An investigation and comparison of machine learning approaches for intrusion detection in IOMT Network. The Journal of Supercomputing, 78(15), 17403–17422. https://doi.org/10.1007/s11227-022-04568-3
25. Saheed, Y. K., &amp; Arowolo, M. O. (2021). Efficient cyber attack detection on the Internet of Medical Things-smart environment based on deep recurrent neural network and machine learning algorithms. IEEE Access, 9, 161546–161554. https://doi.org/10.1109/access.2021.3128837
26. Awotunde, J. B., Abiodun, K. M., Adeniyi, E. A., Folorunso, S. O., &amp; Jimoh, R. G. (2022). A deep learning-based intrusion detection technique for a secured IOMT system. Informatics and Intelligent Applications, 50–62. https://doi.org/10.1007/978-3-030-95630-1_4 .
27. Khan, S., &amp; Akhunzada, A. (2021). A hybrid DL-driven intelligent SDN-enabled malware detection framework for internet of medical things (IOMT). Computer Communications, 170, 209–216. https://doi.org/10.1016/j.comcom.2021.01.013
28. Nandy, S., Adhikari, M., Khan, M. A., Menon, V. G., &amp; Verma, S. (2022). An intrusion detection mechanism for secured IOMT framework based on Swarm-Neural Network. IEEE Journal of Biomedical and Health Informatics, 26(5), 1969–1976. https://doi.org/10.1109/jbhi.2021.3101686
29. Manimurugan, S., Al-Mutairi, S., Aborokbah, M. M., Chilamkurti, N., Ganesan, S., &amp; Patan, R. (2020). Effective attack detection in internet of medical things smart environment using a deep belief neural network. IEEE Access, 8, 77396–77404. https://doi.org/10.1109/access.2020.2986013
30. Su, J., Danilo Vasconcellos, V., Prasad, S., Daniele, S., Feng, Y., &amp; Sakurai, K. (2018). Lightweight Classification of IOT malware based on image recognition. 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC). https://doi.org/10.1109/compsac.2018.10315
31. Nguyen, H.-T., Ngo, Q.-D., &amp; Le, V.-H. (2018). IOT botnet detection approach based on psi graph and DGCNN classifier. 2018 IEEE International Conference on Information Communication and Signal Processing (ICICSP). https://doi.org/10.1109/icicsp.2018.8549713
32. Hussain, F., Abbas, S. G., Fayyaz, U. U., Shah, G. A., Toqeer, A., &amp; Ali, A. (2020). Towards a universal features set for IOT botnet attacks detection. 2020 IEEE 23rd International Multitopic Conference (INMIC). https://doi.org/10.1109/inmic50486.2020.9318106

33. Farhan, R. I., Maolood, A. T., &amp; Hassan, N. F. (2020). Performance analysis of flow-based attacks detection on CSE-CIC-IDS2018 dataset using Deep Learning. Indonesian Journal of Electrical Engineering and Computer Science, 20(3), 1413. https://doi.org/10.11591/ijeecs.v20.i3.pp1413-1418

34. Sarhan, M., Layeghy, S., Moustafa, N., &amp; Portmann, M. (2021). NetFlow datasets for Machine Learning-based network intrusion detection systems. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 117–135. https://doi.org/10.1007/978-3-030-72802-1_9

35. openargus. Retrieved February 20, 2023, from https://openargus.org/

36. Sklearn.preprocessing.StandardScaler. scikit. (n.d.). Retrieved February 20, 2023, from https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

37. Sklearn.preprocessing.PowerTransformer. scikit. (n.d.). Retrieved February 20, 2023, from https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.PowerTransformer.html

38. Jiang, H., Lin, J., &amp; Kang, H. (2022). FGMD: A robust detector against adversarial attacks in the IOT network. Future Generation Computer Systems, 132, 194–210. https://doi.org/10.1016/j.future.2022.02.019.

39. Sklearn.preprocessing.OrdinalEncoder. scikit. (n.d.). Retrieved February 20, 2023, from https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OrdinalEncoder.html

40. Brownlee, J. (2021, March 16). Smote for imbalanced classification with python. MachineLearningMastery.com. Retrieved February 20, 2023, from https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/

41. Imblearn. PyPI. (n.d.). Retrieved February 20, 2023, from https://pypi.org/project/imblearn/

42. Wang, W., Du, X., &amp; Wang, N. (2019). Building a cloud ids using an efficient feature selection method and SVM. IEEE Access, 7, 1345–1354. https://doi.org/10.1109/access.2018.2883142

43. Moon, D., Im, H., Kim, I., &amp; Park, J. H. (2015). DTB-ids: An intrusion detection system based on decision tree using behaviour analysis for preventing apt attacks. The Journal of Supercomputing, 73(7), 2881–2895. https://doi.org/10.1007/s11227-015-1604-8

44. Nayak, J., Meher, S. K., Souri, A., Naik, B., &amp; Vimal, S. (2022). Extreme learning machine and bayesian optimization-driven intelligent framework for IOMT cyber-attack detection. The Journal of Supercomputing, 78(13), 14866–14891. https://doi.org/10.1007/s11227-022-04453-z

45. Kumaran, S. S., Balakannan, S. P., &amp; Li, J. (2021). A deep analysis of object capabilities for intelligence considering wireless IOT devices with the DNN approach. The Journal of Supercomputing, 78(4), 4745–4758. https://doi.org/10.1007/s11227-021-04064-0

46. Mishra, S. (2022). An optimized gradient boost decision tree using enhanced African buffalo optimization method for cyber security intrusion detection. Applied Sciences, 12(24), 12591. https://doi.org/10.3390/app122412591

47. Mantas, C. J., Castellano, J. G., Moral-García, S., &amp; Abellán, J. (2018). A comparison of random forest-based algorithms: Random credal random forest versus oblique random forest. Soft Computing, 23(21), 10739–10754. https://doi.org/10.1007/s00500-018-3628-5

48. Adam¶. Adam - PyTorch 1.13 documentation. (n.d.). Retrieved February 27, 2023, from https://pytorch.org/docs/stable/generated/torch.optim.Adam.html
49. Sirignano, J., &amp; Spiliopoulos, K. (2022, April 12). Scaling limit of neural networks with the Xavier Initialization and Convergence to a global minimum. arXiv.org. Retrieved February 27, 2023, from https://arxiv.org/abs/1907.04108v3
50. What is torch.nn really?¶. What is torch.nn really? - PyTorch Tutorials 1.13.1+cu117 documentation. (n.d.). Retrieved February 20, 2023, from https://pytorch.org/tutorials/beginner/nn_tutorial.html
51. Sklearn.model_selection.train_test_split. scikit. (n.d.). Retrieved February 20, 2023, from https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html
52. Dina, A. S., Siddique, A. B., &amp; Manivannan, D. (2023). A deep learning approach for intrusion detection in internet of things using focal loss function. Internet of Things, 22, 100699. https://doi.org/10.1016/j.iot.2023.100699
53. Gupta, K., Sharma, D. K., Datta Gupta, K., &amp; Kumar, A. (2022). A tree classifier-based network intrusion detection model for internet of medical things. Computers and Electrical Engineering, 102, 108158. https://doi.org/10.1016/j.compeleceng.2022.108158
54. Chaganti, R., Mourade, A., Ravi, V., Vemprala, N., Dua, A., &amp; Bhushan, B. (2022). A particle swarm optimization and deep learning approach for intrusion detection system in the internet of medical things. Sustainability, 14(19), 12828. https://doi.org/10.3390/su141912828
55. Firat Kilincer, I., Ertam, F., Sengur, A., Tan, R.-S., &amp; Rajendra Acharya, U. (2023). Automated detection of cybersecurity attacks in healthcare systems with recursive feature elimination and multilayer perceptron optimization. Biocybernetics and Biomedical Engineering, 43(1), 30–41. https://doi.org/10.1016/j.bbe.2022.11.005.

Figures

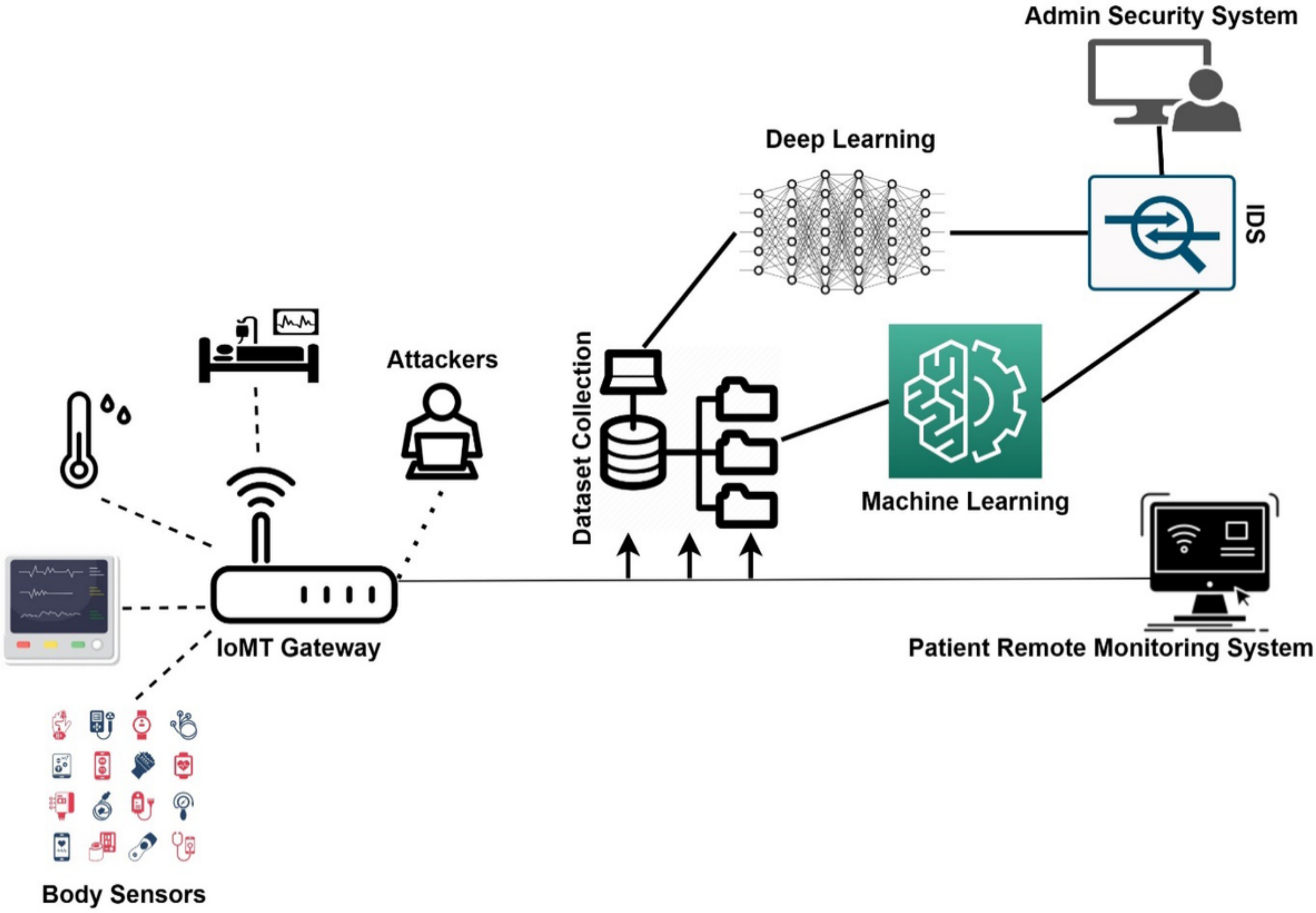


Figure 1

ML and DL-based IDS for IoMT Environment

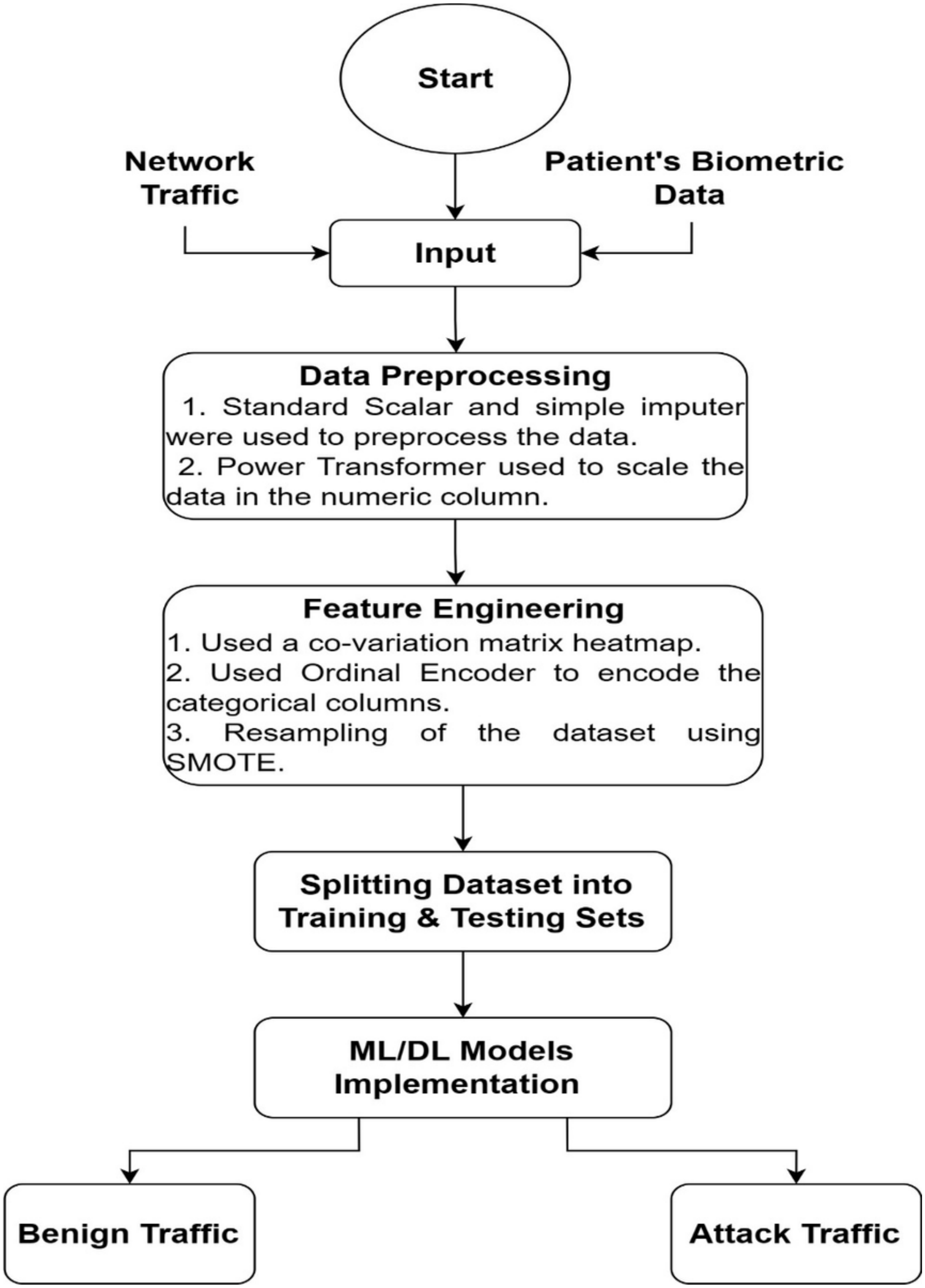

Start
Network Traffic
Patient's Biometric Data
Input
Data Preprocessing
1. Standard Scalar and simple imputer were used to preprocess the data.
2. Power Transformer used to scale the data in the numeric column.
Feature Engineering
1. Used a co-variation matrix heatmap.
2. Used Ordinal Encoder to encode the categorical columns.
3. Resampling of the dataset using SMOTE.
Splitting Dataset into Training & Testing Sets
ML/DL Models Implementation
Benign Traffic
Attack Traffic

Figure 2

Proposed Flowchart to detect IoMT attack

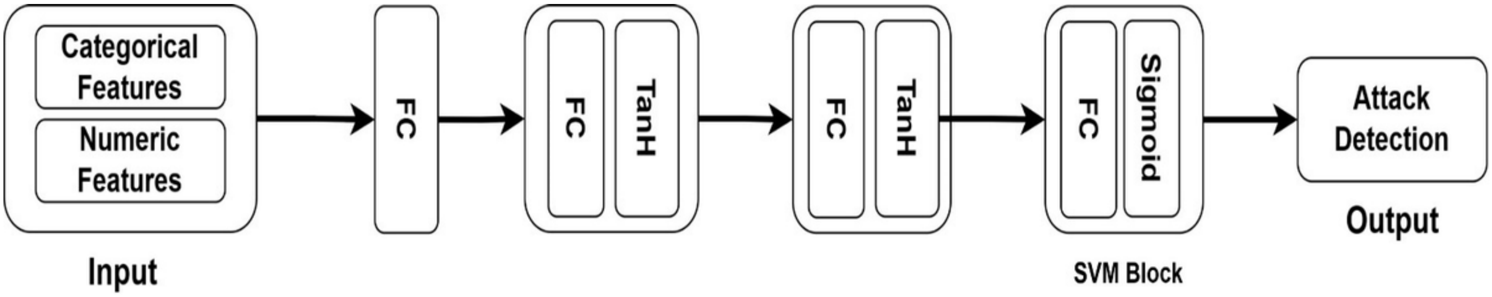


Figure 3

Working flow of LinearSVM for detecting IoMT attack

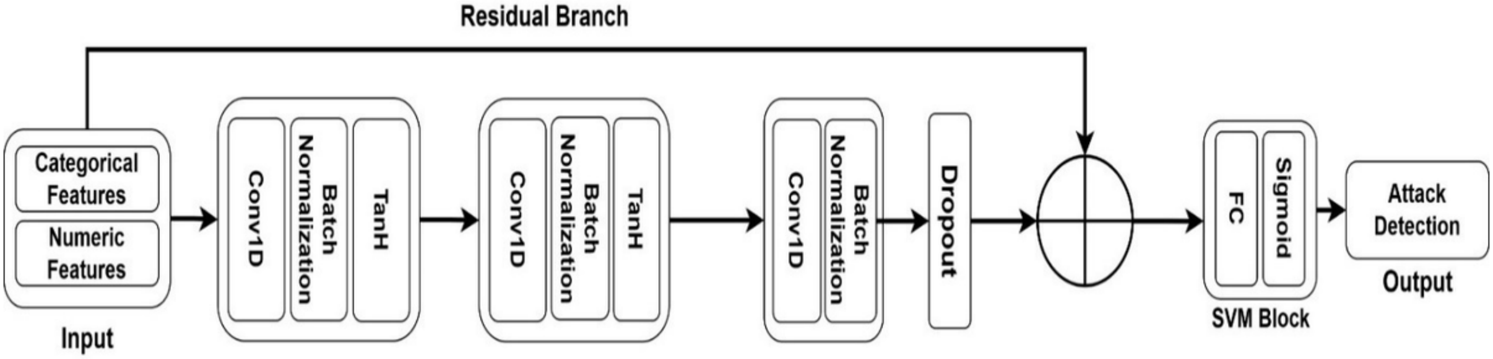


Figure 4

Working flow of ConvSVM for detecting IoMT attack

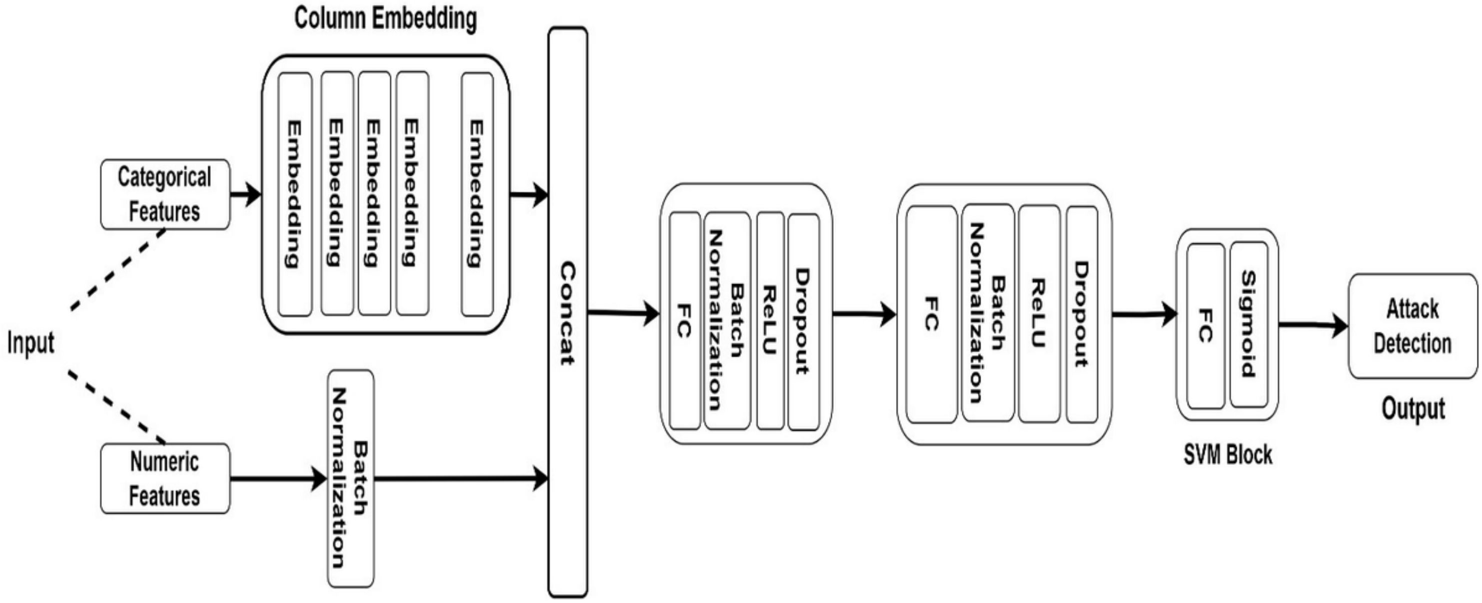


Figure 5

Working flow of Categorical Embedding for detecting IoMT attack

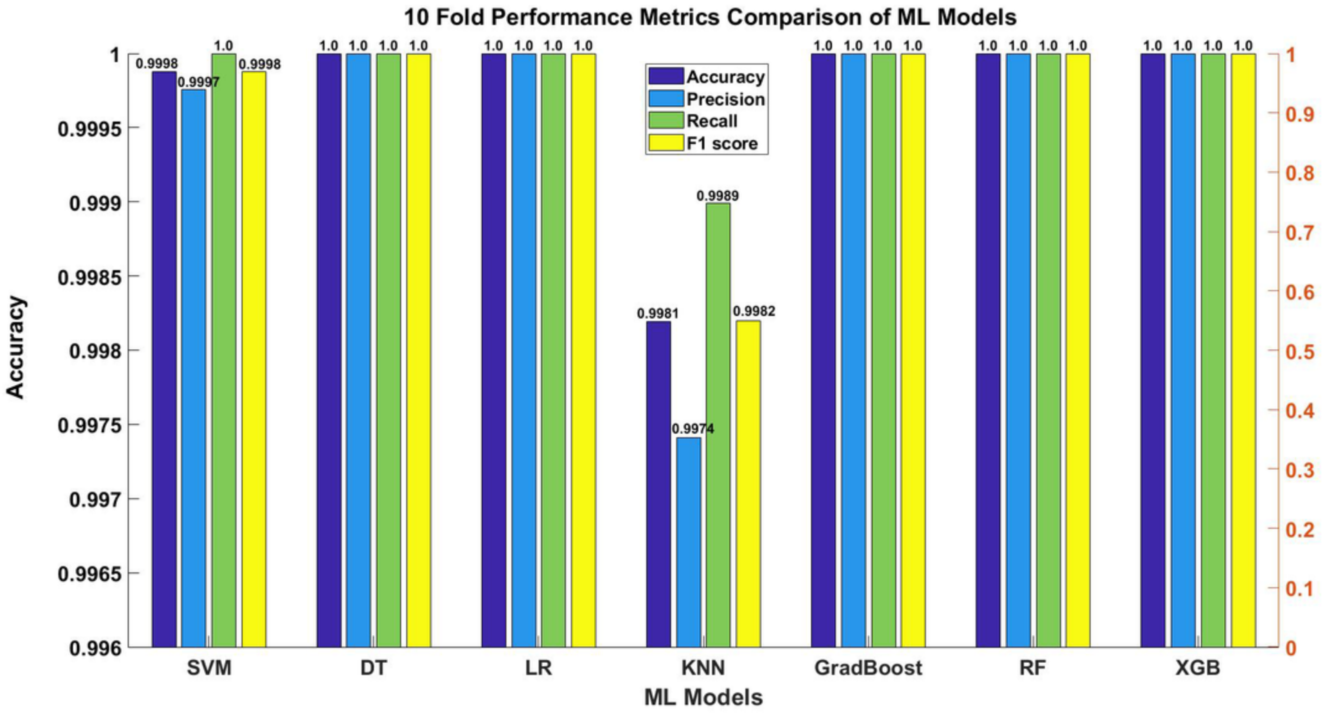


Figure 6

10-Fold Performance Metrics comparison of ML Models

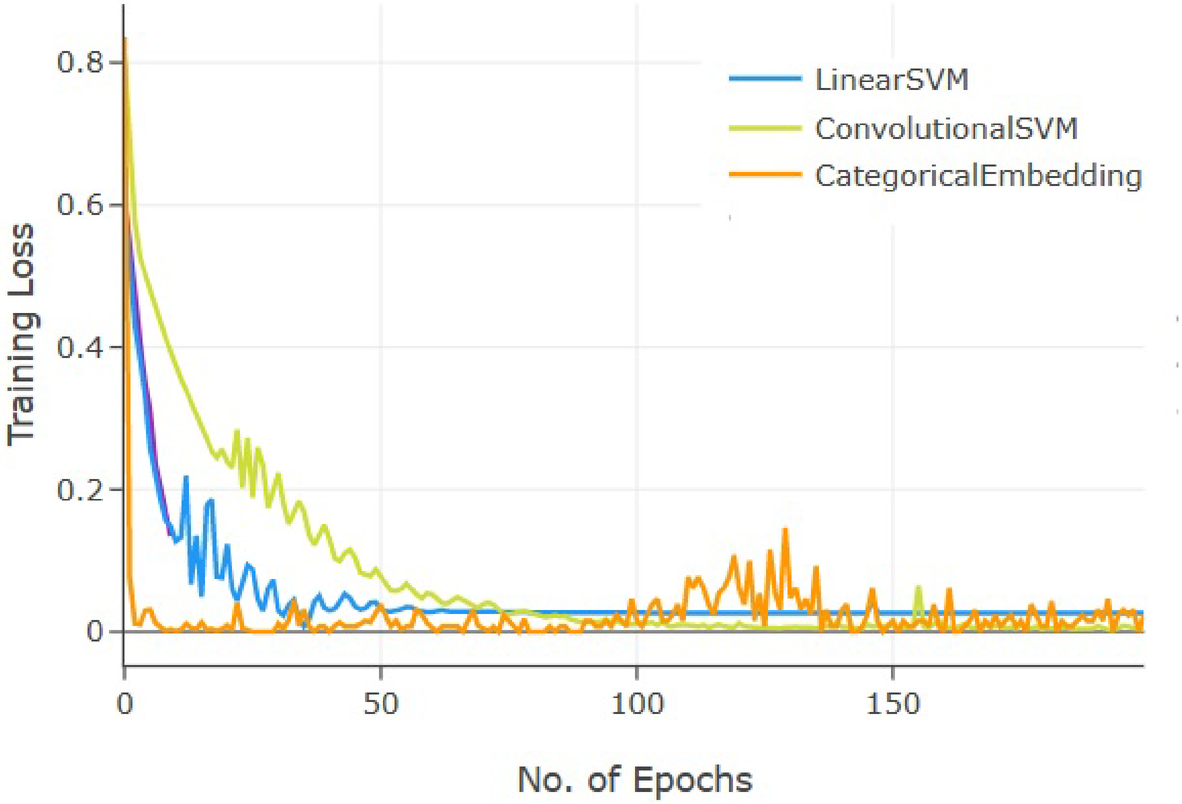


Figure 7

Training Loss for DL Models

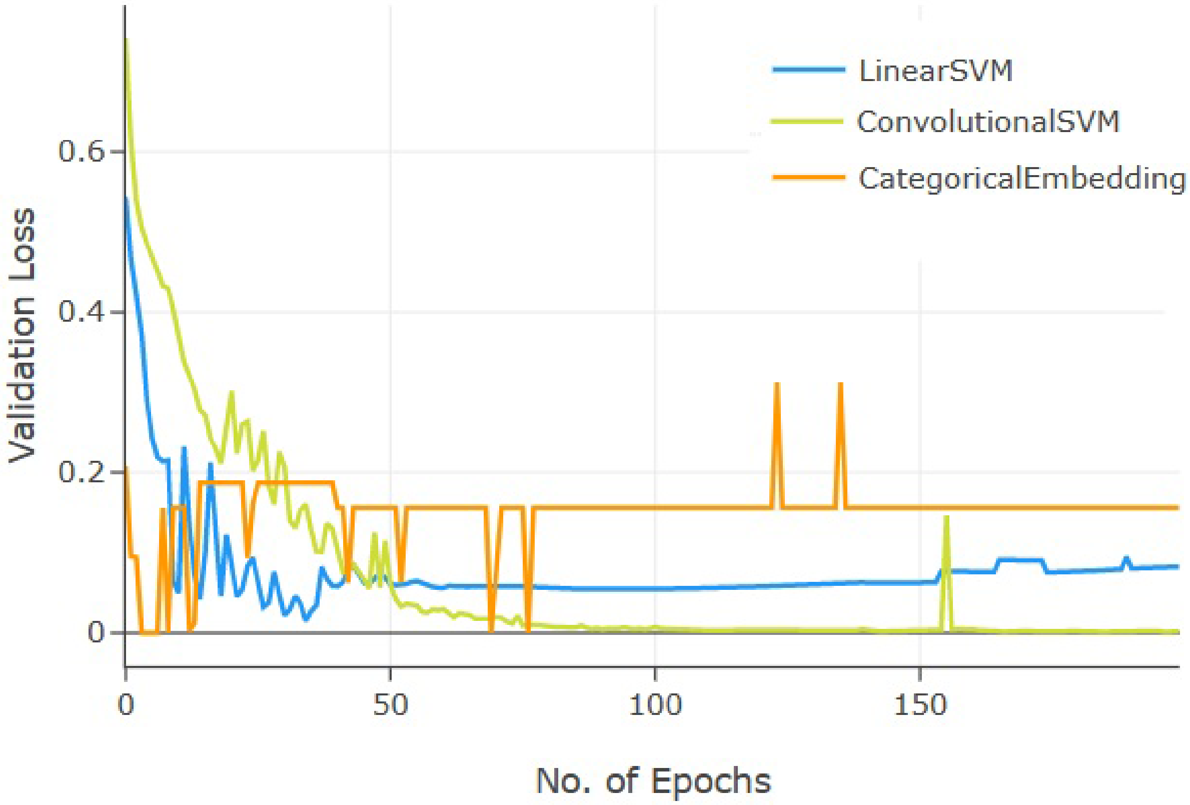


Figure 8

Validation Loss for DL Models

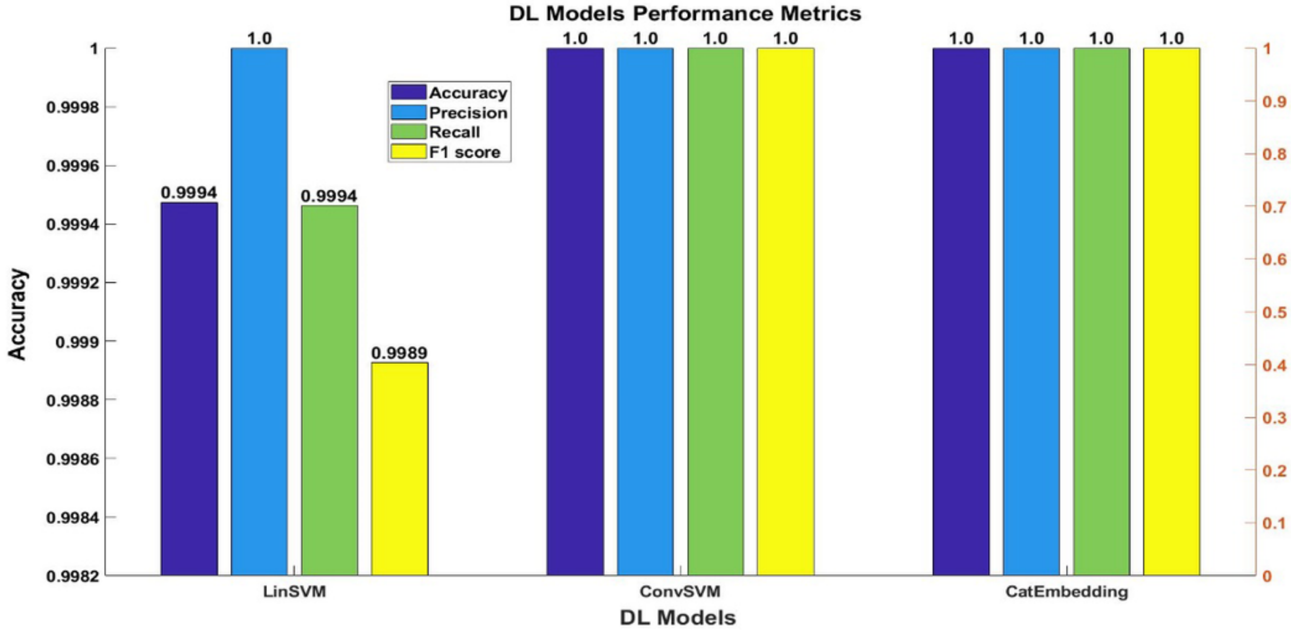


Figure 9

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 297-306 www.ijaem.net ISSN: 2395-5252

Artificial Intelligence and its relation with Computer Aided Manufacturing

Bibek Khadka[1], Bijoy Chouhan[2], Dr. A.K. Madan[3]
[1,2]*Undergraduate Student, Department of Computer Engineering, Delhi Technological University, Delhi, India.*
[3]*Professor, Department of Mechanical Engineering, Delhi Technical University, Delhi, India.*

Date of Submission: 25-02-2023 Date of Acceptance: 05-03-2023

ABSTRACT – Artificial intelligence (AI) research papers are essential for developing the subject and influencing the direction of technology. This paper provides the AI application areas, appraise existing approaches, and recommend new research directions. In order to increase the effectiveness and quality of the manufacturing process, artificial intelligence (AI) is rapidly being employed in computer-aided manufacturing (CAM). Process optimization, quality control, and predictive maintenance are a few CAM applications for AI.The creation of algorithms and methodologies that enhance the functionality, precision, and effectiveness of AI models is one of the key contributions of AI research publications.

This paper covers the fundamental ideas behind artificial intelligence and examines how it is now used in the manufacturing industry. The problems with artificial intelligence are also identified and some possible solutions are suggested. We hope that the information offered in this paper might serve as a valuable set of guidelines and references for future work on artificial intelligence in the manufacturing sector.

KeyWords: Artificial intelligence, computer–aided manufacturing(CAM), Computer Numerical Control(CNC), Computer Aided Design(CAD), Aircraft manufacturing, surface roughness prediction.

I. INTRODUCTION

Over the past few years, the manufacturing industry has exploited the use of AI technology, and has taken advantage in particularly knowledge-based systems, throughout the manufacturing lifecycle. These technologies have been motivated by the competitive challenge of improving quality while at the same time decreasing costs and reducing design and production time. Artificial intelligence has several advantages that are desired in manufacturing practice, including learning and adapting ability, parallel distributed computation, robustness, etc[01].

Artificial Intelligence has received a major focus in both academia and the industry recently due to the competitive advantages that it can provide to manufacturing organizations in creating a more efficient and sustainable operation [02]. The manufacturing sector is going through a period of change with production and AI technology progress setting the pace of this transformation [03]. At the same time, more and more emerging AI technologies such as big data analytics, advanced robotics, expert systems for diagnosis, computer vision and pattern matching for outgoing product quality are creating an impact to the manufacturing industry in a major way [04].

AI algorithms may examine a variety of elements through process optimization, including raw material selection, tool path planning, and cutting parameters, to streamline the production process, decrease waste, and increase productivity. By examining the manufactured product's quality and identifying flaws, surface polish, and other quality characteristics, AI may also be utilized for quality control[05]. Machine vision and image processing techniques may be used to do this and guarantee that the result satisfies the necessary quality requirements.

In general, the application of AI in CAM may result in increased productivity, decreased costs, and higher-quality products, making it a technology that is becoming more and more significant in the industrial sector.

II. RESEARCH METHODOLOGY

The methodology of AI in computer-aided manufacturing (CAM) typically entails identifying the issue or challenge of conventional CAM systems, gathering and preprocessing data from various sources, such as sensors and control systems, developing an AI model using a suitable algorithm and training it on the pre-processed data, integrating the AI model with the CAM software or control system, testing and evaluating the AI-based CAM system to determine its efficacy, and finally, implementing the system.

III. OVERVIEW OF AI TECHNOLOGIES

The manufacturing sector may undergoing a radical transformation thanks to artificial intelligence. The potential advantages include enhanced quality, decreased downtime, lower costs, and higher efficiency[06]. This technology is accessible to smaller firms also.

Though they pertain to two different ideas, the terms artificial intelligence and machine learning are occasionally used interchangeably. By using historical data to show you the odds between several choices and which one obviously worked better in the past, it assists us in solving a specific problem[07]. It explains the significance of everything, the chances that specific outcomes will occur, and their likelihood in the future[05].

3.1 Why adopt AI

Making decisions that can be put into action more quickly and correctly than a human can is what artificial intelligence (AI) in manufacturing refers to. This makes a lot of sense for forecasting and for comprehending anomalies or outliers, to name just two applications[08]. Forecasting can add value in some stages of the production process. There is a good probability that you can make forecasts if you have access to enough historical data as well as information about the decisions and processes around the data.

A human analyst may find the data from one machine to be overwhelming, which is where AI might be useful. Additionally, because manufacturing systems are integrated, one measure in one step of the process can affect another step in the same step. How can you know what's happening in another region if you're just focusing on one? AI may offer a remedy. The four categories listed below are where AI has a big financial influence[09].

- Predictive upkeep. By using historical information from maintenance logs, you may forecast how a machine will perform under a future payload and determine whether you'll need to fix it, when, why, and how based on what fixed that problem in the past[10].
- Reliable prediction. Significant cost savings can be achieved by predicting and minimizing failures.
- Increasing output or yield. You may prevent quality passes by anticipating when a machine or process won't meet requirements and taking proactive steps to bring it back into compliance.
- Forecasting of demand and inventory. It is possible to estimate the demand and movement of essential parts with a complete understanding of plant operations and the production data, leading to significant inventory savings[11].

3.2 How might computer-aided design be improved by artificial intelligence?

The number of manual actions needed to create a design has decreased because of CAD. The drafting process has been significantly expedited by this time savings, which also allowed designers to refocus their efforts[12]. Designers create ever-more complicated ideas as a result. Although the fundamentals have been covered, there are still a lot of other barriers that prevent designers, engineers, and architects from rationally enhancing their workflows. Significant bottlenecks include the following[09]:

- In order to create the optimal design for a project's requirements, designers frequently have to manually adjust model parameters.
- Validating designs after each modification might cause the project to be delayed by days or even weeks because changing just one parameter can significantly affect the attributes of a design.
- A project's progress can be slowed down by feedback loops since gathering data to figure out what needs to be changed takes time.

Since they are simply "glorified drawing boards," one could argue that existing CAD systems conduct computer-aided drafting rather than computer-aided design. Because current technology only helps designers with drafting, the opportunities and challenges of CAD have not yet been fully explored or addressed[12].

AI can address these problems as it develops and becomes more thoroughly integrated into CAD by:

- By producing ideas based on specific criteria, it is possible to expedite the drafting and selection of design solutions (such as weight, size, costs, or material).
- Modifying and altering designs automatically if they don't satisfy performance or aesthetic standards.
- Depending on the user's previous behaviors,

recommending more details to include in the design.

- Adapting current designs to user input, evolving technology, or fresh legal needs.

These processes might be combined into a single solution in future, more sophisticated AI models that would handle the entire design process. AI might free up designers and engineers to concentrate on other, possibly more crucial issues like enhancing the quality, effectiveness, and dependability of their works by handling the labor-intensive tasks.

3.3 What role has AI played in CAD thus far?

AI has already made its way into computer-aided design in one form or an another[17].

"Three most important architectural potentials of the new machine mediation techniques are the expansion of the spatial imagination, and the radical break with a hierarchical design approach, and the introduction of different disciplines into the design process, relating to the design immediately to its final execution."

For instance, a number of software providers have included AI capabilities in their architectural, engineering, and construction solutions. Autodesk provides generative design tools to users in order to assist them optimize their drawing workflows, maybe most prominently. The company's technology accomplishes this by swiftly generating design proposals based on a variety of input factors, including pricing, production processes, materials, or spatial requirements.

Notably, AI in CAD isn't just used for designing and refining designs. Siemens unveiled a new version of its NX CAD software in February 2019 that features a user interface that alters depending on the user and situation. A CAD tool frequently provides the draftsperson or engineer with too many commands. Many people contend that just 10% or less of the available commands are applied in 90% of CAD system operations. When the AI system determines that the engineer would need more commands—commands that might be unknown or infrequently used—a dynamic UI displays them.

The development of CAD datasets for AI training has involved a lot of work, with Sketch Graphs serving as a prime example. Sketch Graphs, which was released in 2020 by academics at Princeton University and Columbia University, has 15 million parametric CAD sketches. The drawings are shown as a geometric constraint graph, where the edges denote the geometric relationships between the primitives (nodes). Sketch Graphs places more emphasis on the relationship structure of its sketch samples than other CAD datasets that stress 3D shape modeling.

IV. DEVELOPMENT OF AI IN CAM

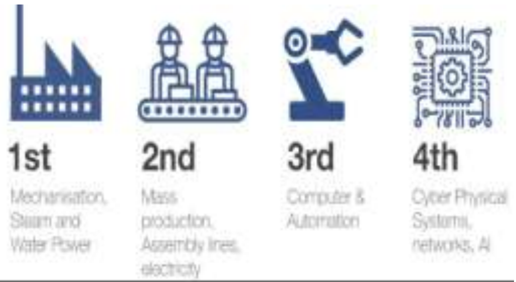

1st
Mechanisation, Steam and Water Power
2nd
Mass production, Assembly lines, electricity
3rd
Computer & Automation
4th
Cyber Physical Systems, networks, AI

Figure: 01 Revolution In Manufacturing Industry with AI

Efficiency, accuracy, and reliability are important needs in the manufacturing sector. Artificial intelligence (AI) can enhance the technology that manufacturers have embraced, such as computer-aided design (CAD) and computer-aided manufacturing (CAM)[13]. The full potential of these technologies can be realized by incorporating AI into 3D modeling. In CAD, digital designs are created, analyzed, and modified using computers before a product is made. These models are then used by CAM to regulate production procedures and equipment in order to produce final goods that adhere to design criteria. The manufacturing sector has benefited greatly from both innovations, but AI has more potential[14].

To achieve some additional functional requirements specification for AI in CAM, one should imagine to a certain extent the performance of an intelligent distributed computer environment. number of users will exploit, update and extend systems's knowledge[15].

So, The first case is to be preferred, for the knowledge-quanta may be supposed to have primarily local significance[16]. Then , the knowledge should be first handled locally and after that consequently included in an existing knowledge base, or used as a foundation of a new one. then, a local knowledge server should be able to determine the significance of each knowledge-quantum in the local pool local or global and in the latter case, to leave it to the common knowledge server[17]. However, that is enough to show the necessity of a higher level intelligence in the computer environment of CAM[18]. Hence, there is a need for tools for building multi level knowledge hybrid knowledge bases and handling systems and also a variety of knowledge handling structures. Further, the needed knowledge handling tools should be able to work in a distributed computer environment with CAM, i.e. to be compatible with a properly designed system layer and as well as to be compatible on the support layer[19]. Both of the above mentioned requirements are directly connected to the openness, flexibility and transparency of the distributed computer environment, so they produce the questions to be asked when one purchases software expert systems building tools for the creation of AI in CAM[20].

4.1 How AI is improving CAD and CAM processes in modern manufacturing ?

Figure: (02) 3D modeling [21].

- **Increased Productivity:** Higher Productivity AI first and foremost enhances 3D modeling by increasing the effectiveness of the procedure. As many as 15 million CAD designs are used as the basic dataset by certain AI design assistants, which influences their forecasts[08]. With that much knowledge, they can make forecasts that are remarkably accurate. By eliminating the need for users to manually sketch several elements, this increases efficiency during the design process. These AI assistants are also capable of automating design decisions[08]. For instance, AI can automatically apply geometry to new projects by aligning them based on how pieces were applied in previous designs[16]. To guarantee that everything lines up properly, this process would often need slow, careful tweaks, but AI can do it in a matter of seconds. Manufacturers can then accelerate the time to market for new products while they can also concentrate on other activities or produce more[19].

- **AI Is the Future of 3D Modeling:**

The Future of 3D Modeling Is AI 3D modeling with AI is still a relatively new technique. The technology is already expanding across CAD and CAM software systems, despite its youth[22]. The reasons why CAD and CAM technologies are so widely used are their effectiveness, accuracy, and dependability. Each of these advantages can be enhanced by AI, resulting in higher-than-expected results[23].

- **Ongoing Improvements:** Continuous Development AI in 3D modeling can facilitate continuing advancements, just like in other industrial processes[22]. Over time, they'll start to notice trends in their triumphs and failures and propose adjustments to maximize the former and decrease the latter. The design and manufacturing processes can be combined with the use of AI in CAD and CAM, opening the door for operational benefits[05]. Both sides' data will show how the production side can change to better support the designs engineers desire to create. As new elements surface, AI can identify these areas for development and adapt them[08].

V. AN APPLICATION OF AI FOR CAM AND CAD TO INTEGRATE AIRCRAFT MANUFACTURING

A single engineer lacks the expertise needed to incorporate the restrictions and manufacturing characteristics of aircraft into the structural design process. Concurrent Engineering (CE) makes it possible to integrate design and production to allow trades based not just on product performance but also on other difficult to evaluate factors, such production and support[08]. System designers would benefit greatly from a decision support system, or knowledge- based system, that helps guide manufacturing concerns throughout the preliminary design process. To illustrate the KBS's (knowledge Based System) functionality as a design tool, it will be used in an integrated design environment with other tools already available[19].

5.1 KNOWLEDGE-BASED SYSTEM DEVELOPMENT

5.1.1 Problem Domain

The High Speed Civil Test (HSCT) is the unique test case, and the area of study is the integration of design and manufacturing. The focus of this research will be on a significant airframe component, exactly the same one that caused designers the most difficulty in the 1970s.

The KBS's task is to choose the production procedures for the structural parts of the wings. In this area, it is not feasible to pre enumerate all of the potential outcomes and then choose the best one based on the data. that is normally available as a first-level structural analysis. Instead of, a set of workable techniques that satisfy the external restrictions imposed by material specifications, fabrication and assembly issues, and cost considerations.

Regarding product design, a few assumptions have been made.

The **first**, prior to structural modeling and optimization, the materials from which the wing structural components will be made are preselected from a database of potential possibilities. This presumption is made because it is impossible to predict the weight of each structural component accurately without modeling its unique material qualities. **Second,** when utilizing weight-complexity based parametric cost models, calculated weights and dimensions of the structural components will vary dramatically with different materials depending on performance requirements and load conditions. When using commercially available parametric cost models, this factor is often ignored. **Third,** some related to process selection will be abstracted to the functional level. Before precise models of the parts are constructed in CAD systems, the manufacturing procedures in the aerospace manufacturing sector are chosen.

The information about the components that will be known before modeling, after structural analysis and optimization, and before the KBS chooses the processes is shown in **Table 1[19]**.

Product & process parameters	Skin panel	Rib	Spar	Spar cap
before modeling and structural analysis/optimization:				
material & associated properties, constraints, & max. service temp.	✔	✔	✔	✔
grid coordinates	✔	✔	✔	✔
modeled as membrane element		✔	✔	
modeled as rod element				✔
after structural analysis and optimization:				
thickness	✔	✔	✔	
cross sectional area				✔
part weight (mass)	✔	✔	✔	✔
production considerations and decisions:				
manufacturing process	✔	✔	✔	✔
fasteners	✔		✔	✔
stiffener type	✔			

stiffener material	✔			
solid or honeycomb construction	✔			

Table 1: Wing Component Modeling [19]

In order for the KBS to operate effectively, the knowledge regarding material selection, manufacturing procedures, stiffener types and materials, fasteners, and fundamental part configuration must be represented in a suitable manner. In a KBS, rules are the most typical form of domain knowledge representation. The frames used to define the items that appear within the rules are frequently combined with the rules themselves [24].

5.1.2 Knowledge Base Development

The knowledge and rule bases raise a number of significant difficulties. Several of the data sets required for the knowledge base building are not collocated.

The relevant information must be acquired through a lengthy process of knowledge acquisition. The most suitable format must be used for compiling and coding historical data on material usage and process selection factors as well as current design guidelines and norms. With the use of frames, it is possible to categorize the data that represents technical information[19].

5.2. SYSTEM INTEGRATION WITH EXISTING TOOLS

Knowledge-Based Engineering (KBE) is a subset of KBS and AI technologies that focuses on automating the generation of support information, engineering analysis, and CAD geometry. The system must function inside an integrated design environment in order to be helpful and show its functioning. Without the proper interface automation techniques, the system's intended functionality won't be visible. As **Figure 03** shows the proposed integrated design environment in which the Knowledge-Based Engineering (KBS) will function.

The system executive scripts will be written in the Tk/tcl (toolkit/tool command language) interpretive shell system. To enable the creation of fully complete, fully functional graphical user interfaces, Tk/tcl combines an interpretative language core with windowing applications. The creation and application of parametric, intelligent CAD systems is a special objective of several aerospace industries. Even if the combination of a KBS and a CAD programme is not a next-generation system in and of itself, it is a step in the right direction [19].

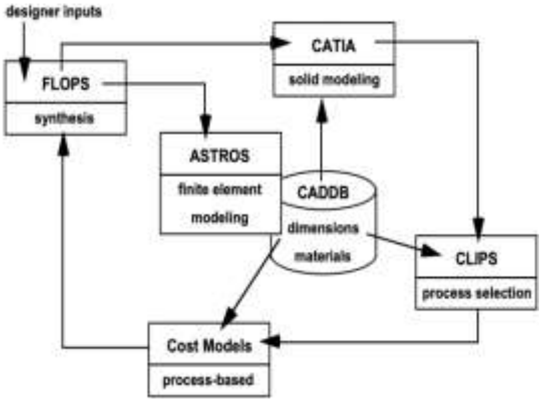


Figure: (03) Integrated Design Environment

As shown in **Figure 03**, the system will be given direct links to CATIA for the purpose of retrieving or storing data regarding the structural elements of the wing. A single function that dynamically accesses all of the internal CATGEO functions will be utilized to access the CATIA resources using Tk/tcl. The Automated STRuctural Optimization System of the USAF and NASA Langley's FLight OPtimization System are two other technologies that are now employed for the product design analyses **(ASTROS)** [19].

5.3 COST MODELING

The stages of the wing product design in this study are decomposed traditionally from the system level to the sub-system level to the part level. Take-off gross weight (TOGW), range, payload, cruise speed, and passenger count are typical examples at the system level. Any decisions made about funding a specific programme will be based on the Life Cycle Cost (LCC) of the intended system, which includes all manufacturing expenses (both recurring and nonrecurring) [19].

VI. AI BASED SURFACE ROUGHNESS PREDICTION

For recently machined objects, there are currently no commercial software tools that assist the accomplishment of preset surface roughness. Machine operators and Computer Aided

Manufacturing (CAM) programmers rely on experience to determine the ideal balance between time and quality optimization[25].

Computer- aided manufacturing programs, which calculate the tool path from input parameters such tool geometry, feed rate, spindle speed, but also part and blank geometries, help manufacturing planning within the Computerized Numerical Control (CNC) milling domain[07]. Professional CNC programmers choose those parameters, but skilled machine operators still need to execute at least one test run to adjust the feed rate in the event of chatter vibrations.[09].

As mentioned, both online and offline chatter avoidance technologies have made significant strides towards optimizing the surface roughness of machined objects. A machine-learning system will be trained using planning, process, and quality data pertaining to features[24].

6.1 Feature based Database Design and Sensor Integration

The data will be compiled into a database in the form of parameter sets and time series, comprising measurements of the surface's roughness, vibration information, cutting depth, feed rate, and spindle speed.

The proposal for integrating the data from the various PLM sources (CAD, CAM, Process, and In- spection) and supplying it to the AI, which is fed by a database categorized by features, is shown in **Figure 04**.

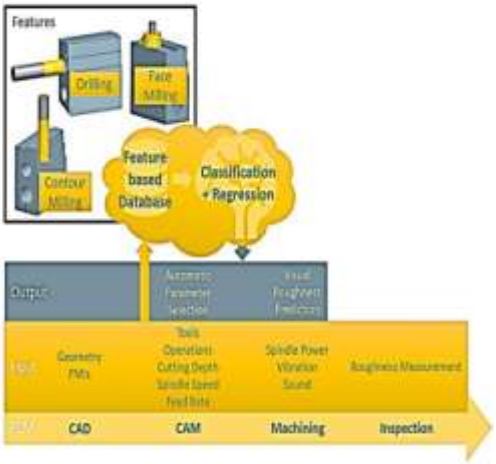

Features
Drilling
Face Milling
Contour Milling
Feature based Database
Classification + Regression
Output
Automatic Parameter Selection
Visual Roughness Prediction
Geometry PMIs
Tools Operations Cutting Depth Spindle Speed Feed Rate
Spindle Power Vibration Sound
Roughness Measurement
CAD
CAM
Machining
Inspection

Figure:04

The manufacturing features, which serve as the database's central entities, are displayed at the top. As a core information source for the CAM optimization and the live visualization, which are depicted in the Output layer, the classification and regression model feeds off of the database[25].

6.2 Experimental Data Generation

The defining of the experimental variables, such as feed rate, spindle speed, and cutting depth, is the first stage in the design of experiments. But it's also important to identify the static conditions, such as the component and tool material, geometry, kind of tooling, and tool wear[25]. Simply put, the spindle speed, feed rate, their combination, and the insert radius have a substantial impact while the tool diameter and depth of cut have little to no effect.

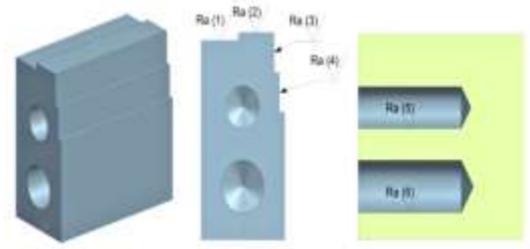


Figure: 05

The Part design demanded roughness values Ra (1) to Ra (6) of three different manufacturing features (face milling, contour milling, and drilling) for two specific geometries per feature.

A part with three distinct production features—drilling, face milling, and contour milling—is shown in **Figure 05**. Each feature, there are two distinct operations included. While the drilling operations have different tool diameters, the milling features have different cutting depths. Before being post-processed into an NC programme, the CAM programme is developed and its cutting parameters automatically adjusted with predetermined offsets and within the required range[09].

The physical parts are categorized for data traceability and readied for identification when the CAM and CNC is programmed, which including tool change, are generated[07]. Finally, the machine learning platform, database connectivity, and data flow must all be well documented and continuously monitored while the machining experiments are being conducted. In this project, a Fanuc 31i numerical control made by Fanuc Austria is combined with a 5-axis milling center manufactured by DMG Mori called the DMU 75 Monoblock[25].

6.3 Data Preparation Analysis and Model Building

A machine learning method with a classification and regression component, like a random forest, is constructed and trained by the collected data sets based on the final data input structure. The experiment execution phase and the simultaneous model development phase will employ an existing data lake structure and machine learning platform.

VII. CONCLUSIONS

Based on the research paper evaluation, "Artificial Intelligence in computer assisted manufacturing," it can be stated that AI has tremendous potential to enhance computer aided manufacturing (CAM) operations. The article gives an overview of the present level of AI in CAM and examines several uses of AI in CAM, such as improving manufacturing processes, quality control, and predictive maintenance.

The report outlines the benefits of adopting AI in CAM, including enhanced productivity, decreased costs, and increased accuracy. The authors do point out that careful planning and implementation are necessary for the integration of AI into CAM as well as that there may be issues with data management and privacy that need to be resolved.

The article contends that through allowing more effective and efficient CAM processes, AI has the potential to revolutionize the manufacturing sector. To fully exploit the advantages of AI in CAM and to address the issues and constraints related to its application, more research and development is necessary.

In conclusion, AI research articles are essential for developing AI, enhancing the functionality of AI models, addressing ethical and societal concerns, and encouraging global partnerships. They are essential in determining how technology will develop and how it will affect society.

ACKNOWLEDGEMENT

We would like to express our special gratitude to our teacher **Dr. Ashok Kumar Madan** who gave us the golden opportunity to do this wonderful Research on the topic **“Artificial Intelligence and its relation with Computer Aided Manufacturing”**, which also helped us to know about so many new things in the manufacturing unit. While making this project he helped and guided us in solving any queries and issues we had.

In addition, we would like to thank the **Department of Mechanical Engineering of Delhi Technological University** for giving us the opportunity to collaborate and work together on this Research project.

REFERENCES

[1]. G. H. Schaffer, “Artificial intelligence: A tool for smart manufacturing,” American Machinist and Automared Manufacturing, vol. 130, no. 8, p. 83, 1986,ISSN: 1741-038X, doi.10.1108/JMTM-09-2018-0325.

[2]. Definition of Computer Aided Manufacturing Engineering and its Place in CA Systems Chain , J. Novak-Marcincin*, J. Barna*, J. Torok* and L. Novakova- Marcincinova, Technical University of Kosice/Department of Manufacturing Technologies, Presov, Slovakia jozef. SAMI 2013 • IEEE 11th International Symposium on Applied Machine Intelligence and Informatics • January 31 - February 2, 2013 • doi: 10.1109/SAMI.2013.6480989

[3]. Rich, E., and Knight, K., Artificial Intelligence, McGraw-Hill Book Company, USA, 1991, Journal of Power and Energy Engineering, Vol.6 No.12, December 18, 2018,doi: 10.4236/jpee.2018.612002.

[4]. Warburton, L. M. and Glatfelter, J. W., "Development and Utilization of a KnowledgeBased (KB) Environment on a Production Helicopter Program", presented at the American Helicopter Society Vertical Lift Aircraft Design Conference, San Francisco, CA, January 1995, Volume 2.

[5]. Leondes, Cornelius, ed. "Computer-Aided Design, Engineering, and Manufacturing." Vo of The Design of Manufacturing Systems. CRC Press, 2001, vol. 6, doi: 10.1201/9781420050059

[6]. J. Lu, C. Ren, and W. Zhang, "Research on the Application of Artificial Intelligence in CNC Machining Process," in 2018 International Conference on Control, Automation and Diagnosis (ICCAD), Changsha, China, 2018, pp. 856-861, doi: 10.1109/ICCAD.2018.8595828.

[7]. K. H. Kweon, S. J. Kim, and J. W. Park, "Development of an Artificial Intelligence-based Machining System for CNC Milling," in Journal of Mechanical Science and Technology, vol. 30, no. 2, pp. 881-888, Feb. 2016, doi: 10.1007/s12206-016-0138-1.

[8]. Integrated Manufacture and Engineering, No. 1/2, 1996, pp. 44-46. J. Novak-Marcincin, J. Barna, M. Janak, L. Novakova- Marcincinova, J. Torok,

Application of the Open Source, doi: 10.1007/978-3-319-05948-8_4.

[9]. Y. Zhao, Z. Guo, and W. Zhou, "Intelligent Monitoring System for CNC Machine Tools Based on Artificial Intelligence," in 2019 IEEE 5th International Conference on Computer and Communications (ICCC), Chengdu, China, 2019, pp. 120-125.

[10]. S. Zhang, H. Song, and Z. Ren, "Application of Artificial Intelligence in Welding Manufacturing," in 2019 2nd International Conference on Advanced Energy Conservation and Emission Reduction (AEER 2019), Kunming, China, 2019, pp. 107-110, doi: 10.2991/aeer-19.2019.25

[11]. C. Kim, J. H. Kim, and S. Park, "Design and Implementation of a Smart Manufacturing Platform Based on Artificial Intelligence," in 2018 IEEE 6th International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM), Bangkok, Thailand, 2018, pp. 73-78, doi: 10.1109/CIS-RAM.2018.8511864.

[12]. Iwata, Y. and Obama, N. QDES: Quality Design Expert System for Steel Products. In Innovative Applications of Artificial Intelligence 3, Smith and Scott (eds.), AAAI, Menlo Park, California, 1991,Vol. 26, No. 5 (Sep. - Oct., 1996), pp. 50-70,doi:stable/25062168.

[13]. Mouta,F.and E.Oliviera (1987).Cooperative Responses in Question-answer Systems, International Workshop on Expert Systems Avignon'87, 1635-1686.

[14]. Ranky,P. (1986). The potential of expert systems in CIM, Practice Hall International, Computer Integrated Manufacturing, 51-57, vol. 7, no. 2, ISSN 2456-1428.

[15]. Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P.J. Boczkowski, & K. A. Foot (Eds.). Media Technologies: Essays on Communication, Materiality, and Society. Boston: MIT Press. pp. 167-197, doi: 10.7551/mitpress/9780262525374.001.0001

[16]. Petropoulos, G. (2018). The impact of artificial intelligence on employment. In M Neufeind, J. O'Reilly, and F. Ranft, Praise for Work in the Digital Age: Challenges of the Fourth Industrial Revolution, London: Rowan & Littlefield, pp. 119-132, vol. 119, curis.ku.dk, 2018,Page cited: p.121.

[17]. Relevance-of-ai-in-computer-aided-design, aug 07 2021,

[18]. "An Application of Artificial Intelligence for Computer-Aided Design and Manufacturing", Article · December 2000. W. J. Marx1, D. P. Schrage2, and D. N. Mavris3 , Aerospace Systems Design Laboratory, School of Aerospace Engineering Georgia Institute of Technology, Atlanta, GA 30332-0150, U.S.A, doi:10.1007/978-3-642-79654-8_81.

[19]. Importance of Artificial Intelligence in Computer-Aided Design, December 2020. Sakata, I. F., and Davis, G. W., "Evaluation of Structural Design Concepts for an Arrow-Wing Supersonic Cruise Aircraft", NASA CR-2667,page:1-13, May 1977.

[20]. J. U. Pillai, I. Sanghrajka, M. Shunmugavel, T. Muthuramalingam, M. Goldberg, and G. Littlefair, "Optimisation of multiple response characteristics on end milling of aluminium alloy using taguchi-grey relational approach," Measurement, vol. 124, pp. 291–298, 2018,doi: 10.1016/j.m.2018.04.052.

[21]. "datafloq.com/read/how-can-ai-improve-cadcam/", Emily Newton , October 8, 2021

[22]. Vora, L. S., Veres, R. E., Jackson, P. C., and Klahr, P.TIES: An Engineering Design Methodology and System. In Innovative A plications of Artificial Intelligence 2, Rappaport and Smith (eds.! AAAI, Menlo Park, California, 1990, Volume 17 Number 4 (1996),doi: 1233-1-10-20080129.

[23]. Bunny, W., Curson, S., DeSantis, J., Lemmer, J., Scollard, J., Smith, R. ESCAPE: An Expert System for Claims Authorisation and Processing In Innovative Aplications of Artificial Intelligence 2,Rappaport and Smith (eds.! AAAI, Menlo Park, California, 1990,Volume 17 Number 4 (1996),doi: 1233-1-10-20080129..

[24]. C. Guo, Z. Li, and Y. Luo, "Application of Artificial Intelligence in CNC Machining," in 2017 3rd International Conference on Mechatronics and Robotics Engineering (ICMRE), Nice, France,

2017, pp. 1-5, doi: 10.1109/ ICMRE.2017.27.

[25]. “AI-Based Surface Roughness Prediction Model for Automated CAM -Planning Optimization”, Lea Tonejca (ne ́e Plessing), Gernot Mauthner , Thomas Trautner, Valentina Ko ̈nig , Werner Liemberger, Institute of Production Engineering and Photonic Technologies, TU Wen,2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA) | 978-1-6654-9996-5/22/$3100 ©2022 IEEE | DOI: 10.1109/ETFA52439.2022.9921281.

Knowledge and Information Systems
https://doi.org/10.1007/s10115-023-01848-z

REGULAR PAPER

Aspect and orientation-based sentiment analysis of customer feedback using mathematical optimization models

Neha Punetha[1] · Goonjan Jain[1]

Received: 20 September 2022 / Revised: 17 November 2022 / Accepted: 13 February 2023

Abstract
Sentiment analysis is a natural language processing method used to assess data's positivity, negativity, and neutrality. Several techniques were suggested as ways to solve the sentiment analysis task. This study presents a novel multi-criteria decision-making (MCDM) and game theory-based mathematical framework for the sentiment orientation of reviews. We propose two frameworks: sentiment orientation tagger modal (SOTM) and aspect-based ranking modal (ABRM). The SOTM consists of the simple additive weighting (SAW) technique and the principle of Nash equilibrium from game theory to deduce the tag for the review dataset. We identify a review's sentiment as positive, negative, or neutral. In ABRM, we rank the aspects of the review using the preference selection index (PSI). We propose an unsupervised sentiment classification model that combines context, rating, and emotion scores with a mathematical optimization model. The effectiveness of our proposed model is comparable to the state-of-the-art models, as demonstrated by experimental results on three benchmark review datasets. We also establish the significance of the results through statistical analysis. The proposed model ensures rationality and consistency. The novel combination of the MCDM and game theory model with the reviews' context, rating, and emotion scores creates a new paradigm in sentiment analysis. Also, the proposed model is generalizable and can analyze sentiment in many fields.

Keywords Sentiment analysis · Game theory · Mathematical optimization · Multi-criteria decision-making

1 Introduction

Over the past decade, the World Wide Web has become one of the most important resources for consumers and vendors to compare and assess goods and services. Daily massive content production on social networking sites needs automated systems to manage and identify

✉ Goonjan Jain
Goonjan_jain@hotmail.com

Neha Punetha
nehapunetha80@gmail.com

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

Published online: 08 March 2023

opinions. As a result, numerous algorithms and methodologies for sentiment analysis have been proposed in the literature. Sentiment analysis is a subtask of natural language processing (NLP) that focuses on the automated analysis of human sentiments expressed in different situations, such as consumer reviews [1]. Opinion-rich internet reviews are significant for producers and customers as these reviews tend to be opinion-heavy. The most common questions consumers ask are which products are popular, which features are of high quality, and why people think these things are excellent or bad. However, due to the sheer volume of this data, manual analysis is prohibitively time-consuming and hence not feasible; therefore, online data processing is required [2, 3].

Customer satisfaction is a consumer's view or sentiment about the gap between their expectations and reality [4]. Customer reviews on online media become crucial since they may improve the popularity of the seller's goods and services. Many customers express their happiness through online reviews on TripAdvisor, Yelp, etc. Restaurant reviews on TripAdvisor and Yelp are still written in text format, and text mining is used to identify them as good or bad, depending on user response. Product, nutrition, and food preparation information are typically used to gauge restaurant customer satisfaction [5]. Customer satisfaction analysis is used by only a few restaurants, even though it is an effective means of raising service standards. Customer satisfaction surveys in restaurants, such as those conducted through reviews on TripAdvisor, are still comparatively rare [6].

This study provides an optimum strategy based on game theory and MCDM for analyzing the subjective text's sentiment. The proposed model enhances the task's precision and rationality. The fundamental concepts of game theory and MCDM and their relationships are discussed in Sect. 3.

1.1 Research problem

In recent years, there has been a significant increase in the number of articles published in sentiment analysis and opinion mining, making this a highly researched area. Huge machine-readable reviews are readily available online. These bulky data enhance the chance to develop more effective algorithms for sentiment analysis [6, 7]. Determining how emotions should be expressed in content and whether a sentence makes a favorable or unfavorable judgment about the subject is the main task in judging emotions [8, 9]. Sentiment analysis, therefore, requires verification of the sentiments expressed in the words and their relevance to the subject matter. The algorithms use fundamental methods that necessitate a quick processor and appropriate resources but fall short of what is necessary to deliver services focused on results [7]. The current study focuses on the following objectives.

(1) To propose a system that has higher accuracy of sentiment classification.
(2) To reduce the amount of human labor required for content analysis.
(3) To attain efficiency in terms of both space and time complexity.
(4) To propose a model that is unsupervised and independent of language and training.
(5) To propose a system that makes it easier for customers to make a purchase decision when you can anticipate their sentiments under critical situations.

This research aims to develop a novel decision-making system that assigns a polarity sentiment label to a review. Customer feedback is used to generate recommendations for the best alternative to a given feature set for a restaurant, with feedback from more knowledgeable diners providing the most weight. Text mining, sentiment analysis, and MCDM theory are included in the proposed algorithm. First, we analyze online product reviews using aspect-level sentiment analysis to determine the positive or negative trend based on the review comments. The unsupervised methodology comprises applying the MCDM method. Then, we use the non-cooperative game theory model, play games between two reviews, and assign a sentiment tag to each review. The proposed methodology's applicability was tested in a hotel selection scenario based on internet comments considering numerous hotel alternatives. Rating scores, polarity scores of each review comment, and the emotional scores of comments are used in this study.

1.2 Contributions

In our information, this is the first attempt to use an optimization strategy to explore sentiment analysis of reviews. The following are some of the article's significant contributions and novelties:

(1) We developed an unsupervised mathematical framework for analyzing sentiments using optimization techniques based on the MCDM and non-cooperative game models. We performed tertiary class sentiment classification of reviews.

(2) We performed two tasks: first, we performed sentiment orientation tagging of reviews; second, we performed the aspect-based ranking of the food delivery aspects, i.e., food, delivery, and service.

(3) The proposed model employs optimization techniques of Game theory and MCDM to enhance precision while reducing the time and space required to run the algorithm. We used textual feedback and star ratings of reviews to classify sentiment into tertiary classes.

(4) We implemented the proposed algorithm on three different review datasets, taking context, rating, and emotional parameters into account. A domain-independent and language-independent model is presented, which can be used on any domain and any language with minor modifications.

The structure of the paper can be summarized as follows: In Sect. 1, we go over the basics of sentiment analysis, MCDM, and game theory. Research in sentiment analysis and game theory is summarized in Sect. 2. The preliminary theories connected to MCDM and Game theory are covered in Sect. 3. The proposed sentiment analysis algorithm is presented in Sect. 4. In Sect. 5, we evaluate the effectiveness of the proposed model against alternative strategies using three datasets of restaurant reviews. We also validated results using statistical significance. Finally, in Sects. 6 and 7, we summarize key findings, perform error analysis, and make recommendations for future directions.

2 Related work

This section provides a chronological overview of the research conducted on sentiment analysis. We begin by discussing lexicon-based methods. Then, some of the machine learning methods were reviewed. To wrap up, we discuss the topic modeling-based approaches that describe the identified research need and the method offered to address the identified research gap. In the end, we discuss some of the game theory and MCDM-based approaches, which motivated us to explore both these techniques in the sentiment analysis task.

Lexicon-based approaches primarily make use of reference materials like dictionaries and corpora. Many lexicon-based methods have been proposed in the literature by Mohammad et al. [8], Giatsoglou et al. [9], Bravo-Marquez et al. [10], and Bollegala et al. [11]. The approaches by these authors are explained below in detail.

Mohammad et al. [8] compiled a 2012 US Presidential Election tweets dataset. Using punctuation marks, emojis, hashtags, and dictionaries were studied with a total accuracy of 56.84%. Giatsoglou et al. [9] used various vector representations, including lexicon-based, word embedding-based, and hybrid vectorizations. The techniques were compared using four datasets of bilingual user reviews from the web. A hybrid approach combining word2vec and lexicons yielded good results. Using word2vec, it achieved 74.49% accuracy on the Movie Review Polarity dataset and 87.80% on the IMDb Large Movie Review Dataset. Bravo-Marquez et al. [10] employed lexica to classify short microblogging messages. The authors conducted experiments with Sanders and SemEval and employed commonly available lexical. Bollegala et al. [11] created a thesaurus to increase feature vectors for sentiment categorization and used Logistic Regression. The method was tested on the Multidomain Sentiment Amazon Dataset with an overall performance of 80.91%.

The attention and machine learning methods in the literature have shown promising prospects and innovation. Attention-based aspect-based sentiment analysis was proposed by Liu et al. [12]. The authors suggested a co-attention mechanism to acquire missing data information. The authors created a new loss function because features and settings weren't matched. The model was created using pre-trained GloVe and BERT models and evaluated using the Twitter and Restaurant datasets, resulting in a successful aspect-based sentiment analysis. The author [13] proposed the TRG-DAtt model for sentiment analysis. It utilized the combination of a target relational graph (TRG) and a double attention network (DAtt) to decipher and evaluate emotional data using decision-making. Second, to acquire the DAtt and the sentiments behind the keywords and feedback, they created a dependency graph attention network (DGAT) and an interactive attention network (IAT). By compiling the TRG's semantic data, DGAT simulated its interdependencies. Zunic et al. [14] employed a corpus of drug reviews to assess the model's performance in the health and wellbeing domain, where this task was to elicit negative emotions. The study showed that the graph convolution method was superior to the more common deep learning architectures when applied to aspect-based sentiment analysis. For optimal use of syntactic information and emotional dependencies, the author [15] proposed an aspect-gated graph convolutional network (AGGCN) built on a phrase's dependency tree and performed a unique aspect gate to guide the encoding of aspect-specific information at the outset. Donadi [16] demonstrated how to develop a German Sentiment Analysis system using TensorFlow and human-labeled data sets. This study introduced machine learning and Tensor Flow and created a rudimentary RNN using the tools. This research focuses on results from combining data sets.

Numerous studies on the topic modeling technique have been published, and we discussed the most precise and up-to-date approaches. The author [17] introduced a new probabilistic modeling approach based on the joint sentiment topic (JST) model, which can identify sentiment and topic from the text. Reverse JST, a reparametrized variant of the JST model derived by switching the order of sentiment and topic generation in the modeling process, is investigated. To extract a hierarchical structure of aspect-based sentiments from unlabeled internet reviews, the author [18] proposed a hierarchical aspect sentiment model (HASM). Every part of a HASM system is organized like a tree. Each node contains a two-tiered tree, with the root representing a facet and the children representing the related emotional extremes.

Compared to other proposed aspect-sentiment joint models, the results show that the presented model improved the classification accuracy at the sentence level. The author [19] investigated methods to better mine customer reviews for opinions on specific aspects of a product or service offered on the web. Initially, the author employed the extracted aspect-dependent sentiment lexicons in several aspect-level opinion mining tasks, including implicit aspect identification, aspect-based extractive opinion summarization, and aspect-level sentiment categorization. Experimental results demonstrated that the ASUM + model acquired sentiment dictionaries that rely on several aspects. Pablos et al. [20] conducted an aspect-based sentiment analysis utilizing topic modeling and continuous word embedding. The author conducted evaluations for four languages and multiple domains. With a few seed words changed, the suggested method is simple to adapt to various domains and linguistic systems.

Recent years witnessed a proliferation of studies employing game theory and MCDM for sentiment analysis. An efficient system for predicting the emotional future based on game theory was proposed by Bu et al. [21]. Game theory was applied to word sense disambiguation by Tripodi et al. [22] and Jain and Lobiyal [23]. Game theory-integrated Wikipedia is the basis for Ahmad et al. [24] multi-document summarization system. Several approaches for rumor identification based on game theory were introduced in recent years, including work by [25–28] demonstrated models for sentiment analysis based on game theory. Keyword extraction was executed using Game Theory by the author [29]. The study [30] published by the author employed the evolutionary Game Theory and made query expansion more precise and time-efficient. To identify propaganda, Barfar employed Shapley value methods [31]. Researchers used the Bayesian game model for binary class sentiment analysis of written text [32]. Many studies used the MCDM approach for sentiment analysis to recommend the best hotel [33–37]. Some studies used MCDM techniques for selecting electric vehicle batteries [38].

The successful application of game theory and MCDM to various NLP problems inspired us to investigate the applicability of game theory and MCDM for NLP tasks. A novel unsupervised method for sentiment analysis is presented in this study. Using game theory and multi-criteria decision-making, we investigate how to improve text sentiment analysis by incorporating in background information, subjective evaluations, and expert opinion.

3 Research gap

Algorithms developed using machine learning (ML) provide a high degree of accuracy and may be easily modified to fit new circumstances. The major downsides of these methods are that they are laborious, limited in scope, and reliant on both human intervention and information tagging. Poor results and failure occur when there is no probabilistic foundation for the classification and the number of attributes greatly exceeds the number of samples. In comparison with other algorithms, training an ANN requires a large amount of time and frequently a sizeable dataset. Compared to other models, sentiment analysis can be implemented more easily with the help of deep neural network techniques like (DNNs), (CNNs), and (RNNs). Their complex architecture, high computational cost, and overfitting issues make training them more time-consuming [19, 39–43].

Sentiment analysis based on topic modeling is feasible. The goal of topic modeling is to extract the most salient themes, entities, or subjects from a text corpus by recognizing patterns within that corpus. But it fails when the text is shorter, and topic modeling is infamously harder to perform, such as when the corpus is made up of tweets rather than traditional news items. Due to a lack of context, brief texts present a challenge for tasks like topic detection and sentiment extraction due to a lack of data. The topic model's quality depends on its manipulation and refinement, which is often manual and necessitates time-consuming fine-tuning of model parameters. The problem of configuration is one of the most significant obstacles in subject modeling. Though several high-quality research projects have tackled the problems of topic modeling and sentiment analysis of brief texts, there is still room for improvement in terms of accuracy and efficiency in the models' output [44, 45].

Various studies have proposed supervised, semi-supervised, and unsupervised techniques for sentiment analysis. So this is the first time MCDM techniques have been applied to the review dataset. This study developed a sentiment orientation tagger for text, classified reviews as positive, negative, or neutral, and the priority of various aspects of restaurants. The proposed study is explained in depth in Sect. 3.

3.1 How the proposed methodology differs from others?

We designed a mathematical framework using MCDM and game theory in the proposed work. We performed two tasks: sentiment orientation tagging and aspect-based review ranking. We used the Simple Additive Weighting (SAW) method for sentiment orientation tagging. This method aggregates the context, rating, and emotion scores extracted from reviews and gives the ranking scores. Then, we play the non-cooperative game between two reviews and deduce their sentiment tags using the principle of Nash equilibrium. For aspect-based ranking, we consider three aspects of reviews, i.e., food, service, and delivery experienced by the customer. For this purpose, we used the Preference Selection Index (PSI) MCDM ranking method and ranked different aspects. The proposed model was tested with various evaluation metric measures. The validity of the proposed model was checked using statistical significance. The proposed model presents state-of-the-art performance using integrated MCDM and game theory.

4 Preliminaries

This section provides an overview of the methodologies utilized in this study. Section 3.1 introduces the game theory and non-cooperative games in sub-Sect. 3.2. Sections 3.3 and 3.4 provides the structure for multi-criteria decision-making and explain how they relate to game theory. There is a quick overview of SAW and PSI methods in Sects. 3.5 and 3.6, respectively.

4.1 Game theory

The study of mathematical models of tactical interactions between rational agents is known as game theory. It can be used in areas as diverse as computer science, logic, and systems analysis. According to the definition given by [46], to put it simply, game theory is "a set of situations in which each player needs to take into account the actions of other players to make effective decisions. In this scenario, everyone is considered to act reasonably. What happens next depends entirely on the choices made by the other participants [47]. Game theory's predictive power is beneficial in cooperative decision-making. A game has the following components.

- **Players:** Competitors whose sole motivation is to achieve victory.
- **Strategy profile:** The term "strategy profile" *(A*$_i$*)* refers to a group of players' decisions and actions.

$$A_i = \{s_1, s_2, ..., s_i\}$$

- **Payoff:** The reward a player receives for executing a specific strategy.
- **Best Response:** The best response function (*br*) of each player is the best strategy *(s_i)* one can make on a strategy profile *(A_i)*.

$$br_i(s_{-i}) = \{s_i \in A_i : u_i(s_i, s_{-i}) \geq u_i(s_i^k, s_{-i}); \forall s_i^k \in A_i\}$$

- **Nash Equilibrium:** A decision-making principle known as the Nash equilibrium states that players can attain the desired point, after which they cannot deviate from their initial strategy. i.e., if s^* is a Nash equilibrium, then $s_i^* \in br_i(s_{-i}^*)$. It is also known as the point where the players' best strategies intersect.
- **Strongly Dominated Strategy:** Irrespective of what other players are doing, a player who continuously employs a dominant approach has the best results. Every significantly dominant approach indicates Nash equilibrium and, therefore, the optimal response. In a two-player game (A, B) $\in \mathbb{R}^{m\times n}$, a strategy strongly dominates *s* if Eq. 1 holds for all strategies of the other player *t*.

$$u(s,t) < u(s^*, t) \tag{1}$$

- **Weakly Dominant Strategy:** In a two-player game, (A, B) $\in \mathbb{R}^{m\times n}$, A strategy s is weakly dominated by $s*$ if for all strategies of the other player *t* as depicted in Eq. 2.

$$u(s,t) \leq u(s^*, t) \tag{2}$$

4.2 Non-cooperative game

A non-cooperative game is a method for modeling and evaluating scenarios in which each player's best choices rely on their beliefs or expectations about their opponents' behavior. Non-cooperative games are generally analyzed through the game theory model, which tries to predict players' strategies and payoffs and find Nash equilibria. Non-cooperative models assume that players do not have a cooperation mechanism (to make binding commitments) outside the specified rule of the game.

4.3 MCDM

MCDM is a process that combines the performance of various alternatives based on a variety of qualitative and quantitative criteria to arrive at a solution that can be agreed upon by all alternatives involved [48]. There are two types of criteria in an MCDM problem. The benefit criteria are those that should be maximized. Cost criteria, on the other hand, are those that must be reduced to the minimum level. A typical MCDM problem with *m* alternatives (A_1, A_2, ..., A_m) and *n* criteria (C_1, C_2, ..., C_n) can be presented with the below expression.

$$M = [m_{ij}]_{m\times n},\ \ W = [w_j]_n$$

where *M* stands for the decision matrix, m_{ij} stands for the i^{th} alternative's performance concerning the j^{th} criterion, *W* stands for the weight vector, and w_j denotes the weight of the j^{th} criterion. Typically, the original decision matrix *M* is incomparable because multiple units of measure are used to convey various criteria. Since the 1960s, numerous MCDM techniques and approaches have been effectively applied in various areas. MCDM's objective is to help decision-makers choose an alternative that satisfies their needs and is consistent with their preferences, not just to recommend the best alternative. A thorough understanding

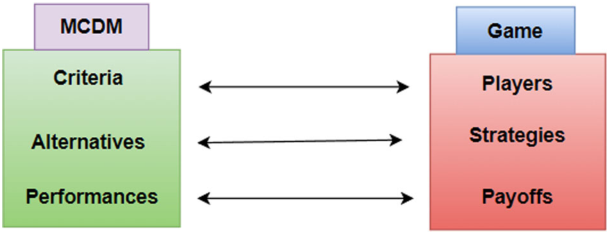


Fig. 1 MCDM–game relationship (color figure online)

of MCDM approaches and the participants' views is essential for efficient and successful decision-making in the early phases of the process. Multiple python[1] libraries are available to execute MCDM techniques smoothly.

4.4 MCDM as a game of strategy

An MCDM problem with p alternatives and q criteria is typically defined in the cardinal form. Equation 3 introduces the decision matrix of MCDM [49].

$$M = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1q} \\ m_{21} & m_{22} & \dots & m_{2q} \\ \dots & \dots & \dots & \dots \\ m_{p1} & m_{p2} & \dots & m_{pq} \end{bmatrix} \tag{3}$$

where m_{ij} is the performance of the i^{th} alternative under the j^{th} criterion for $i = 1, 2, \dots, m$ and $j = 1,2, \dots, n$.

The essential parts of an MCDM model are the criteria, the alternatives, and different ways each alternative meets each criterion. These components relate to the essential components of strategic games: participants, strategies, and payoff from expected outcomes. Figure 1 shows that the MCDM and game model have a direct correlation.

4.5 Preference selection index (PSI) method

Maniya and Bhatt [50] presented the PSI approach for material selection MCDM challenges. The PSI method determines criteria weights entirely on the information presented in the decision matrix, i.e., it uses an objective method, such as standard deviation or entropy, to determine criterion weights. The PSI technique, unlike the majority of MCDM methods, does not need a determination of the relative weight of the criteria. As a result, the technique is beneficial when determining the relative relevance of numerous criteria [51]. The PSI approach for handling MCDM problems has several steps in its primary application procedure and determines the purpose and the appropriate criteria for evaluating the alternatives.

[1] https://pypi.org/project/pymc/.

4.6 SAW method

SAW is commonly referred to as the method of adding weights. The underlying principle of the SAW technique is to find the weighted sum of the performance ratings for each alternative across all attributes [52]. The decision matrix *(M)* must be normalized for the SAW approach so that all possible alternative ratings can be compared on a single scale. The weighted sum of all attribute values calculates a candidate solution's final SAW score. The SAW method has three essential steps: normalizing the decision matrix *(M)*, applying the weight vector W, and figuring out the total score for each alternative [34].

5 Proposed methodology

In this work, we propose a Sentiment Orientation Tagger Model (SOTM) and an Aspect Based Ranking Model (ABRM). The objective of SOTM is to give a sentiment tag for each review. Similarly, the ABRM is designed to rank the three aspects of reviews, i.e., food, delivery, and service, and give an excellent, worst, and average tag based on the ranking. Figure 2 shows the pipeline of the proposed methodology. We took three parameters of reviews, i.e., context, rating, and emotion associated with the reviews. When these three parameters are given as input to SOTM, we get the sentiment orientation of each review. Similarly, when given as input to ABRM, we get a ranking of different aspects of the restaurant.

5.1 Sentiment orientation tagger model (SOTM)

SOTM aims to give a sentiment tag to a review. The pipeline of the SOTM model is shown in Fig. 3. The first step is to create a decision matrix consisting of three alternatives and

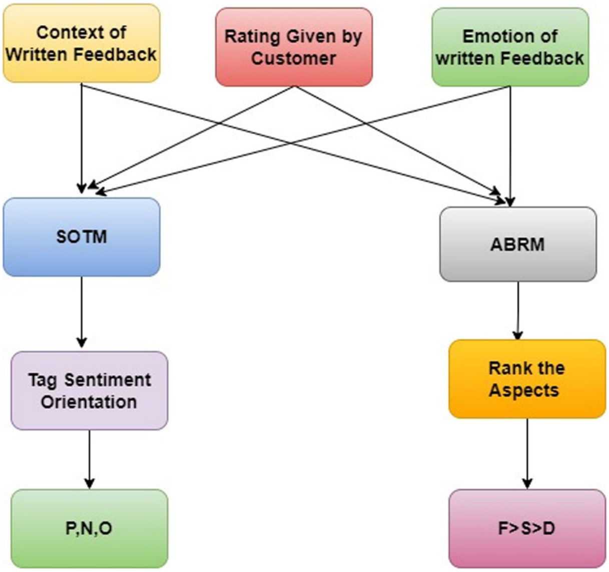
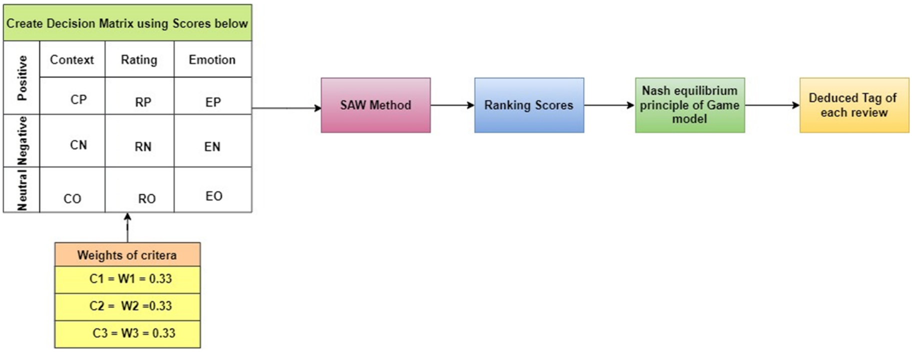

Context of Written Feedback
Rating Given by Customer
Emotion of written Feedback
SOTM
ABRM
Tag Sentiment Orientation
Rank the Aspects
P,N,O
F>S>D

Fig. 2 Pipeline of the proposed model (color figure online)



Create Decision Matrix using Scores below

	Context	Rating	Emotion
Positive	CP	RP	EP
Negative	CN	RN	EN
Neutral	CO	RO	EO

Weights of critera
C1 = W1 = 0.33
C2 = W2 =0.33
C3 = W3 = 0.33

SAW Method
Ranking Scores
Nash equilibrium principle of Game model
Deduced Tag of each review

Fig. 3 Pipeline of the SOTM model for sentiment tagging (color figure online)

three criteria. Three criteria are context, rating, and emotion, and the three alternatives are positive, negative, and neutral. Evaluations of scores (CP, CN, CO, RP, RN, RO, EP, EN, EO) corresponding to these criteria and alternatives are explained in detail in further steps. After creating the 3 × 3 matrices, the SAW method is implemented to get the ranking scores of each review. After the game, the model is applied to get the Nash equilibrium, and then, the final tag is deduced.

Step 1: Construct a decision matrix

To construct a decision matrix for SOTM, we consider the context of reviews, the emotion of reviews, and the rating given by the customer. The numeric scores are calculated by using Eqs. 4–9.

(i) Evaluation of Rating Score

The rating given with the review ranges from 1 to 5 stars. We use these star ratings to evaluate rating scores for positive, negative, and neutral sentiments of a review. We first calculate the degree of positive rating (DRP), degree of negative rating (DRN), and degree of neutral rating (DRO). In Eq. 4, DRP = *p,* where p is the actual rating of a review given by the customer. In Eq. 5, we calculate DRN, where we subtract DRP from 5 because 5 is the maximum rating of a review. We calculate DRO using Eq. 6. Then, we normalize DRP, DRN, and DRO to nullify one score's dominance over the other. After normalization, all scores are between 0 and 1. Normalization of rating scores (RP, RN, and RO) is done using Eqs. 7, 8, and 9.

$$\mathrm{DRP} = p \tag{4}$$

$$\mathrm{DRN} = (5 - \mathrm{DRP}) \tag{5}$$

$$\mathrm{DRO} = (5 - |\mathrm{DRP} - \mathrm{DRN}|) \tag{6}$$

$$\mathrm{RP} = \frac{\mathrm{DRP}}{\mathrm{DRP} + \mathrm{DRN} + \mathrm{DRO}} \tag{7}$$

$$\mathrm{RN} = \frac{5 - \mathrm{DRP}}{\mathrm{DRP} + \mathrm{DRN} + \mathrm{DRO}} \tag{8}$$

$$\mathrm{RO} = \frac{5 - |\mathrm{DRP} - \mathrm{DRN}|}{\mathrm{DRP} + \mathrm{DRN} + \mathrm{DRO}} \tag{9}$$

(ii) Evaluation of Emotion Scores

Emotions are classified as happy, angry, surprised, sad, or fear. In the current study, we neglect the fear emotion. Equation 10 shows the set of four emotions, i.e., Happy (H), Angry (A), Sad (S), and Surprise (S_{p}). We evaluate emotion scores using the text2emotion[2] library in python. Then, we categorize emotions E into three categories: positive emotion (EP), negative emotion (EN), and neutral emotion (EO). The values of EP, EN, and EO are evaluated using Eqs. 11–13. The range of values of EP, EN, and EO is between 0 and 1.

$$E = \{H,\ A,\ S,\ S_p\} \tag{10}$$

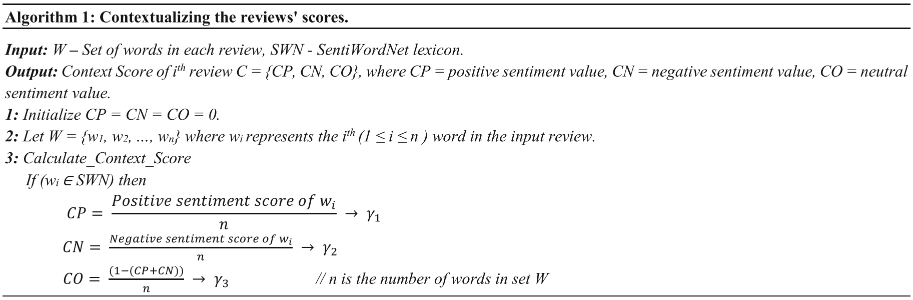
$$\mathrm{EP} = H + S_P \tag{11}$$

$$\mathrm{EN} = A + S + S_P \tag{12}$$

$$\mathrm{EO} = \frac{H + S + S_P + A}{2} \tag{13}$$

(iii) Evaluation of Context Scores

Context score is a numerical score that we create from textual reviewer feedback. We assign weights to each of the three possible contexts, positive (CP), negative (CN), and neutral (CO) context scores. To determine the context score for the textual comments, we used SentiWordNet (SWN) [35]. Words that have positive or negative connotations are recorded in the SWN database. We take the POS-tagged words in a review comment and add their positive and negative polarity values. Values for the context score can range from 0 to 1. We calculate the CP, CN, and CO values by following Algorithm 1. After calculating these scores, we named CP, CN, and CO as *γ_1, γ_2, and γ_3*. The input in Algorithm 1 is the customer's written reviews, and as the output, Algorithm 1 returns the normalized context scores (CP, CN, and CO).



Algorithm 1: Contextualizing the reviews' scores.

```
Input: W – Set of words in each review, SWN - SentiWordNet lexicon.
Output: Context Score of i^th review C = {CP, CN, CO}, where CP = positive sentiment value, CN = negative sentiment value, CO = neutral sentiment value.
1: Initialize CP = CN = CO = 0.
2: Let W = {$w_1, w_2, …, w_n$} where $w_i$ represents the i^th ($1 \le i \le n$) word in the input review.
3: Calculate_Context_Score
  If ($w_i \in$ SWN) then
      $CP = \frac{Positive\ sentiment\ score\ of\ w_i}{n} \rightarrow \gamma_1$
      $CN = \frac{Negative\ sentiment\ score\ of\ w_i}{n} \rightarrow \gamma_2$
      $CO = \frac{(1-(CP+CN))}{n} \rightarrow \gamma_3$          // n is the number of words in set W
```

Once all the scores are evaluated, we construct a decision matrix. Table 1 shows the constructed decision matrix. It consists of positive, negative, and neutral as the alternatives, and context, rating, and emotion as the criteria of the matrix.

Step 2: SAW Technique

In this step, we apply Algorithm 2 to tag the sentiment of reviews. There are two criteria in Algorithm 2. A beneficial criterion (BC) is one for which the highest value is aimed. The criteria are considered non-beneficial (NBC) when minimum values are sought. We

[2] https://pypi.org/project/text2emotion/.

Table 1 Alternatives and criteria of the decision matrix

Criteria → Alternative ↓	Context	Rating	Emotion
Positive	CP	RP	EP
Negative	CN	RN	EN
Neutral	CO	RO	EO

Table 2 Ranking scores of three sentiment orientations

Orientation	Ranking scores
Positive	λ_1
Negative	λ_2
Neutral	λ_3

considered three criteria, viz., rating, context, and emotions, and all are BC. Since none of the criteria are NBC, thus NBC is equal to zero. Next, we assign weights (W). We chose $W = 0.33$ because all criteria have equal weightage. Lastly, we evaluate the ranking scores, as shown in Table 2.

Algorithm 2: SAW technique to retrieve ranking scores

```
Input: positive, negative, and neutral score of rating (RP, RN, RO), emotion (EP, EN, EO), and context (CP, CN, CO)
Output: The combined score is the ranking Score ($\lambda_i$) of each review $R_i$.

1: Construct decision matrix $M = [M_{ij}]_{3\times 3}$ using the input scores.
2: Calculate normalized value ($\theta_{ij}$).
       $\theta_{ij} = \begin{cases} \frac{m_{ij}}{m_j^{\max}} = \wp \in BC \\ \frac{m_j^{\min}}{m_{ij}} = \xi = 0 \in NBC \end{cases} \quad \forall i, j \in \{1,2,3\}$
3: Assign weight, $w_j = 0.33$.
4: Calculate ranking scores for each review ($\lambda_i$).
       $\lambda_i = \sum_{j=1}^{n} w_j.(\theta_{ij})$
```

Step 3: Sentiment orientation tagging using the game model

Now, a non-cooperative game is played between players. For a non-cooperative game, we need two players *(R_1* and *R_2)* with positive, negative, and neutral strategies. Ranking scores in Table 2 are taken as the payoff for players. Ranking scores of R_1 are λ_1, λ_2, λ_3 and ranking scores of R_2 are ω_1, ω_2, ω_3. The possible combinations of ranking scores of R_1 and R_2 are shown in Table 3. To achieve the Nash equilibrium, we apply dominant strategies (DR_i) as given in Eq. 1 and Eq. 2. The strategies corresponding to the payoffs of Nash equilibrium are the deduced tag of each review. We follow Algorithm 3, implemented using Nashpy[3] library of python, to reach the Nash equilibrium.

[3] https://github.com/drvinceknight/Nashpy.

Table 3 Normal form representation of game played between two reviews

Players ↓→	R_2			
	Strategies ↓→	Positive	Negative	Neutral
R_1	Positive	(λ_1, ω_1)	(λ_1, ω_2)	(λ_1, ω_3)
	Negative	(λ_2, ω_1)	(λ_2, ω_2)	(λ_2, ω_3)
	Neutral	(λ_3, ω_1)	(λ_3, ω_2)	(λ_3, ω_3)

Algorithm 3: Deduce sentiment tag for review

```
Input: Ranking scores { $\lambda_1$, $\lambda_2$, $\lambda_3$} for review $R_i$ and { $\omega_1$, $\omega_2$, $\omega_3$} of review $R_j$.
Output: Sentiment Tag for $R_i$ and $R_j$, i.e., Ri, Rj ∈ $\{P, N, O\}$.

1: Generate a Normal form matrix for players $R_i$ and $R_j$ using the Ranking scores.
2: Compute dominant strategies $DR_i$ for $R_i$ and $R_j$.
3: Compute Nash equilibrium (NE), where NE = $DR_i \bigcap DR_j$.
4: The strategies corresponding to NE are the sentiment tags for reviews $R_i$ and $R_j$.
```

5.2 Aspect-based ranking model (ABRM)

ABRM is designed to rank aspects so that customers can conveniently get help making the right decision. This model includes two steps. The first step includes evaluating food, service, and delivery quality scores. The second step aggregates these three scores to calculate the PSI score and rank each aspect in descending order. Figure 4 shows the steps required in the ABRM model.**Step 1: Evaluation of Aspect-based scores**

We consider three aspects that improve restaurants' performance: food quality, service quality, and delivery quality of reviews. For implementing MCDM, we need the numeric scores of these aspects. So the first step in ABRM is the evaluation of these scores.

(i) Evaluation of Food Quality Score

We get to know the food quality of the restaurant using written feedback, so we used SWN lexicon-based approach to evaluate whether the sentiment behind the food quality is positive (FP), negative (FN), or neutral (FO).

(ii) Evaluation of Service Quality Score

The service quality of restaurants is evaluated using emotions in the written feedback. We use the python-based library to calculate the emotions. We evaluate whether the sentiment

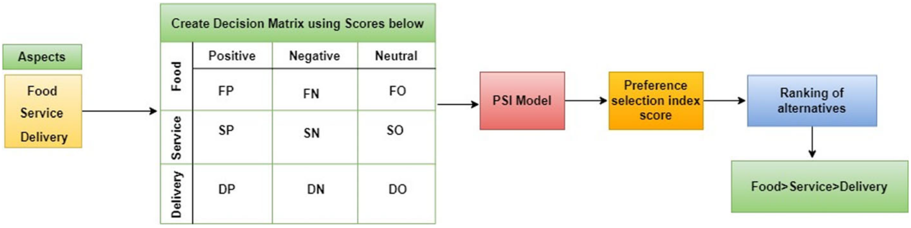


Fig. 4 Pipeline of the ABRM model for aspects ranking (color figure online)

of service quality is positive (SP), negative (SN), and neutral (SO) using Eqs. 14 - 16.

$$\mathrm{SP} = H + S_P \tag{14}$$

$$\mathrm{SN} = S + A + S_P \tag{15}$$

$$\mathrm{SO} = \frac{S + A + S_P + H}{2} \tag{16}$$

iii) Evaluation of delivery quality score

To evaluate the delivery quality score, we use the delivery rating given by customers to the delivery services. The sentiment about the delivery service can be positive (DP), negative (DN), or neutral (DO). The values are evaluated using Eqs. 17 - 19. In the given equations, *p* is the rating given for the product. The range of DP, DN, and DO is between 0 and 1.

$$\mathrm{DP} = \mathrm{p} \tag{17}$$

$$\mathrm{DN} = 5 - \mathrm{DP} \tag{18}$$

$$\mathrm{DO} = 5 - |\mathrm{DP} - \mathrm{DN}| \tag{19}$$

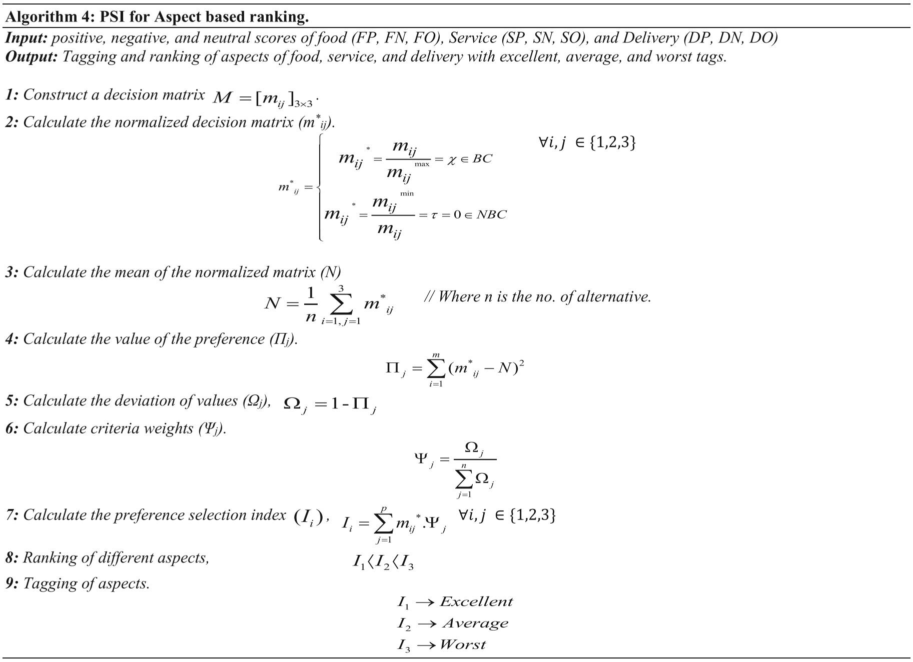
The decision matrix containing the three alternatives (Food, Service, and Delivery) and their criteria (Positive, Negative, and Neutral) is given in Table 4. Here alternatives are aspects, and criteria are the sentiment tags of each review.**Step 2:** We apply the PSI MCDM technique for ranking the aspects. These rankings help us get priority while ordering food from an online service. This step is implemented in Algorithm 4. Three criteria, i.e., positive, negative, and neutral, were considered in this step. All three criteria are BC; thus, NBC equals zero. Then, as a next step, we normalize these criteria. Depending upon the ranking, the alternative with the highest rank gets the excellent tag, and the last-ranked alternative gets the worst.

Table 4 Decision matrix consists of alternatives and criteria

Criteria → Alternative ↓	Positive	Negative	Neutral
Food	FP	FN	FO
Service	SP	SN	SO
Delivery	DP	DN	DO

Table 5 Preference selection index of three aspects

Aspects	Preference selection indexes (I_i)
Food	I_1
Service	I_2
Delivery	I_3



```
Algorithm 4: PSI for Aspect based ranking.
Input: positive, negative, and neutral scores of food (FP, FN, FO), Service (SP, SN, SO), and Delivery (DP, DN, DO)
Output: Tagging and ranking of aspects of food, service, and delivery with excellent, average, and worst tags.

1: Construct a decision matrix $M = [m_{ij}]_{3\times3}$.
2: Calculate the normalized decision matrix ($m^*_{ij}$).
       $m^*_{ij} = \begin{cases} m_{ij}^* = \frac{m_{ij}}{m_{ij}^{\max}} = \chi \in BC \\ m_{ij}^* = \frac{m_{ij}^{\min}}{m_{ij}} = \tau = 0 \in NBC \end{cases}$    $\forall i, j \in \{1,2,3\}$
3: Calculate the mean of the normalized matrix (N)
       $N = \frac{1}{n} \sum_{i=1, j=1}^{3} m^*_{ij}$    // Where n is the no. of alternative.
4: Calculate the value of the preference ($\Pi_j$).
       $\Pi_j = \sum_{i=1}^{m} (m^*_{ij} - N)^2$
5: Calculate the deviation of values ($\Omega_j$), $\Omega_j = 1 - \Pi_j$
6: Calculate criteria weights ($\Psi_j$).
       $\Psi_j = \frac{\Omega_j}{\sum_{j=1}^{n} \Omega_j}$
7: Calculate the preference selection index $(I_i)$, $I_i = \sum_{j=1}^{p} m_{ij}^* . \Psi_j$    $\forall i, j \in \{1,2,3\}$
8: Ranking of different aspects,    $I_1 \langle I_2 \langle I_3$
9: Tagging of aspects.
       $I_1 \rightarrow Excellent$
       $I_2 \rightarrow Average$
       $I_3 \rightarrow Worst$
```

After applying Algorithm 4, we get the results illustrated in Table 5. We rank them accordingly and give them suitable tags depending upon the ranking of alternatives. The highest PSI score gets the excellent tag, and the lowest PSI score gets the worst tag.

$$I_1 \succ I_2 \succ I_3 \text{ and } \begin{array}{l} I_1 \rightarrow \text{Excellent} \\ I_2 \rightarrow \text{Average} \\ I_3 \rightarrow \text{Worst} \end{array}$$

5.3 Illustrative Example

Consider the following two reviews: R1 and R2. Each review consists of written feedback from customers and corresponding ratings based on individual experiences.

R1 (1 star): "Very bad service provides by swiggy no any customer care help and very rude answer given by riders very poor service."

R2 (4 star): "Absolutely fantastic platform for online food ordering & delivery within estimate time. have a great deal with every time to give better & best satisfaction."

5.3.1 SOTM algorithm

The SOTM algorithm gives the sentiment orientation of the written feedback by the customer. We first construct the decision matrix shown in Table 6 by following Algorithm 1. Following Algorithm 2, we get the ranking scores mentioned in Table 7 for *R1* and *R2*. We apply the non-cooperative game model using Algorithm 3. Table 8 shows the normal form representation of the game model; then, using the principle of Nash equilibrium, we deduce the sentiment orientation of both reviews highlighted in bold in Table 8.

Table 6 Decision matrix consists of alternatives and criteria of R1 and R2

R1 (1 stars)				R2 (4 stars)			
Alternatives	Context	Rating	Emotion	Alternatives	Context	Rating	Emotion
Positive	0.0127	0.1250	0.0000	Positive	0.1266	0.5000	0.6600
Negative	0.0706	0.5000	0.1100	Negative	0.0588	0.1250	0.3300
Neutral	0.0579	0.3750	0.0825	Neutral	0.0678	0.3750	0.1650

Table 7 Ranking scores of R1 and R2

R1		R2	
Alternatives	Appraisement score	Alternatives	Appraisement score
Positive	0.0127	Positive	0.8010
Negative	0.0706	Negative	0.1667
Neutral	0.0579	Neutral	0.4166

Table 8 Non-cooperative game played between R1 and R2

R1	R2			
		Positive	Negative	Neutral
	Positive	(0.0127,0.8010)	(0.0127, 0.1665)	(0.0127, 0.4166)
	Negative	**(0.0707, 0.8010)**	(0.0706, 0.1666)	(0.0705, 0.4166)
	Neutral	(0.0579, 0.8010)	(0.0579, 0.1666)	(0.0579, 0.4166)

Bold value indicates the Nash equilibrium of the game played

5.3.2 Implementation of ABRM

The objective of constructing ABRM is to rank restaurants' different aspects (Food, Service, and Delivery). Table 9 shows the decision matrix with numeric values and all the alternatives and criteria. Following Algorithm 4, we generate PSI scores illustrated in Table 10, then we rank them, and using Algorithm 4, we arrange them in descending order and give a tag according to the ranking.

$$\text{Delivery} \succ \text{Food} \succ \text{Service} \quad \begin{array}{l} \text{Delivery} \rightarrow \text{Excellent} \\ \text{Food} \rightarrow \text{Average} \\ \text{Service} \rightarrow \text{Worst} \end{array} \quad \text{For R1} \tag{20}$$

$$\text{Delivery} \succ \text{Service} \succ \text{Food} \quad \begin{array}{l} \text{Delivery} \rightarrow \text{Excellent} \\ \text{Food} \rightarrow \text{Average} \\ \text{Service} \rightarrow \text{Worst} \end{array} \quad \text{For R2} \tag{21}$$

Equation 20 shows the ranking of each review's aspect, which shows that according to reviewer *R1*, delivery is excellent, food quality is average, and service is worst. Similarly, Eq. 21 shows that according to reviewer *R2,* delivery is excellent, service is average, and food is worst.

6 Result and performance evaluation

6.1 Data collection

To evaluate the efficacy and efficiency of the proposed model, we collected three datasets. The first dataset was collected from two sources, i.e., Zomato and Swiggy [36]. The second and third datasets consist of yelp and TripAdvisor restaurant reviews, respectively [4]. Ratings

Table 9 Numeric value alternative and criteria

	R1 (1 stars)				R2 (4 stars)		
Alternatives	Context	Rating	Emotion	Alternatives	Context	Rating	Emotion
Positive	0.0127	0.1250	0.0000	Positive	0.1266	0.5000	0.6600
Negative	0.0706	0.5000	0.1100	Negative	0.0588	0.1250	0.3300
Neutral	0.0579	0.3750	0.0825	Neutral	0.0678	0.3750	0.1650

Table 10 Preference selection indexes values of different aspects of the delivery app

	R1 (1 stars)			R2 (4 stars)	
Alternatives	PSI	Rank	Alternatives	PSI	Rank
Service	0.6480	3	Service	0.1622	2
Delivery	0.8520	1	Delivery	1.0000	1
Food	0.7397	2	Food	0.1356	3

Table 11 Data statistics of the crawled dataset

Crawled dataset	Language	Positive	Negative	Neutral
Zomato + Swiggy	English	644	145	211
Yelp	English	567	245	188
TripAdvisor	English	369	456	175

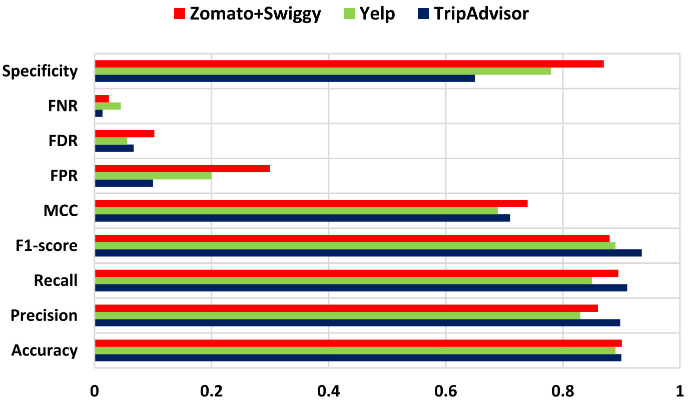


Fig. 5 Performance of evaluation metrics over three datasets (color figure online)

and comments are included with reviews in each dataset. Table 11 provides data statistics for the three datasets. After collecting data, we pre-process the data. The term "pre-processing" describes the operations performed on the data before it is fed into the algorithm. It is a method used to clean up unstructured data. That is to say, the data are continuously gathered in a raw format that is impractical for analysis if it is gathered from several sources. A well-formatted data set is essential for getting the most out of the applied model. We performed tokenization[4], lemmatization[5], and stop word[6] removal over the datasets to clean the reviews.

6.2 Statistical measurements over three datasets

Numerous evaluation indicators were used to calculate the proposed framework's robustness. Among these are the F1-measure, accuracy, precision, MCC, etc. Because a model's performance may be satisfactory according to one evaluation criteria but subpar according to another, it is essential to analyze using many metrics. The performance of the proposed model on the three datasets, as measured by various assessment indicators, is depicted in Fig. 5. The formulas to calculate the metrics are given from Eqs. 22–30.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \tag{22}$$

[4] From nltk tokenize import word_tokenize.

[5] From nltk.stem import WordNetLemmatizer.

[6] From nltk.corpus import stopwords.

$$\text{Precision} = \frac{\text{TP}}{\text{TP + FP}} \tag{23}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}} \tag{24}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN + FP}} \tag{25}$$

$$\text{False positive rate (FPR)}=\frac{\text{FP}}{\text{FP + TN}} \tag{26}$$

$$\text{False negative rate (FNR)}= \frac{\text{FN}}{\text{TP + FN}} \tag{27}$$

$$\text{False discovery rate (FDR)}= \frac{\text{FP}}{(\text{TP + FP})} \tag{28}$$

$$F1_{\text{score}} = \frac{(2 * \text{precision} * \text{recall})}{(\text{precision} + \text{recall})} \tag{29}$$

$$\text{Matthews correlation coefficient (MCC)} = \frac{(\text{TP*TN - FP*FN})}{\sqrt{(\text{TP + FP})*(\text{TN + FN})*(\text{FP + TN})*(\text{TP + FN})}} \tag{30}$$

where TP = True positive, TN = True Negative, FP = False Positive, and FN = False Negative.

When evaluating a model's performance, accuracy is defined as the proportion of correct predictions (positive negative and neutral) relative to the total number of predictions. Zomato + Swiggy dataset has 0.9 accuracy, and the TripAdvisor dataset is 0.89 accurate. The recall is the proportion of positive cases adequately predicted by the model out of the total number of positive cases. The recall is as follows: 0.895 for the Zomato + Swiggy dataset, 0.85 for Yelp, and 0.91 for TripAdvisor. Specificity is the percentage of negative cases accurately predicted by the model out of the total number of negative cases. Specificity for the Zomato + Swiggy dataset, Yelp, and TripAdvisor is 0.87, 0.78, and 0.65, respectively. Precision measures how often a model correctly predicts a positive instance out of the total number of positive cases anticipated by the model. Precision is recorded at its highest for TripAdvisor, around 0.898; for Yelp, it is 0.83; and for Zomato + Swiggy, it is 0.86. The F1 score represents the harmonic mean of the precision and recall scores. The F1 score provides a simple way to evaluate and compare the results of different models. The F1-score for TripAdvisor is 0.935 for the Zomato + Swiggy dataset, or 8.89 and 0.89, respectively.

Alternatively, the Matthews correlation coefficient (MCC) is a more trustworthy statistical rate that yields a high score only if the forecast is accurate. The MCC score is recorded highest for the Zomato + Swiggy dataset, which is 0.7399; 0.69 for Yelp; and 0.71 for TripAdvisor. The FPRs for the Zomato + Swiggy dataset, Yelp, and TripAdvisor are 0.3, 0.2, and 0.1. The FDRs for the Zomato + Swiggy dataset, Yelp, and TripAdvisor are 0.1, 0.5, and 0.067. The FNR for the Yelp dataset is the highest at 0.045; for the Zomato + Swiggy dataset, it is 0.0247; and for the TripAdvisor dataset, it is 0.0136.

6.3 Macro-, Micro-evaluation

We employed macro- and micro-averages to measure overall performance while working with various datasets. Macro-average independently computes metrics for each class before averaging them (thus treating all the classes equally). The macro precision, recall, and F1-score are 0.57, 0.58, and 0.58. Micro-average computes the mean metric by adding each class's

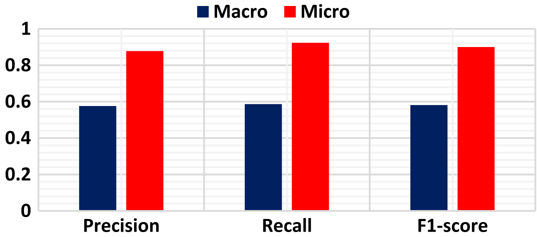


Fig. 6 Macro-, Micro-performance evaluation metrics over three datasets (color figure online)

contributions. It is a helpful statistic for determining performance when the size of datasets fluctuates. The micro-precision, recall, and F1-score are 0.87, 0.92, and 0.90. Equations 31–36 describe the macro- and micro-averaged accuracy, F-score, and recall across the n datasets. Overall performance estimates are calculated using global and local averages as we navigate many data sets. The results across all three datasets are shown in Fig. 6.

$$\text{Macro - averaged Precision} = \frac{\sum_{i=1}^{n} \text{Precision}_i}{n} \tag{31}$$

$$\text{Macro - averaged Recall} = \frac{\sum_{i=1}^{n} \text{Recall}_i}{n} \tag{32}$$

$$\text{Macro - F - score} = 2.\frac{\text{Macro} - \text{Precision}_i \times \text{Macro} - \text{Recall}}{\text{Macro} - \text{Precision}_i + \text{Macro} - \text{Recall}} \tag{33}$$

$$\text{Micro - averaged precision} = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} (TP_i + FP_i)} \tag{34}$$

$$\text{Micro - averaged recall} = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} (TP_i + FN_i)} \tag{35}$$

$$\text{Micro - averaged F - score} = 2.\frac{\text{Micro} - \text{Precision}_i \times \text{Micro} - \text{Recall}}{\text{Micro} - \text{Precision}_i + \text{Micro} - \text{Recall}} \tag{36}$$

We implemented the proposed models on three review datasets to test the performance. Figure 7 shows the performance of SOTM over three datasets. Zomato and swiggy have the highest negative feedback from customers, and Yelp shows the maximum neutral feedback. Figure 8 shows the aspect-based performance of ABRM over three datasets. TripAdvisor is excellent in food quality and has a least rank in delivery. Similarly, Yelp is best in service and worst in delivery. Zomato and Swiggy have the best delivery of food. The primary objective of the SOTM + ABRM is to give the overall quality of aspects and sentiment orientation of the written feedback.

Figure 7 illustrates the performance of the SOTM across three datasets that yield sentiment classification outputs (positive, negative, and neutral). The largest proportion of Yelp reviews

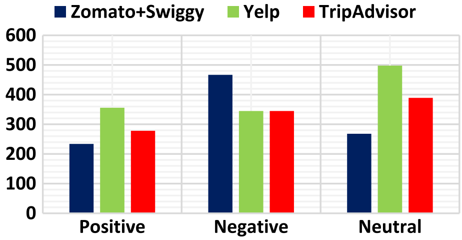


Fig. 7 Output of the SOTM across three datasets in terms of sentiment orientation (color figure online)

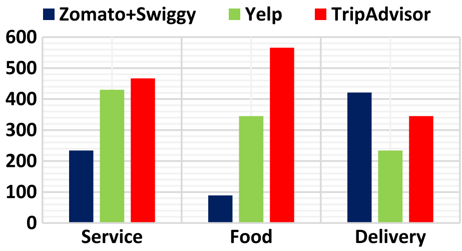


Fig. 8 Output of the ABRM across three datasets in terms of aspects ranking (color figure online)

is positive or neutral. The number of negative reviews for Zomato and Swiggy is the highest. The proportion of positive, negative, and neutral TripAdvisor reviews is about average. The results of ABRM on three datasets, which generate rankings for three aspects, are displayed in Fig. 8. Most people seem to rate their delivery experience using Zomato and Swiggy, whereas TripAdvisor is the best option for rating their overall experience. Yelp has more service-related comments than any other platform in the sample.

It is noted that TripAdvisor ranks relatively low for delivery yet high for the quality of its meal reviews. In a similar vein, Yelp excels in service but fails miserably at delivery. The top food delivery services are Zomato and Swiggy. The primary function of the SOTM + ABRM is to provide an overarching assessment of the ideas' quality in all areas and the emotional orientation of the written response.

6.4 Comparison over zomato + swiggy dataset

In this section, we compare the proposed model with various approaches by Anas and Kumari [37], Gojali and Khodra [53], Al Omari et al. [39], and Jagdale and Deshmukh [36]. Anas and Kumari [37] used Naïve Bayes and the random forest method for opinion mining of reviews. Gojali and Khodra [53] used the WordNet approach to predict the reviews' orientation and aspect of the reviews, and the recorded F-measure is 0.783 (precision and recall are the same) over the Zomato dataset [40] proposed sentiment attribution analysis with hierarchical classification and automatic aspect categorization to improve social listening for diligent marketing. He proposed five models, out of which SVM on Hierarchical Classification (Hybrid) gives the best result. Al Omari et al. [39] perform a logistic regression algorithm on the Zomato dataset and perform a sentiment orientation task *(P, N, O)*. Jagdale and Deshmukh [36] performed sentiment analysis of the Zomato dataset using a supervised machine learning classification algorithm like gradient boosting. All the comparison based on four evaluation metrics is depicted in Fig. 9. The proposed model outperforms in all respects

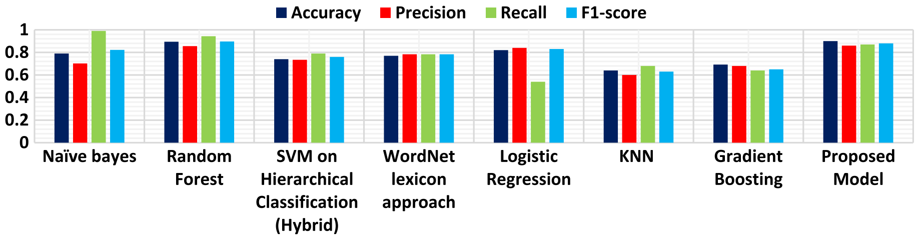


Fig. 9 Performance comparison of the proposed model with existing approaches (color figure online)

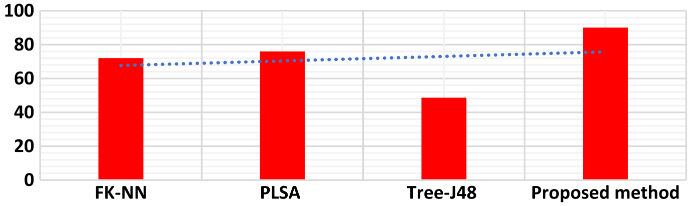


Fig. 10 Performance comparison with the supervised approaches (color figure online)

having an accuracy 90%, F1-measure 0.88, the recorded precision of the proposed model is 0.86, and recall is 0.87.

6.5 Comparison on trip advisor dataset

We analyzed the TripAdvisor dataset and compared the suggested model to other popular models. To evaluate these methods, we used supervised models such as PLSA [41], FK-NN [42], and Decision Tree-J48 [43]. Probabilistic Latent Semantic Analysis (PLSA) was proposed by Khotimah et al. [41] to study customer evaluations by counting how often certain words appeared within specific documents. PLSA assigned positive and negative connotations to words so they could be sorted into categories. The Fuzzy K-Nearest Neighbor (FK-NN) method, which combines the Fuzzy and K-Nearest Neighbor approaches, was applied for sentiment analysis by Billiyan et al. [42]. Positive and negative evaluations from customers are separated using this mixed-sentiment model. On this data set, PLSA outperformed the FK-NN. There is a 0.72 percent difference between PLSA and FK-NN accuracy. We also compared our findings with aksono et al. [43].'s Decision Tree-J48. The comparison is shown visually in Fig. 10.

6.5.1 Comparison with the unsupervised approaches

Afzaal et al. [54] suggested fuzzy logic models, including FURIA, FLR, FNN, FRNN, and VQNN, which extracted several features to categorize users' viewpoints. Fuzzy Lattice Reasoning (FLR) performs the best in comparison with the other five models. FLR is constituted of fuzzy lattice rules. When compared to the proposed model, FLR performs poorly. The inconsistent application of Fuzzy Logic models to solve problems may contribute to unsatisfactory results. Extensive testing is required for system validation and verification. Zuheros et al. [55] offered DOC-ABSADeepL as a more creative decision-making tool. Expert judgments were derived from numerical ratings and reviews written in a natural language. Multi-tasking deep learning models like DOC-sentiment ABSADeepL's analysis capability allow it to break down expert reviews into their constituent categories and pull out the most relevant findings. Figure 11 shows the comparisons of various models based on four parameters.

6.6 Comparison on yelp dataset

The performance of the proposed model is compared with various supervised models, i.e., Naïve Bayes [56], Support Vector Machine, and Genetic algorithm [57], XGBoost [58],

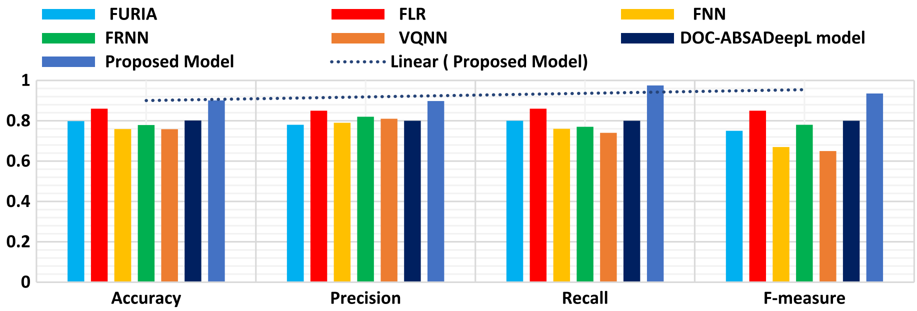


Fig. 11 Evaluation of the proposed model in comparison with an unsupervised method (color figure online)

Table 12 Comparison of the proposed model with the supervised model

Supervised Algorithm	Accuracy	Error rate
Naïve Bayes [56]	79.12	20.88
Support Vector Machine [57]	85.20	14.80
Genetic Algorithm [57]	85.30	14.70
XGBoost [58]	83.00	17.00
SVM_FDO (Fuzzy Domain Ontology) [59]	79.59	20.41
Proposed Model	89.01	11.00

and SVM_FDO (Fuzzy Domain Ontology) [59]. Hemalatha and Ramathmika [56] implemented the machine learning algorithms over the yelp dataset. Govindarajan [57] proposed a comparative study of the effectiveness of ensemble techniques for sentiment classification in which SVM and Genetic algorithms worked best. Nasim and Haider [58] presented the Aspect-Based Sentiment Analysis (ABSA) Toolkit developed to perform aspect-level sentiment analysis on customer reviews. Luo and Xu [59] proposed a classification algorithm based on SVM and FDO algorithms. Table 12 shows the results of the experiment. The proposed model has the least error rate as compared to other approaches. The maximum error rate was recorded by Naïve Bayes and SVM_FDO, i.e., 20.88.

6.6.1 Performance on unsupervised learning models

To evaluate the performance of the proposed model, we compared our polarity classification result with JST [17], ASUM [60], HASM [18], TSM [61], and ASUM + [19]. For evaluating maximum and minimum accuracy, we make the two samples of the dataset, i.e., a small sample of 1000 reviews (all of which were given 1 or 5 stars) and a larger sample of 1000 reviews (all of which were given 2 or 4 stars). Figure 12 represents the maximum and minimum accuracy range of different models over big and small datasets. ASUM has the range 76–79%, ASUM + has the range 84–86%, JST + has the range 60–61%, TSM + has the range 52–54%, and proposed model has range 85–89%. Compared to other unsupervised approaches, the proposed model outperformed those shown in Fig. 12. The enhanced accuracy of the proposed model is due to its independency of the training and language. The decreased performance of these models is because the manipulation and refining of the topic model are essential to

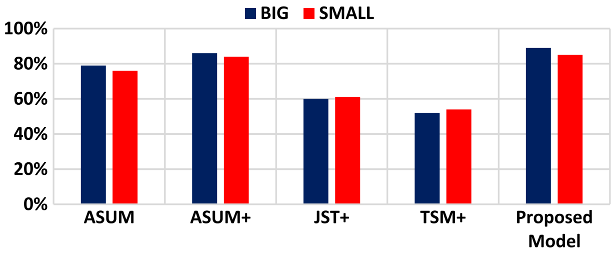


Fig. 12 Performance evaluation on the yelp dataset of different sentiment classification methods (color figure online)

Table 13 Z-test statistics for comparing two proportions across data sets

Parameters	Sample 1	Sample 2
Sample Size(n_I)	1000	500
Sample Proportion (p_I)	0.91	0.89
Favorable Cases (X_I)	910	448

its quality, but they are often performed manually and require extensive fine-tuning of model parameters, both of which take considerable time. In topic modeling, the configuration is a major obstacle.

6.7 Significance of statistics in two different data sets

Two proportion tests were conducted using data from hotel reviews and movie reviews, and both were subjected to the Z test. We used a random sample of 1000 reviews (sample proportion = 0.91), of which 910 were correctly labeled. Sample 2 also had many reviews (n_2) = *252*, but only 410 out of 500 were correctly labeled (a proportion of 0.896). For two population proportions (p1 and p2), we used the Z-test, with H_o and H_a as possible null hypotheses. Table 13 provides a concise summary of the information provided**.**

H$_\text{o}$ $p_1 = p_{2,}$ i.e., the accuracy of sample 1 = accuracy of sample 2.

H$_\text{a}$ $p_1 \neq p_{2,}$ i.e., the accuracy of sample 1 = accuracy of sample 2.

Calculating the value of the aggregated proportions

$$P = \frac{X_1 + X_2}{N_1 + N_2} = \frac{910 + 448}{1000 + 500} = 0.9053 \tag{37}$$

For this purpose, we used a two-tailed test called a z-test on two population proportions. The z-score was determined with the help of Eq. 38.

$$z = \frac{p_1 - p_2}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}} = \frac{0.91 - 0.896}{\sqrt{0.9053 \times (1 - 0.9053) \times (1/1000 + 1/500)}} = 0.873 \tag{38}$$

The accepted and crucial regions of the previously stated theory are depicted graphically in Fig. 13. The null hypothesis Ho could not be rejected. As a result, there is insufficient

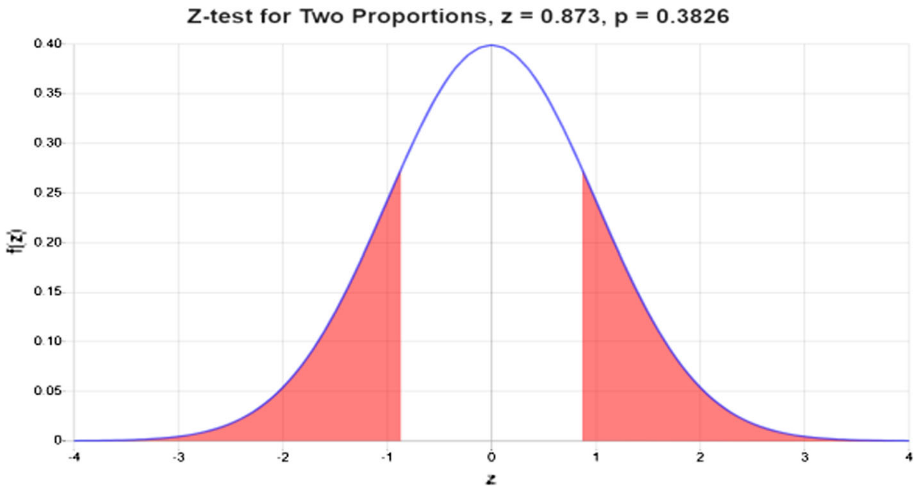


Fig. 13 A graphical illustration of the hypothesis's critical region (color figure online)

data to claim that, at $\alpha = 0.05$ level of significance, the population proportion p1 varies from p2. At a significance level of $\alpha = 0.05$, our null hypothesis is accepted, indicating that proportions 1 and 2, or the proposed model's accuracy over the hotel review dataset and the proposed model's accuracy over the movie review dataset, are identical. The proposed model's accuracy is constant across diverse sample sizes and datasets.

7 Discussion and error analysis

The proposed algorithm has its benefits and limitations. The following are the most notable advantages and disadvantages of the proposed model.

(1) Time complexity determines how many operations it has to conduct concerning the input dataset size to fulfill its task. The algorithm's space complexity measures how a small amount of space it consumes during execution for varying input sizes. Table 14 shows the algorithm's run time and space complexity in different cases where *m* is the number of alternatives and *n* is the number of criteria. The complexity comes under the P-class problem solvable in polynomial time. It is also a deterministic algorithm which means the algorithm that continuously computes the correct answer. The time and space complexity are linear and calculated in polynomial time, which is reduced by half using game theory. This is because, in the current study, we played the game between two players (*R1* and *R2*) using game theory, halving the time required to tag two reviews.

(2) A critical flaw in the proposed method is that if a term is unavailable in the sentiment lexicon, the system cannot classify customer textual feedback accurately. The sentiment analysis of English text's opinion terms is built on the SWN lexicon. The shortcomings of SWN are that not enough words are covered, and some words are not given the appropriate polarity score. Table 15 presents a few examples where the proposed model fails to give the correct deduction.

Table 14 Efficiency of the proposed model in different cases

Efficiency	Best case	Average case	Worst case
Time Complexity	$\Omega(m+n)$	$\theta(mn)$	$\mathrm{O}(mn)$
Space Complexity	$\Omega(m+n)$	$\theta(mn)$	$\mathrm{O}(mn)$

Table 15 Examples where the proposed model fails

Reviews	Actual	Predicted
("I do not dislike noodles") (phrases with negation)	Positive	Negative
("Thnk u 4 the treat in @ Phonenix Palladium") (special characters and slang terms)	Neutral	Negative
("someone who works as a pizza man does not like pizza?") (Irony)	Neutral	Positive

8 Conclusion

In this article, we propose an English language sentiment classification system. The system's efficacy was tested on the datasets of movies, restaurants, and electronic product reviews. The proposed system deploys SWN lexical resources. It follows effective customer feedback and a rating system to extract all relevant performance. SWN is used to determine the sentiment scores of written feedback. The system calculates the ranking score for each review using the SAW method.

Additionally, the algorithm creates a sentiment tag for each review based on the game theory concept of Nash equilibrium. Observations indicate that the system's accuracy is 91%. Another task consists of aspect ranking using the PSI method so that the customer decides according to his priority. The outcomes of the proposed methodology are contrasted with those of its baseline frameworks, such as lexicon-based and ML. By addressing the issues mentioned in the discussion section, the estimated performance of the proposed system will be enhanced in the future. In addition, we will attempt to improve the accuracy of the proposed method for extending sentiment lexicons by employing a WordNet-based approach. In addition to providing encouraging results, the proposed system could serve as a springboard for other emerging applications, such as the sentiment analysis of product reviews and the use of social media for effective policing, government accountability and participation by the public in decision-making, empowerment of women and the prevention of riots and other forms of crime. Future works to improve the proposed system's performance, such as emoticon and slang lexicon with proper sentiment scoring, need to be investigated. We can increase the criteria and alternatives by including other polarity classifications (strong positive, positive, weak, positive, neutral strong negative, negative, weak negative) in future work to increase the efficiency of the proposed work.

Acknowledgements Not Applicable.

Author contributions All authors contributed equally in this manuscript.

Funding No funding was received to assist with the preparation of this manuscript.

Data availability Zomato reviews: https://www.zomato.com/ncr/33-food-malviya-nagar-new-delhi/reviews Swiggy reviews: https://www.kaggle.com/code/residentmario/exploring-tripadvisor-uk-restaurant-reviews/notebook. Yelp reviews https://www.kaggle.com/datasets/omkarsabnis/yelp-reviews-dataset. TripAdvisor reviews https://www.kaggle.com/code/residentmario/exploring-tripadvisor-uk-restaurant-reviews/notebook.

Code availability The code generated during the current study is available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors state that they have no known competing financial interests or personal ties that could have appeared to affect the work reported in this study.

Consent to participate Not Applicable.

Human and animal ethics No humans or animals were harmed in any way.

Consent for publication Not applicable.

Credit authorship contribution statement All authors contributed equally to this study.

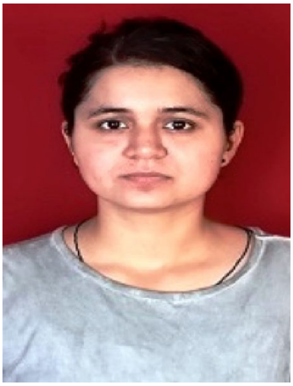
References

1. Athanasiou V, Maragoudakis M (2017) A novel, gradient boosting framework for sentiment analysis in languages where NLP resources are not plentiful: a case study for modern greek. Algorithms 10:34. https://doi.org/10.3390/a10010034
2. Berka P (2020) Sentiment analysis using rule-based and case-based reasoning. J Intell Inform Syst 55:51–66. https://doi.org/10.1007/S10844-019-00591-8/TABLES/1
3. Zhou T, Law KMY (2022) Semantic relatedness enhanced graph network for aspect category sentiment analysis. Expert Syst Appl 195:116560. https://doi.org/10.1016/J.ESWA.2022.116560
4. Zhang S, Ly L, Mach N, Amaya C (2022) Topic modeling and sentiment analysis of yelp restaurant reviews. Int J Inform Syst Serv Sect 14:1–16. https://doi.org/10.4018/ijisss.295872
5. Fikri M, Sarno R (2019) A comparative study of sentiment analysis using SVM and SentiWordNet. Indones J Electr Eng Comput Sci 13:902–909. https://doi.org/10.11591/IJEECS.V13.I3.PP902-909
6. Sangkaew N, Zhu H (2022) Understanding tourists' experiences at local markets in phuket: an analysis of tripadvisor reviews. J Qual Assur Hosp Tour 23:89–114. https://doi.org/10.1080/1528008X.2020.1848747
7. Huang F, Yuan C, Bi Y et al (2022) Multi-granular document-level sentiment topic analysis for online reviews. Appl Intell 52:7723–7733. https://doi.org/10.1007/S10489-021-02817-1/TABLES/6
8. Mohammad SM, Zhu X, Kiritchenko S, Martin J (2015) Sentiment, emotion, purpose, and style in electoral tweets. Inf Process Manag 51:480–499. https://doi.org/10.1016/J.IPM.2014.09.003
9. Giatsoglou M, Vozalis MG, Diamantaras K et al (2017) Sentiment analysis leveraging emotions and word embeddings. Expert Syst Appl 69:214–224. https://doi.org/10.1016/J.ESWA.2016.10.043
10. Bravo-Marquez F, Mendoza M, Poblete B (2014) Meta-level sentiment models for big social data analysis. Knowl-Based Syst 69:86–99. https://doi.org/10.1016/J.KNOSYS.2014.05.016
11. Bollegala D, Weir D, Carroll J (2013) Cross-domain sentiment classification using a sentiment sensitive thesaurus. IEEE Trans Knowl Data Eng 25:1719–1731. https://doi.org/10.1109/TKDE.2012.103
12. Liu M, Zhou F, Chen K, Zhao Y (2021) Co-attention networks based on aspect and context for aspect-level sentiment analysis. Knowl-Based Syst 217:106810. https://doi.org/10.1016/J.KNOSYS.2021.106810
13. Chen F, Xia J, Gao H et al (2021) TRG-DAtt: the target relational graph and double attention network based sentiment analysis and prediction for supporting decision making. ACM Trans Manag Inform Syst (TMIS) 13:1–25. https://doi.org/10.1145/3462442
14. Žunić A, Corcoran P, Spasić I (2021) Aspect-based sentiment analysis with graph convolution over syntactic dependencies. Artif Intell Med 119:102138. https://doi.org/10.1016/J.ARTMED.2021.102138
15. Lu Q, Zhu Z, Zhang G et al (2021) Aspect-gated graph convolutional networks for aspect-based sentiment analysis. Appl Intell 51:4408–4419. https://doi.org/10.1007/S10489-020-02095-3/FIGURES/5
16. Donadi M (2018) A system for sentiment analysis of online-media with tensorflow. 1–44

17. Lin C, He Y, Everson R, Rüger S (2012) Weakly supervised joint sentiment-topic detection from text. IEEE Trans Knowl Data Eng 24:1134–1145. https://doi.org/10.1109/TKDE.2011.48
18. Kim S, Zhang J, Chen Z, et al (2013) A hierarchical aspect-sentiment model for online reviews. In: Proceedings of the 27th AAAI Conference on Artificial Intelligence, AAAI 2013 526–533. https://doi.org/10.1609/aaai.v27i1.8700
19. Xu X, Cheng X, Tan S et al (2013) Aspect-level opinion mining of online customer reviews. China Commun 10:25–41. https://doi.org/10.1109/CC.2013.6488828
20. García-Pablos A, Cuadros M, Rigau G (2017) W2VLDA: almost unsupervised system for aspect based sentiment analysis. Expert Syst Appl 91:127–137. https://doi.org/10.1016/j.eswa.2017.08.049
21. Bu Z, Li H, Cao J et al (2016) Game theory based emotional evolution analysis for Chinese online reviews. Knowl-Based Syst 103:60–72. https://doi.org/10.1016/j.knosys.2016.03.026
22. Tripodi R, Linguistics MP-C (2017) Undefined A game-theoretic approach to word sense disambiguation. direct.mit.edu
23. Jain G, Lobiyal DK (2022) Word sense disambiguation using cooperative game theory and fuzzy hindi wordnet based on ConceptNet. Trans Asian Low-Resour Languag Inform Proce 21:1–25. https://doi.org/10.1145/3502739
24. Ahmad A, Ahmad T (2019) A Game Theory Approach for Multi-document Summarization. Arab J Sci Eng 44:3655–3667. https://doi.org/10.1007/S13369-018-3619-Y
25. Hossain N, Bhuiyan MR, Tumpa ZN, Hossain SA (2020) Sentiment analysis of restaurant reviews using combined CNN-LSTM. In: 2020 11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020. https://doi.org/10.1109/ICCCNT49239.2020.9225328
26. Basiri ME, Nemati S, Abdar M et al (2021) ABCDM: an attention-based bidirectional CNN-RNN deep model for sentiment analysis. Futur Gener Comput Syst 115:279–294. https://doi.org/10.1016/J.FUTURE.2020.08.005
27. Tripathy A, Anand A, Rath SK (2017) Document-level sentiment classification using hybrid machine learning approach. Knowl Inf Syst 53:805–831. https://doi.org/10.1007/S10115-017-1055-Z/FIGURES/5
28. Feng S, Wang D, Yu G et al (2010) Extracting common emotions from blogs based on fine-grained sentiment clustering. Knowl Inform Syst 27:281–302. https://doi.org/10.1007/S10115-010-0325-9
29. Saxena A, Mangal M, Jain G (2021) KeyGames: a game theoretic approach to automatic keyphrase extraction. 2037–2048. https://doi.org/10.18653/v1/2020.coling-main.184
30. Jain M, Suvarna A, Jain A (2021) An evolutionary game theory based approach for query expansion. Multimed Tools Appl. https://doi.org/10.1007/S11042-021-11297-X
31. Barfar A (2022) A linguistic/game-theoretic approach to detection/explanation of propaganda. Expert Syst with Appl 189:116069. https://doi.org/10.1016/J.ESWA.2021.116069
32. Punetha N, Jain G (2023) Bayesian game model based unsupervised sentiment analysis of product reviews. Expert Syst Appl 214:119128. https://doi.org/10.1016/J.ESWA.2022.119128
33. Mardani A, Jusoh A, Zavadskas EK et al (2016) Proposing a new hierarchical framework for the evaluation of quality management practices: a new combined fuzzy hybrid MCDM approach. Taylor Francis 17:1–16. https://doi.org/10.3846/16111699.2015.1061589
34. Afshari A, Mojahed M, Yusuff R (2010) Simple additive weighting approach to personnel selection problem. Int J Innov Manage Technol 1:511–515
35. Esuli A, Sebastiani F (2006) SENTIWORDNET: A publicly available lexical resource for opinion mining. In: Proceedings of the 5th International Conference on Language Resources and Evaluation, LREC 2006 417–422
36. Jagdale RS, Deshmukh SS (2020) Sentiment Classification on Twitter and Zomato Dataset Using Supervised Learning Algorithms. In: Proceedings of the 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing, ICSIDEMPC 2020 330–334. https://doi.org/10.1109/ICSIDEMPC49020.2020.9299582
37. Anas SM, Kumari S (2021) Opinion mining based fake product review monitoring and removal system. In:Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021 985–988. https://doi.org/10.1109/ICICT50816.2021.9358716
38. Ren X, Sun S, Yuan R (2021) A study on selection strategies for battery electric vehicles based on sentiments, analysis, and the MCDM model. Math Probl Eng. https://doi.org/10.1155/2021/9984343
39. Al Omari M, Al-Hajj M, Hammami N, Sabra A (2019) Sentiment classifier: logistic regression for arabic services' reviews in lebanon. In: 2019 International Conference on Computer and Information Sciences, ICCIS 2019. https://doi.org/10.1109/ICCISci.2019.8716394
40. Win MN, Ravana SDR, Shuib L (2022) Sentiment attribution analysis with hierarchical classification and automatic aspect categorization on online user reviews. Malays J Comput Sci 35:89–110. https://doi.org/10.22452/MJCS.VOL35NO2.1

41. Khotimah DAK, Sarno R (2018) Sentiment detection of comment titles in booking.com using probabilistic latent semantic analysis
42. Billyan B, Sarno R, Sungkono KR, Tangkawarow IRHT (2019) Fuzzy k-nearest neighbor for restaurants business sentiment analysis on tripadvisor. In: 2019 International Conference on Information and Communications Technology, ICOIACT 2019 543–548. https://doi.org/10.1109/ICOIACT46704.2019.8938564
43. Laksono RA, Sungkono KR, Sarno R, Wahyuni CS (2019) Sentiment analysis of restaurant customer reviews on tripadvisor using naïve bayes. In: Proceedings of 2019 International Conference on Information and Communication Technology and Systems, ICTS 2019 49–54. https://doi.org/10.1109/ICTS.2019.8850982
44. Yu SM, Wang J, Wang JQ (2017) An interval type-2 fuzzy likelihood-based MABAC approach and its application in selecting hotels on a tourism website. Int J Fuzzy Syst 19:47–61. https://doi.org/10.1007/S40815-016-0217-6/TABLES/7
45. Vyas V, Uma V, Ravi K (2020) Aspect-based approach to measure performance of financial services using voice of customer. J King Saud Univ Comput Inform Sci. https://doi.org/10.1016/j.jksuci.2019.12.009
46. Biaou BOS, Oluwatope AO, Odukoya HO et al (2020) Ayo game approach to mitigate free riding in peer-to-peer networks. J King Saud Univ Comput Inform Sci. https://doi.org/10.1016/j.jksuci.2020.09.015
47. Vincent TL, Brown JS (2005) Evolutionary game theory, natural selection, and darwinian dynamics
48. Seydel J (2006) Data envelopment analysis for decision support. Ind Manag Data Syst 106:81–95. https://doi.org/10.1108/02635570610641004
49. Madani K, Lund JR (2012) California's sacramento-san joaquin delta conflict: from cooperation to chicken. J Water Resour Plan Manag 138:90–99. https://doi.org/10.1061/(asce)wr.1943-5452.0000164
50. Maniya K, Bhatt MG (2010) A selection of material using a novel type decision-making method: preference selection index method. Mater Des 31:1785–1789. https://doi.org/10.1016/J.MATDES.2009.11.020
51. Singh T, Patnaik A, Gangil B, Chauhan R (2015) Optimization of tribo-performance of brake friction materials: effect of nano filler. Wear 324–325:10–16. https://doi.org/10.1016/J.WEAR.2014.11.020
52. Rasiulis R, Ustinovichius L, Vilutiene T, Popov V (2016) Decision model for selection of modernization measures: public building case. J Civ Eng Manag 22:124–133. https://doi.org/10.3846/13923730.2015.1117018
53. Gojali S, Khodra ML (2016) Aspect based sentiment analysis for review rating prediction; Aspect based sentiment analysis for review rating prediction
54. Afzaal M, Usman M, Fong ACM et al (2016) Fuzzy aspect based opinion classification system for mining tourist reviews. Adv Fuzzy Syst 2016. https://doi.org/10.1155/2016/6965725
55. Zuheros C, Martínez-Cámara E, Herrera-Viedma E, Herrera F (2021) Sentiment analysis based multi-person multi-criteria decision making methodology using natural language processing and deep learning for smarter decision aid. case study of restaurant choice using tripadvisor reviews. Inform Fusion 68:22–36. https://doi.org/10.1016/J.INFFUS.2020.10.019
56. Hemalatha S, Ramathmika R (2019) Sentiment analysis of yelp reviews by machine learning. In: 2019 International Conference on Intelligent Computing and Control Systems, ICCS 2019 700–704. https://doi.org/10.1109/ICCS45141.2019.9065812
57. Govindarajan M (2014) Sentiment Analysis Of Restaurant Reviews Using Hybrid Classification Method. Chennai India ISBN: 978–93
58. Nasim Z, Haider S (2017) ABSA toolkit: an open source tool for aspect based sentiment analysis. International Journal on Artificial Intelligence Tools. https://doi.org/10.1142/S0218213017500233
59. Luo Y, Xu X (2019) Predicting the helpfulness of online restaurant reviews using different machine learning algorithms: a case study of yelp. Sustainability 11:5254. https://doi.org/10.3390/SU11195254
60. Jo Y, Oh A (2011) Aspect and sentiment unification model for online review analysis. In: Proceedings of the 4th ACM International Conference on Web Search and Data Mining, WSDM 2011 815–824. https://doi.org/10.1145/1935826.1935932
61. Mei Q, Ling X, Wondra M, et al (2007) Topic sentiment mixture: modeling facets and opinions in weblogs. In: 16th International World Wide Web Conference, WWW2007 171–180. https://doi.org/10.1145/1242572.1242596

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



Neha Punetha has been working as a research scholar under the guidance of Dr. Goonjan Jain in the Department of Applied Mathematics of Delhi Technological University (DTU) since 2020. She received her M.Sc. degree with a major in Mathematics and minor in computer science from G.B Pant University of Agriculture and Technology (GBPUAT) and B.Sc. from D.S.B Campus K.U.

Goonjan Jain has been an assistant professor in the Department of Applied Mathematics of Delhi Technological University (DTU) since 2017. She has more than 4 years of teaching and administrative experience. Before joining academia, she worked in Infosys as a Systems Engineer from 2009–2012. She received a Ph.D. degree in Natural Language Processing (2015–2020) and M. Tech degree in Computer Science and Technology (2013–2015) from Jawaharlal Nehru University (JNU), Delhi. She was awarded Junior Research Fellowship by UGC (2015) and CSIR (2013). She completed her B.E. from Vaish College of Engineering, Rohtak, Haryana (2004–2008). Her research interests include Natural Language Processing, Artificial Intelligence, Graph Theory, and Game Theory. She has published many research papers in reputed international journals like Natural Language Engineering, ACM Transactions on Asian and Low-Resource Language Information Processing, and proceedings of international conferences like COLING (2020). She is a lifetime member of the Computer Society of India (CSI), Indian Society of Technical Education (ISTE).

Environmental Science and Pollution Research
https://doi.org/10.1007/s11356-023-26204-0

RESEARCH ARTICLE

Assessment of groundwater quality and human health risks of nitrate and fluoride contamination in a rapidly urbanizing region of India

Riki Sarma[1] · Santosh Kumar Singh[1]

Received: 18 August 2022 / Accepted: 25 February 2023

Abstract

Groundwater contamination studies are important to understand the risks to public health. In this study, groundwater quality, major ion chemistry, sources of contaminants, and related health risks were evaluated for North-West Delhi, India, a region with a rapidly growing urban population. Groundwater samples collected from the study area were analysed for physicochemical parameters — pH, electrical conductivity, total dissolved solids, total hardness, total alkalinity, carbonate, bicarbonate, chloride, nitrate, sulphate, fluoride, phosphate, calcium, magnesium, sodium and potassium. Investigation of hydrochemical facies revealed that bicarbonate was the dominant anion while magnesium was the dominant cation. Multivariate analysis using principal component analysis and Pearson correlation matrix indicated that major ion chemistry in the aquifer under study is primarily due to mineral dissolution, rock-water interactions and anthropogenic factors. Water quality index values showed that only 20% of the samples were acceptable for drinking. Due to high salinity, 54% of the samples were unfit for irrigation purposes. Nitrate and fluoride concentrations ranged from 0.24 to 380.19 mg/l and 0.05 to 7.90 mg/l, respectively due to fertilizer use, wastewater infiltration and geogenic processes. The health risks from high levels of nitrate and fluoride were calculated for males, females, and children. It was found that health risk from nitrate is more than fluoride in the study region. However, the spatial extent of risk from fluoride is more indicating that more people suffer from fluoride pollution in the study area. The total hazard index for children was found to be more than adults. Continuous monitoring of groundwater and application of remedial measures are recommended to improve the water quality and public health in the region.

Keywords Groundwater quality · Hydrogeochemistry · Health risk assessment · Groundwater nitrate · Fluoride contamination · Multivariate analyses

Introduction

Groundwater, as an easily accessible resource, not only meets the domestic water needs of people but also supports agricultural and industrial activities (Jiang et al. 2022). This dependence on groundwater is expected to increase in the future owing to the water demands of a rapidly rising global population (Xiao et al. 2022a). Excessive abstraction of groundwater exceeding the natural recharge inevitably leads to declining groundwater levels, seawater intrusion, land subsidence, and pollution (Wilopo et al. 2021; Orhan 2021). Groundwater pollution is also caused by contaminants released by anthropogenic activities that percolate into subsurface aquifers (Goyal et al. 2021; Motlagh et al. 2020). Assessment of groundwater quality and human health risks related to groundwater pollutants are thus essential research themes for scientists and scholars worldwide (Varol et al. 2021; Snousy et al. 2022).

Major contaminants detected in groundwater are nitrate, fluoride, toxic metals, pesticides, pharmaceuticals, hydrocarbons, and radioactive substances (Bedi et al. 2020; Li et al. 2021a). Researchers around the globe have tested groundwater samples for the presence of such contaminants and evaluated their suitability for drinking and irrigation purposes using geospatial tools, multivariate statistics, and index-based approaches (Sarma and Singh 2021). For example, Rahman et al. (2018) assessed the groundwater quality of Gopalganj district in Bangladesh and reported that most hydrochemical parameters exceeded the limits for

Responsible Editor: Xianliang Yi

Santosh Kumar Singh
sksinghdce@gmail.com

[1] Department of Environmental Engineering, Delhi Technological University, Delhi, India

Published online: 09 March 2023

drinking water standards. For the Guanzhong Basin of northwestern China, Li et al. (2021b) analysed 25 groundwater samples and reported that residents of their study area are at risk from high fluoride levels. Ram et al. (2021) applied water quality index and GIS methods to samples from Uttar Pradesh, India, to categorize the groundwater quality as excellent, good, poor, and unsuitable. For Malda district in Eastern India, water quality assessment showed that 14% of the groundwater samples fell in poor category (Sarkar et al. 2022).

The occurrence of high levels of nitrate and fluoride in groundwater has been reported in many studies (Rezaei et al. 2017; Rufino et al. 2019; Makubalo and Diamond 2020; He et al. 2021; Liu et al. 2022). Nitrate enters the groundwater system from excessive fertilizer use, agricultural runoff, sewage and septic tank leaks, manure systems and animal wastes (Duvva et al. 2022; Dhakate et al. 2023). Consumption of groundwater with nitrate levels greater than 45 mg/L can lead to methemoglobinemia. Also known as blue baby syndrome, this condition reduces the ability of blood to transport oxygen, causing breathlessness, cardiac arrest or death, especially in infants (Ceballos et al. 2021; Golaki et al. 2022; Gugulothu et al. 2022; Panneerselvam et al. 2022). Fluoride naturally occurs in groundwater from geogenic sources and weathering of fluoride-bearing minerals (Subba Rao 2017; Mukherjee and Singh 2018). Surplus application of phosphate fertilizers further enhances the fluoride pollution of groundwater (Karunanidhi et al. 2020; Subba Rao et al. 2021). The intake of fluoride within the permissible limits for drinking water prevents tooth decay and dental cavities and helps in bone formation (Adimalla and Venkatayogi 2017; Sathe et al. 2021). But the long-term intake of excessive fluoride (≥ 1.5 mg/l) may lead to neurological effects, and dental and skeletal fluorosis (Mukherjee and Singh 2018; Ambastha and Haritash 2021; Liu et al. 2022).

In India, high levels of nitrate and fluoride have been reported in many regions — Maharashtra (Nawale et al. 2021), Telangana (Adimalla and Li 2019; Subba Rao et al. 2021; Duvva et al. 2022), Punjab (Singh et al. 2020), Haryana (Kaur et al. 2020; Rishi et al. 2020), Assam (Sathe et al. 2021), Rajasthan (Rahman et al. 2021; Jandu et al. 2021), Tamil Nadu (Karunanidhi et al. 2020; Khan et al. 2021), Jharkhand (Giri et al. 2021) and Uttar Pradesh (Maurya et al. 2020). The United States Environmental Protection Agency has developed a framework to assess the health risks from the usage of fluoride and nitrate contaminated groundwater (USEPA 1989; 1997; 2004; 2014). Many researchers around the world have adopted this model in their studies to evaluate the hazard index (HI) for different categories of people — males, females, children and infants (Singh et al. 2020; Chen et al. 2021; Reddy et al. 2022). The acceptable limit of non-carcinogenic risk is when HI $\leq$ 1. If the HI is more than 1, then exposure to contaminated groundwater has serious adverse effects on health (Adimalla et al. 2019). Many studies have reported that infants and children are more at risk than adults (Gao et al. 2020; Adimalla and Qian 2021).

The North-West region of Delhi has industrial, residential and agricultural areas. In the last two decades, this region has experienced major changes in land use. Much of the agricultural lands and rural built-up area have been converted to urban areas. Identifying the mechanisms of groundwater pollution arising from the rapid urbanization in this region is important. This study was carried out in the North-West region of Delhi, India, to (a) evaluate the hydrogeochemistry of groundwater and its suitability for drinking and irrigation, (b) assess the spatial extent of fluoride and nitrate contamination, and (c) estimate the corresponding non-carcinogenic health risks for men, women and children using the USEPA method. The results of this study will be helpful in understanding how increasing urbanization influences groundwater quality and affects human health.

Materials and methods

Study area

The investigated area lies in the North-West region of the National Capital Territory (NCT) of Delhi. The NCT covers a geographical area of 1483 km^2 and falls in the Yamuna River sub-basin, which controls its drainage system. The NCT has adjoining smaller cities — Faridabad, Gurugram, Ghaziabad and Noida which contribute to a total of 3000 km^2 of urban area (Chaudhuri and Sharma 2020). The region is characterized by hot summers and cold winters. The average rainfall is 581 mm. July, August and September are the main monsoon months that receive 81% of the total rainfall. There are planned residential and industrial areas in the North-West region of NCT with some agricultural lands near the adjoining state of Haryana (CGWB n.d). Thus, this area has both urban as well as rural populations. Land use maps from Bhuvan (2021) show that agricultural regions have decreased in the last 15 years while urban areas have increased in this region.

In Delhi, the aquifer geology is complex, varying from Quartzite to Older and Younger Alluvium (CGWB 2021; Sarma and Singh 2022). North West District is characterized by unconsolidated Quaternary alluvium deposits from the Middle to Late Pleistocene Age (CGWB n.d). Sand, silt, and clay are the major soil types in the region in varying proportions. In most of the district, water levels are 5–10 m below ground level, with deeper water levels (> 15 mbgl) observed in the northern part. The district is bordered by the Yamuna River in the northeast which controls the drainage system. The total annual groundwater recharge has been estimated as 8630.7 ham and total annual ground water draft for all uses has been estimated as 9015.2 ham as on 2011 (CGWB n.d). Groundwater exploration studies by the Central Ground Water Board, India, showed that discharge in exploratory wells and piezometers ranged from 150 to 2816 lpm and drawdown ranged from 0.72 to 17.23 m (CGWB n.d). The overall stage of ground water development of the area is 112.36%. The Central Ground Water Board has classified the sub-regions of the district as semi-critical or over-exploited.

Sample collection and analysis

Groundwater samples from 58 locations in the study area were collected from handpumps and bore wells with a depth range of 15–35 mbgl in January 2021. The coordinates of the sampling locations were recorded using a portable GPS device. Location map of the study area and sampling points were prepared by GIS software ArcMap 10.7.1 (Fig. 1). The wells were pumped for 5–10 min to remove the interference from any stagnant water. The water samples were collected in distilled water rinsed polyethylene bottles of 1 l capacity. The sample bottles were sealed, labelled and stored at 4 °C. The analytical procedures for estimating the groundwater parameters were carried out according to the standard methods given by the American Public Health Association (APHA 2017).

The physical parameters — pH, electrical conductivity (EC) and total dissolved solids (TDS) were measured on site using a portable multi-parameter meter (Orion Star A320). Prior to use, the pH meter was calibrated using buffer solutions of pH 4.0, 7.0 and 10.0 and the EC meter was calibrated using standard solutions with EC = 1413 µS/cm and 12.9 mS/cm. Total hardness (TH as $CaCO_3$), total alkalinity (TA as $CaCO_3$), chloride (Cl^-), carbonate (CO_3^{2-}), bicarbonate (HCO_3^-), calcium (Ca^{2+}) and magnesium (Mg^{2+}) ions were determined by titrimetric methods. Sulphate (SO_4^{2-}), nitrate (NO_3^-), and phosphate (PO_4^{3-}) were measured using UV–Visible Spectrophotometers (Labtronics 290 and LabIndia Analytical UV 3092). Fluoride (F^-) was measured using an electrode meter. Sodium (Na^+) and potassium (K^+) ions were determined using a flame photometer (Systronics 128). The analytical test methods, their corresponding reagents and detection limits are presented in Table S1. The accuracy of the chemical analysis was validated by charge balance errors (CBE), and samples with ± 10% error were considered only (Domenico and Schwartz 1990; Adimalla et al. 2019; Rahman et al. 2021; Panneerselvam et al. 2022). Eliminating samples above this error, 52 samples were considered for further analysis (Fig. S1). The CBE was calculated as CBE $= \frac{\sum \text{cations} - \sum \text{anions}}{\sum \text{cations} + \sum \text{anions}} \times 100$.

The groundwater samples were evaluated for drinking purpose by comparing the observed value against the recommended limits given by the Bureau of Indian Standards (BIS) and World Health Organization (WHO). The spatial distribution maps of the groundwater parameters were created using the inverse-distance weighted (IDW) interpolation technique in ArcMap 10.7.1 software. The

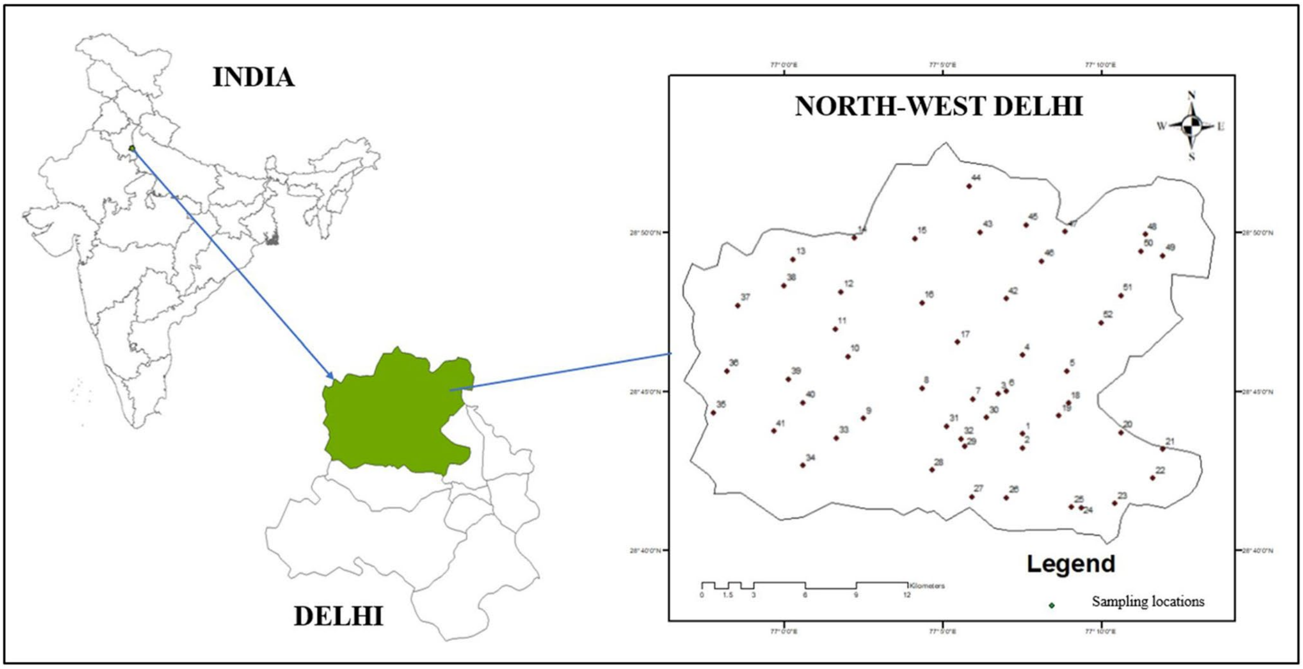


Fig. 1 Groundwater sampling locations in the study area

hydrogeochemical characteristics of the groundwater samples were studied by plotting Piper trilinear diagram (Piper 1944) in AquaChem software. Chloro-alkaline indices CAI-1 and CAI-2 (Schoeller 1965) were calculated to understand the mechanisms of ion-exchange and rock-water interactions (Subba Rao 2017). CAI-1 and CAI-2 were calculated as per the following equations (all ions in meq/l).

$$\mathrm{CAI}-1=\frac{\mathrm{Cl}^{-}-\mathrm{Na}^{+}+\mathrm{K}^{+}}{\mathrm{Cl}^{-}} \tag{1}$$

$$\mathrm{CAI}-2=\frac{\mathrm{Cl}^{-}-\mathrm{Na}^{+}+\mathrm{K}^{+}}{\mathrm{SO}_4^{2-}+\mathrm{HCO}_3^{-}+\mathrm{CO}_3^{2-}+\mathrm{NO}_3^{-}} \tag{2}$$

Water quality index values (WQI) were calculated to determine the suitability of the groundwater samples for drinking. The WQI was based on the values of pH, TDS, TH, TA, Cl^-, F^-, $\mathrm{SO^{2-}}_4$, $\mathrm{NO^-}_3$, Ca^{2+}, and Mg^{2+}. The following equation was used to calculate the WQI.

$$\mathrm{WQI}=\frac{\sum W_iQ_i}{\sum W_i} \tag{3}$$

where Q_i is the quality rating for each parameter given by $Q_i = 100 * [(V_i - V_o)/(S_i - V_o)]$, V_i is the observed value of *i*th parameter, V_o is the ideal value of parameter in pure water (0 for all parameters; 7.0 for pH), S_i is the recommended standard value of *i*th parameter and W_i is the unit weight of each parameter ($W_i = K/S_i$). For calculation of W_i, K is proportionality constant given by $K = 1/\sum(1/\mathrm{S_i})$).

In order to determine the suitability of the samples for irrigation, the parameters such as soluble sodium percentage (SSP), residual sodium carbonate (RSC), sodium absorption ratio (SAR), permeability index (PI), Kelley's ratio (KR) and magnesium hazard (MH) were calculated as per their respective formula (Aravinthasamy et al. 2021) (Table 1). The suitability of samples for irrigation was also determined using the US Salinity Laboratory classification (USSL 1954). IBM SPSS Statistics software version 26 was used for multivariate statistical techniques — principal component analysis (PCA) and Pearson correlation matrix.

Health risk assessment

The USEPA considers high nitrate and fluoride in drinking water as non-carcinogenic risks to human health (USEPA 1989). The exposure routes to such contaminated water may be either through oral ingestion (drinking) and/or dermal contact (bathing). Considering these exposure pathways, the chronic daily intake (CDI in mg/kg/day) through oral ingestion, and dermally absorbed dose (DAD in mg/kg/day) through bathing were calculated. The non-carcinogenic risk through drinking water exposure route in terms of CDI was calculated by Eq. (4).

$$\mathrm{CDI}=\frac{\mathrm{CPW}*\mathrm{IR}*\mathrm{ED}*\mathrm{EF}}{\mathrm{ABW}*\mathrm{AET}} \tag{4}$$

where CDI is the chronic daily intake (mg/kg/day), CPW is the concentration of a particular contaminant in groundwater (mg/L), IR is the human ingestion rate (L/day: 2.5 L/day for adults and 0.78 L/day for children), ED is the exposure duration (years: 64, 67 and 12 for men, women and children respectively), EF is the exposure frequency (days/years: 365 days for children and adults), ABW is the average body weight (Kg: 65, 55 and 15 for males, women and children respectively), and AET is the average time (days: 23,360, 24,455 and 4380 for males, women and children, respectively) (USEPA 2014). The health risk due to dermal exposure was calculated by using the following equation.

$$\mathrm{DAD}=\frac{\mathrm{CPW}*\mathrm{TC}*\mathrm{Ki}*\mathrm{EV}*\mathrm{SSA}*\mathrm{CF}*\mathrm{ED}*\mathrm{EF}}{\mathrm{ABW}*\mathrm{AET}} \tag{5}$$

where DAD is the dermally absorbed dose (mg/kg/day), TC indicates the contact duration (h/d: 0.4 h per day for adults and children), Ki is the dermal adsorption parameters (cm/h: 0.001 cm/h), EV is the bathing frequency (times/day: considered as 1 time in a day), SSA is the skin surface area available for contact (cm^2: 16,600 and 12,000 cm^2 for adults and children, respectively), CF is the unit conversion factors (0.001), ED is the exposure duration (years: 64, 67 and

Table 1 Equations to calculate suitability of water for irrigation

Equations	References
Soluble sodium percentage (SSP) = $[\mathrm{Na^+}/(\mathrm{Ca^{2+}}+\mathrm{Mg^{2+}}+\mathrm{Na^+}+\mathrm{K^+})] *100$	Wilcox (1955)
Residual sodium carbonate (RSC) = $(\mathrm{HCO_3}^- + \mathrm{CO_3}^{2-}) - (\mathrm{Ca^{2+}}+\mathrm{Mg^{2+}})$	Eaton (1950)
Sodium absorption ratio (SAR) = $\mathrm{Na^+}/[(\mathrm{Ca^{2+}}+\mathrm{Mg^{2+}})/2]^{1/2}$	Richards (1954)
Permeability index (PI) = $[\mathrm{Na^+}+(\mathrm{HCO_3}^-)^{1/2}/(\mathrm{Ca^{2+}}+\mathrm{Mg^{2+}}+\mathrm{Na^+})] * 100$	Doneen (1964)
Kelley's ratio (KR) = $\mathrm{Na^+}/(\mathrm{Ca^{2+}}+\mathrm{Mg^{2+}})$	Kelley (1940)
Magnesium hazard (MH) = $[\mathrm{Mg^{2+}}/(\mathrm{Ca^{2+}}+\mathrm{Mg^{2+}})] *100$	Raghunath (1987)

All ions in meq/l

12 for males, women and children, respectively), EF is the exposure frequency (days/years: 365 days for children and adults), ABW is the average body weight (Kg: 65, 55 and 15 for men, women and children respectively), and AET is the average time (days: 23,360, 24,455 and 4380 for males, women and children, respectively) (USEPA 1997; Adimalla and Qian 2021).

Oral and dermal hazard quotient for the nitrate and fluoride were computed by the following equations:

$$HQ_{\mathrm{oral}} = \frac{\mathrm{CDI}}{\mathrm{RfD}} \tag{6}$$

$$HQ_{\mathrm{dermal}} = \frac{\mathrm{DAD}}{\mathrm{RfD}} \tag{7}$$

where HQ_{oral} and HQ_{dermal} are the non-carcinogenic oral and dermal hazard quotient, respectively. CDI and DAD are chronic daily intake (mg/kg/day) and the dermally absorbed dose (mg/kg/day), respectively, and RfD represents the reference dose of a specific contaminant (USEPA 1989). The oral reference dose of nitrate is 1.6 mg/kg/day and that of fluoride is 0.06 mg/kg/day, obtained from the database of IRIS (Integrated Risk Information System) (USEPA 1989). The HQ values can be used to evaluate the health risk alone where adverse health effects are seen if HQ > 1. However, the hazard index (HI) gives the total hazard presented by exposure to multiple contaminants through multiple pathways. In this study, it is calculated as the sum of the hazard quotients calculated for oral and dermal risk exposure (HQ_{oral} and HQ_{dermal}) to nitrate and fluoride given by:

$$HI_i = HQ_{\mathrm{oral}} + HQ_{\mathrm{dermal}} \tag{8}$$

$$HI_{\mathrm{total}} = \sum_{i=1}^{n} HI_i \tag{9}$$

Based on the HI_{total} values, no significant non-carcinogenic risk occurs if $HI_{\mathrm{total}} \leq 1$. However, if $HI_{\mathrm{total}} > 1$, then there is significant non-carcinogenic risk (USEPA 1991; 2004).

Results and discussion

Groundwater chemistry

The statistics of the physicochemical parameters of the groundwater samples — minimum, maximum, mean and standard deviation are summarized in Table 2. The pH of the samples ranges between 7.5 and 8.4 with a mean value of 8.0, indicating slightly alkaline conditions. The pH values of the water samples fall within the acceptable limits set by BIS (2012). The EC values vary significantly, with a range of 254–15,440 µS/cm and a mean value of 3699 µS/cm. The

Table 2 Statistics of groundwater quality parameters ($n = 52$) and comparison with drinking water standards (BIS and WHO)

Parameters	Statistics of groundwater samples				Drinking water standards		Percentage of samples exceeding the standard		Maximum multiple of exceeding the standard	
	Minimum	Maximum	Mean	Std. Dev	BIS (2012)	WHO (2017)	BIS	WHO	BIS	WHO
pH	7.5	8.4	8.0	0.22	6.5–8.5	7.0–8.0	–	–	–	–
EC (µS/cm)	254	15,440	3699	3837.68	–	–	-	-	–	–
TDS (mg/l)	128	7770	1854	1922.43	500	600	83%	77%	15.54	12.95
TA as $CaCO_3$ (mg/l)	220	1550	880	312.79	200	–	52%	–	7.75	–
TH as $CaCO_3$ (mg/l)	180	7108	1883	1717.35	200	200	98%	98%	35.54	35.54
Cl^- (mg/l)	20	4700	704	904.81	250	250	58%	58%	18.8	18.8
CO_3^{2-} (mg/l)	0	180	62	41.57	–	–	–	–	–	–
HCO_3^- (mg/l)	268	1696	949	352.80	–	–	–	–	–	–
SO_4^{2-} (mg/l)	35	2840	443	597.78	200	250	58%	50%	14.2	11.36
NO_3^- (mg/l)	0.24	380.19	65.29	89.37	45	50	40%	40%	8.44	7.60
F^- (mg/l)	0.05	7.90	2.23	1.90	1.5	1.5	58%	58%	5.26	5.26
PO_4^{3-} (mg/l)	BDL	0.61	0.13	0.105	–	–	–	–	–	–
Na^+ (mg/l)	4	1006	296	247.47	–	–	–	–	–	–
K^+ (mg/l)	1.4	71.4	12.2	15.09	–	–	–	–	–	–
Ca^{2+} (mg/l)	20	872	170	146.62	75	100	77%	65%	11.62	8.72
Mg^{2+} (mg/l)	20	1580	356	347.15	30	–	96%	–	52.66	–

BDL below detection limit

elevated values of EC indicate high ionic strength, mineral content and dissolved solids. TDS values range from 128 to 7770 mg/l, with a mean value of 1854 mg/l. Only 17% of the samples are within the BIS acceptable limit of 500 mg/l. According to the classification of TDS given by Freeze and Cherry (1979), TDS < 1000 mg/l indicates fresh water while TDS between 1000 to 10,000 mg/l indicates brackish water. Based on this classification, 44% and 56% of the samples fall in the fresh and brackish water categories respectively. Davis and DeWiest (1966) classified groundwater as desirable for drinking if TDS < 500 mg/l, permissible for drinking if TDS is between 500 to 1000 mg/l, useful for irrigation if TDS is between 1000 to 3000 mg/l and unsuitable for drinking and irrigation if TDS > 3000 mg/l. Based on this classification, 17% of the samples were desirable for drinking, 27% were permissible for drinking, 37% were suitable for irrigation and 19% were unfit for both drinking and irrigation. The spatial distribution maps of pH, EC and TDS are given in Fig. S2.

The concentrations of the cations Ca^{2+}, Mg^{2+}, Na^{+} and K^{+} range from 20 to 872 mg/l, 20–1580, 4–1006 and 1.4–71.4 mg/l respectively with mean values of 170, 356, 296 and 12.2 mg/l respectively. The concentrations of dissolved anions such as HCO_3^{-}, Cl^{-}, PO_4^{3-} and SO_4^{2-} vary from 268 to 1696, 20 to 4700, 0.00 to 0.61 and 35 to 2840 mg/l respectively with the mean concentrations of 949, 704, 0.13 and 443 mg/l, respectively. The TH values range from 180 to 7108 mg/l as $CaCO_3$ with mean of 1883 mg/l as $CaCO_{3.}$ According to the classification for total hardness by Sawyer and McCarty (1967), water is termed “very hard” if TH > 300 mg/l as $CaCO_3$ and “hard” if TH is between 150 and 300 mg/l as $CaCO_3$. Based on this classification, 92% of the samples have “very hard” water, and 8% of the samples fall in “hard water” categories (Table S2). This is evident from the high levels of bicarbonate ions present in the samples. The standard deviation of SO_4^{2-} is higher than its mean which indicates that sulphate levels in the water samples fluctuate randomly. The dominant major cations in the groundwater samples are in the order of $Mg^{2+} > Na^{+} > Ca^{2+} > K^{+}$, while the dominant anions are $HCO_3^{-} > Cl^{-} > SO_4^{2-} > NO_3^{-} > CO_3^{2-} > F^{-}$. The elevated concentrations of HCO_3^{-} along with Mg^{2+} and Ca^{2+} ions in some samples indicate that the study area might be affected by dissolution of carbonate minerals (like calcite and dolomite) and/or silicate minerals by carbonic acid (CGWB 2016; Snousy et al. 2022). Excess Na^{+} over Cl^{-} indicates rock weathering (or cation exchange) while the vice versa indicates reverse ion exchange (Subba Rao et al. 2017; Gugulothu et al. 2022). For the studied samples, about 85% had excess Cl^{-} over Na^{+} indicating that reverse ion exchange was the primary source of these ions. High sodium intake (> 200 mg/l) leads to problems of hypertension, kidney and nerves (Rishi et al. 2020). Na^{+}, Mg^{2+} and K^{+} arise from anthropogenic sources such as wastewater, return flows from irrigation and potassium fertilizers (Subba Rao et al. 2021). The high Cl^{-} concentration may be due to the release of untreated sewage and industrial effluents in the region. Chloride imparts a salty taste to the water and may have laxative effects. The industrial activities in the study region may also be the reason for the high SO_4^{2-} levels found in the water samples. High sulphate concentrations along with high Mg^{2+} are known to cause gastro-intestinal problems (CGWB 2016).

Nitrate levels in the samples range from 0.24 to 380.19 mg/L, with a mean of 65.29 mg/L (Fig. S3(a)). According to WHO (2011), there is no health risk for humans if nitrate levels are below 45 mg/l. However, nitrate between 45 and 100 mg/L causes health effects on children and adults and > 100 mg/L have very high health risk. As per this classification of nitrate, 60% of groundwater samples fall under the “no health risk” category, while 19% and 21% of groundwater samples fall under the “high health risk” and “very high health risk” categories. The spatial distribution map of nitrate is presented in Fig. 2a. Nitrate is predominant in shallow aquifers and easily reaches the groundwater from the surface owing to its high solubility in water (Adimalla and Qian 2021). Nitrate is thus largely anthropogenic in nature and majorly sourced from agrochemicals, open land dumping, domestic, animal and manufacturing wastes (Duvva et al. 2022; Panneerselvam et al. 2022). The high levels of nitrate in the study region may be due to fertilizers such as diammonium phosphate and urea which are commonly utilized in North India. Because of the widespread use of such fertilizers, nitrate can drain away from soils and percolate into the groundwater. Rahman et al. (2021) lists landfill leachate as one of the contributors to nitrate contamination of groundwater. The Bhalaswa landfill in the study region has been operational since 1993 (Sidhu et al. 2015), and its leachate percolating into the groundwater may also contribute to high nitrate levels.

The fluoride concentration ranged from 0.05 to 7.90 mg/l with a mean of 2.23 mg/l (Fig. S3b). Fluoride concentration less than 0.6 mg/l may cause dental caries while greater than 1.5 mg/l may cause severe problems of fluorosis. The concentration of fluoride was below 0.6 mg/l in 21% of the samples and exceeded the permissible limit (1.5 mg/l) in 58% of the groundwater samples. The spatial distribution map of fluoride is presented in Fig. 2b. The high fluoride distribution is identified in northern, southern, central and western parts of the region. Fluoride-rich minerals and usage of phosphate fertilizers are the chief sources of elevated fluoride levels. The anionic exchange controlling the fluoride content in the study region is enhanced by the alkaline nature of water (Duvva et al. 2022; Xiao et al. 2022b). Several studies have reported high concentrations of NO_3^{-} and F^{-} in north-west Delhi and the neighbouring state of Haryana

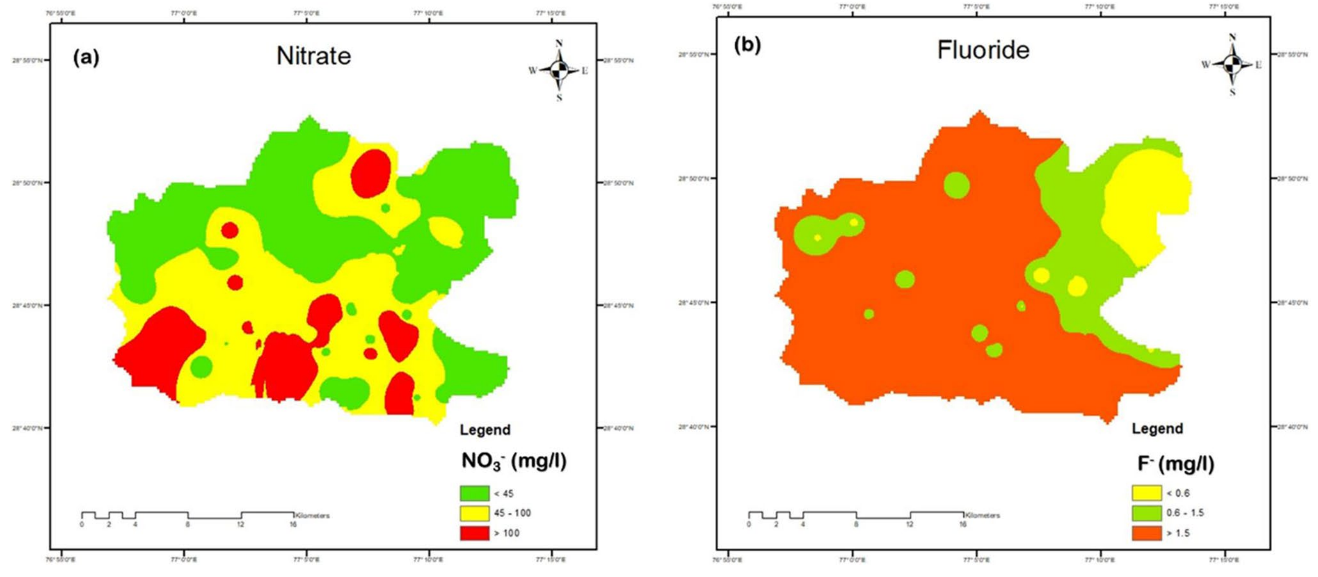


Fig. 2 Spatial distribution maps of **a** nitrate and **b** fluoride in the study area

(Singh et al. 2017; Kaur et al. 2020; Ambastha and Haritash 2021; Masood et al. 2022).

Hydrochemical facies

The Piper trilinear diagram (Piper 1944) suggests the dominance of groundwater chemistry. For the collected groundwater samples, Piper diagram was plotted using AquaChem software, version 10 (Fig. 3). In the cation plot, maximum samples fall in magnesium type (67%) while in the anion plot, maximum samples fall in the bicarbonate type (56%). Mg^{2+} is the dominant ion as a result of weathering of silicate rocks (Adimalla 2019). In the diamond shape, maximum samples (48%) fall in $CaHCO_3$ type followed by mixed CaMgCl type (30%). $CaHCO_3$ type water indicates that River Yamuna and irrigation canals are primarily responsible for the aquifer recharge in the absence of adequate rainfall. The Piper classification indicates that major processes regulating groundwater chemistry in the study region are ion exchanges, rock-water interactions, mineral weathering and anthropogenic influences (Snousy et al. 2022; Panneerselvam et al. 2022).

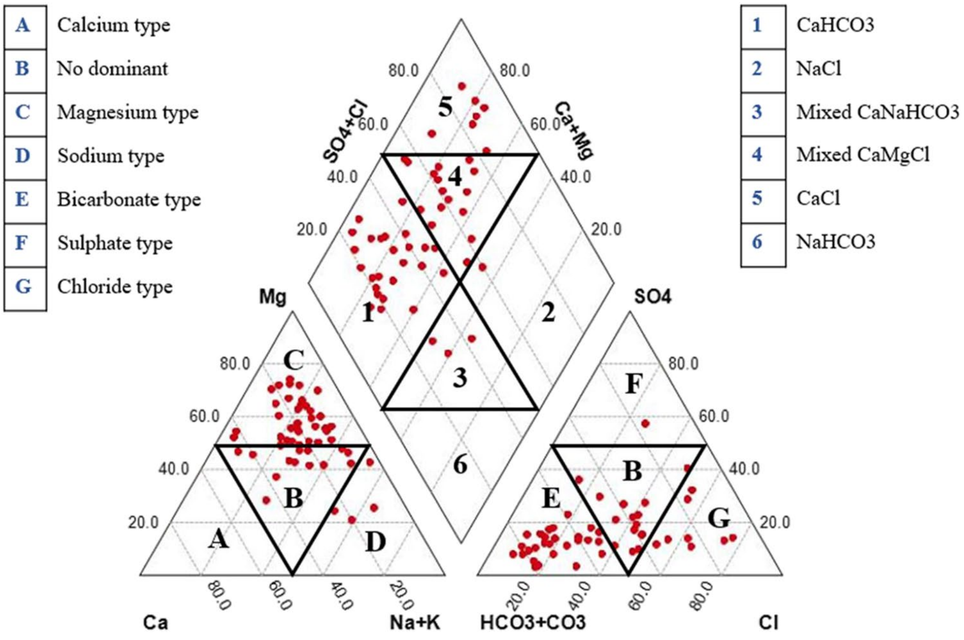


Fig. 3 Piper trilinear classification of groundwater samples

The chloro-alkaline indices CAI-1 and CAI-2 help in understanding the mechanism of ion exchange. If the index is positive, it implies an exchange of sodium and potassium ions from the water with calcium and magnesium ions of the rocks (base–exchange reaction). If it is negative, it indicates vice versa, i.e., calcium and magnesium of water exchanging with sodium and potassium from rocks (cation–anion exchange reaction) (Subba Rao 2017). For the studied groundwater samples, most of the samples demonstrated positive CAI values (Fig. 4a) indicating the cation–anion exchange reaction where Na^+ and K^+ from the water continuously exchanges with Ca^{2+} and Mg^{2+} from aquifer materials due to rock-water interactions (Rashid et al. 2022). Moreover, plot of $(Na^+ + K^+)–Cl^-$ against $(Ca^{2+} + Mg^{2+})–(HCO_3{}^- + SO_4{}^{2-})$ can be expressed as $y = -1.1249x + 5.5638$ with a correlation coefficient of 0.9226 (Fig. 4b). The negative slope of − 1.1249 confirms that the relationship between Na^+, K^+, Ca^{2+} and Mg^{2+} is influenced by reverse ion exchange process (Kumar and James 2016):

$$2Na^+ + Ca^{2+}(Mg)–Clay \rightarrow Na^+–Clay + Ca^{2+} + Mg^{2+}$$

Multivariate statistical analysis

Principal component analysis

The application of PCA was first done by checking the Bartlett's test of sphericity and Kaiser–Meyer–Olkin (KMO) sampling adequacy. PCA requires KMO sampling adequacy to be > 0.50 for the dataset (Snousy et al. 2022). The Bartlett test of sphericity was in accordance with p value < 0.0001, and KMO sampling adequacy was 0.671 for the groundwater samples. These values confirm that the dataset is suitable for PCA. The PCA was performed in SPSS software using varimax rotation method with Kaiser normalization. Factors loading values are classified as weak (0.30–0.50), moderate (0.50–0.75), and strong (> 0.75) (Wu et al. 2020). Table 3 reveals that five significant components are calculated (with eigenvalues > 1), which represent 83% of the total variance. The eigenvalues represent how much variance there is in the dataset, and the variance represents the amount of variation in the dataset that can be attributed to each principal component. Component 1 explains 46.6% of the total variance and has positive loading of EC, TDS, TH, Cl^-, Mg^{2+}, Ca^{2+}, Na^+ and $SO_4{}^{2-}$ implying that EC and TDS are primary governed by the major cations and anions through mineral dissolution, rock-water interaction, ion-exchange and anthropogenic

Table 3 The main five principal components extracted form groundwater samples

	Component				
	1	2	3	4	5
TDS	0.966				
EC	0.966				
TH	0.944				
Mg	0.929				
Cl	0.916				
Na	0.859				
SO4	0.855				0.331
Ca	0.816			− 0.323	
TA		0.975			
HCO3		0.923			
CO3		0.551		0.361	− 0.458
PO4			0.839		
K			0.764		
F				0.817	
pH	− 0.366			0.551	
NO_3					0.878
Eigenvalue	7.459	2.160	1.383	1.197	1.085
% of variance	46.617	13.497	8.642	7.484	6.783

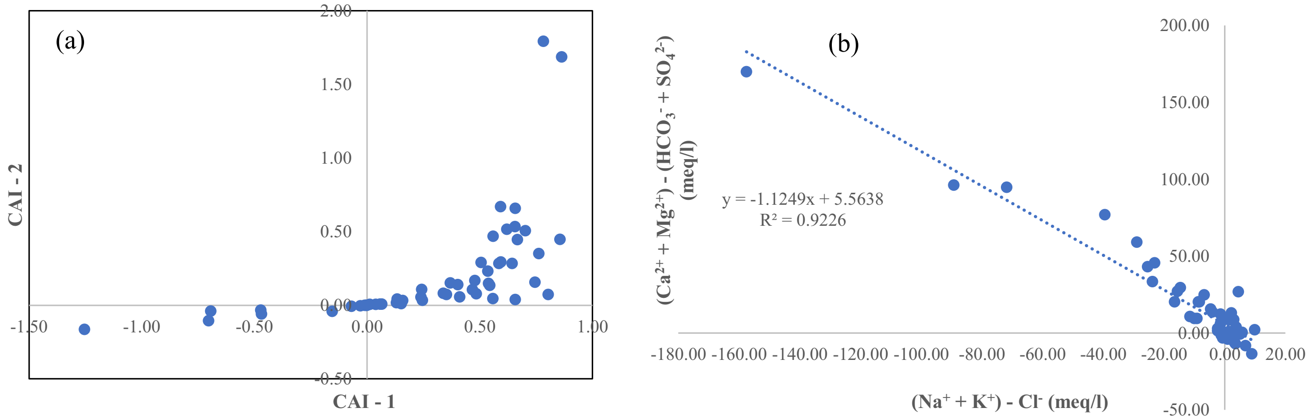


Fig. 4 Plot of **a** CAI-1 against CAI-2 and **b** $(Na^+ + K^+)–Cl^-$ against $(Ca^{2+} + Mg^{2+})–(HCO_3{}^- + SO_4{}^{2}.^-)$

factors (Elemile et al. 2021). Component 2 explains 13.5% of the total variance and has positive loading of carbonate, bicarbonate and total alkalinity. This implies that TA is driven by dissolution of carbonate and bicarbonate minerals in the study area. Component 3 explains 8.6% of the total variance and has positive loading of K^+ and PO_4^{3-} indicating the use of potash and phosphate fertilizers. Component 4 explains 7.5% of the total variance and has positive loading of F^-, with moderate loading of CO_3^{2-} and pH and negative loading of Ca. This implies that the concentration of fluoride is due to weathering of fluorite minerals (CaF_2) enhanced by carbonate weathering and alkaline conditions (Barzegar et al. 2017; Xiao et al. 2022b). Finally, component 5 explains 6.8% of the total variance and has strong positive loading of nitrate indicating that origins of nitrate in the water samples may be purely anthropogenic — fertilizer use, sewage and animal wastes. Figure S4 represents the PCA plot of the components in rotated space.

Pearson correlation matrix analysis

The relationships between the physicochemical parameters were analysed by PCMA. The correlation matrix is presented in Table 4. pH has negative correlation with EC, TDS, TH, Cl^-, SO_4^{2-}, Ca^{2+}, Mg^{2+}, and Na^+, consistent with studies by Swain et al. (2022) and Panneerselvam et al. (2021). EC shows identical liner correlation with TDS ($r=1.000$) with a 99% confidence level and significant positive correlation with Na^+ ($r=0.825$), Ca^{2+} ($r=0.835$), Mg^{2+} ($r=0.923$), Cl^- ($r=0.935$) and SO_4^{2-} ($r=0.769$). This is consistent with the results of PCA. The TDS has a strong positive correlation with Na^+ ($r=0.825$) and Cl^- ($r=0.939$) indicating that rock weathering and sewage seepage have caused the salinity to increase. Ca^{2+} shows significant positive correlation with Mg^{2+} ($r=0.756$), Cl^- ($r=0.828$) and SO_4^{2-} ($r=0.693$). Mg^{2+} also shows significant positive correlation with Cl^- ($r=0.867$) and SO_4^{2-} ($r=0.823$). These correlations indicate that major ion chemistry in the groundwater samples is influenced by the dissolution of aquifer materials, rock-water interactions and domestic wastewater infiltration (Snousy et al. 2022). NO_3^- shows negative correlation with pH which is also reported in Stylianoudaki et al. (2022) and Glass and Silverstein (1998).

Water quality index for drinking

Based on the classification given in the study by Masood et al. (2022), the WQI obtained for the groundwater samples were evaluated (Table S3). The WQI < 50 is beneficial for health ("excellent" category) which is calculated for 12% of the samples, located in some isolated pockets in the study region. WQI between 50 and 100 is acceptable for drinking use ("good" category) which is calculated for 8% of the samples. Forty percent of the samples were impure with WQI 100–200 ("poor" category), and 25% of the samples needed treatment prior to use ("very poor" category) with WQI 200–300. The WQI > 300 were found in 15% of the samples which were completely unsuitable for drinking. The spatial distribution map of WQI is presented in Fig. 5. Poor, very poor and unsuitable water quality can be observed in most parts of the study region — central, northern, western, eastern and southern. Only a small area in the north eastern region has good water quality.

Irrigation water quality

Agricultural areas in the study region are situated in the extreme northwest and western regions, where groundwater is the primary source of irrigation. Evaluating the suitability of groundwater for irrigation purposes was done by comparing the irrigation quality parameters with the recommended values (Table 5). The quality of water for irrigation is dependent on its mineral constituents which affect both plants and soil (Wilcox 1955; Alam et al. 2012). The EC is an indicator of the salinity of the groundwater which can influence crop growth. High levels of salinity can negatively affect crop development (Subba Rao 2017; Gugulothu et al. 2022). The salinity is low if EC < 250 μS/cm and very high if EC > 2250 μS/cm. For the study region, 54% of the samples have very high salinity. The sodium absorption ratio (SAR) values indicate the cation–exchange reaction in the soil. High values of SAR specify a situation where the absorbed calcium and magnesium have been replaced by sodium, posing a risk to soil structure (Saha et al. 2019). All the studied samples present a low sodic hazard in terms of the SAR (SAR values < 10). The USSL classification (USSL 1954) plots EC values against the SAR values (Fig. 6). The USSL diagram shows that majority of the samples fall in S1C2, S1C3, S1C4 and S2C4 classes, indicating low to medium sodium hazard and medium to very high salinity hazard in the study region.

The residual sodium carbonate (RSC) is an indicator of the hazardous effects of carbonate and bicarbonate ions for irrigation purposes (Saha et al. 2019; Rishi et al. 2020). RSC values < 1.25 meq/l are fit for irrigation while RSC > 2.5 meq/l are unsuitable. Based on this classification, 83% of the samples are suitable for irrigation while only 15% are unfit. The soluble sodium percentage (SSP) indicates the sodium content in terms of %Na. The sodium-laden water reacts with soil and accumulates in the air spaces (or voids) in the soil. This leads to clogging of the soil particles and reduction in soil permeability which can affect the growth of plants (Todd 1980). The permissible limit of SSP is 60% for irrigation water. Based on this classification, 98% of the samples were permissible for irrigation. Kelley's ratio (KR) measures

Table 4 Pearson correlation coefficient matrix among physicochemical parameters in groundwater samples

	pH	EC	TDS	Cl	CO3	HCO3	SO4	NO3	PO4	F	Ca	Mg	Na	K	TH	TA
pH	1	−0.384**	−0.385**	−0.403**	0.078	−0.215	−0.296*	−0.111	−0.067	0.204	−0.458**	−0.364**	−0.310*	−0.132	−0.400**	−0.182
EC		1	1.000**	0.935**	−0.003	0.203	0.769**	0.169	0.295*	−0.149	0.835**	0.923**	0.825**	0.263	0.943**	0.187
TDS			1	0.939**	−0.001	0.202	0.765**	0.160	0.295*	−0.150	0.834**	0.923**	0.825**	0.264	0.943**	0.186
Cl				1	0.041	0.156	0.619**	0.011	0.274*	−0.185	0.828**	0.867**	0.795**	0.332*	0.896**	0.154
CO3					1	0.235	−0.058	−0.068	0.061	0.182	−0.040	0.026	0.160	−0.008	0.013	0.439**
HCO3						1	0.168	0.285*	0.115	0.053	0.226	0.325*	0.266	0.133	0.318*	0.977**
SO4							1	0.350*	0.022	−0.037	0.693**	0.823**	0.737**	0.118	0.830**	0.142
NO3								1	0.116	0.005	0.154	0.197	0.233	0.240	0.196	0.248
PO4									1	−0.005	0.182	0.172	0.278*	0.399**	0.182	0.120
F										1	−0.243	−0.158	0.036	−0.215	−0.183	0.090
Ca											1	0.756**	0.627**	0.417**	0.840**	0.200
Mg												1	0.750**	0.274*	0.990**	0.306*
Na													1	0.171	0.756**	0.282*
K														1	0.316*	0.121
TH															1	0.296*
TA																1

**Correlation is significant at the 0.01 level (2-tailed)

*Correlation is significant at the 0.05 level (2-tailed)

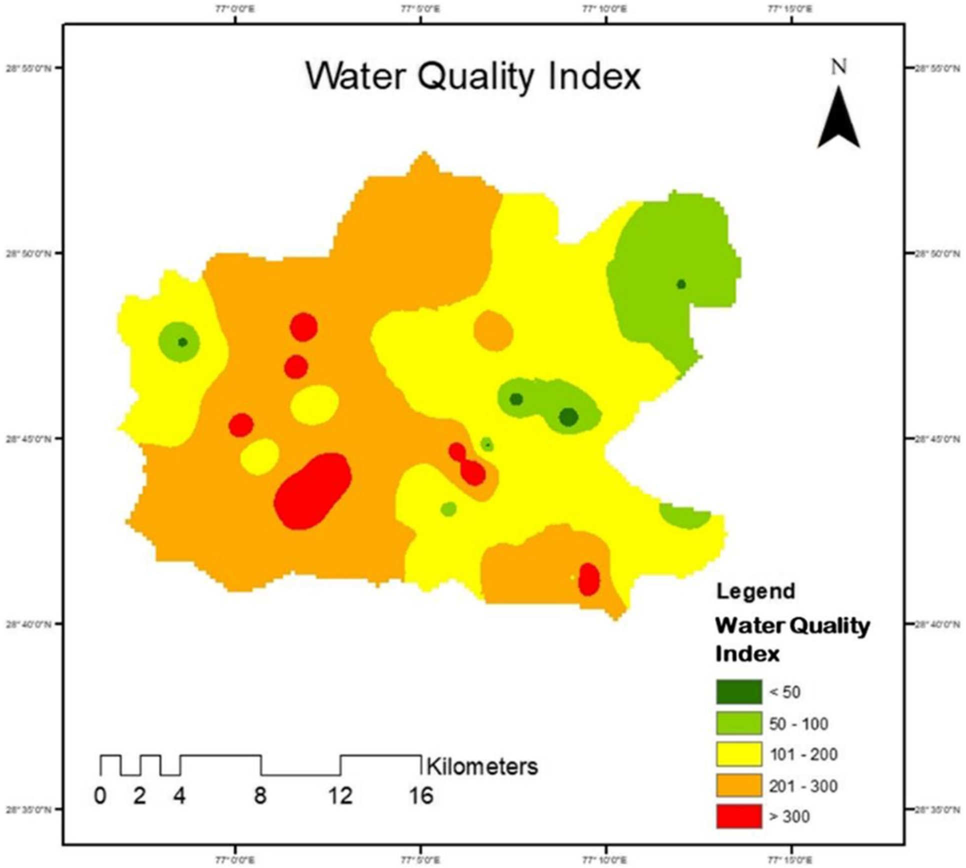


Fig. 5 Spatial distribution map of water quality index

sodium against calcium and magnesium (Kelley 1940). Water with KR > 1 indicates high sodium content and is unsuitable for irrigation. For the present study, KR values ranged from 0.04 to 1.77, with only 4 samples above KR value 1. Ninety-two percent of the samples were within the acceptable limit of KR < 1. Magnesium is important for soil productivity and maintaining soil structure. High levels of magnesium may result due to exchanges with Na^+. This in turn renders the soil alkaline which causes loss of phosphorus (Paliwal 1972; Saha et al. 2019). The magnesium hazard index classifies water for irrigation as suitable if it is < 50 and unsuitable if it is > 50. In the present study, groundwater samples had high levels of Mg^{2+}. Thus 98% of the samples were unsuitable for irrigation (Mg hazard > 50). Permeability of soil is affected by the continuous and long-term use of irrigation water and is regulated by soil Na^+, Mg^{2+}, Ca^{2+} and HCO_3^- (Snousy et al. 2022). The permeability index given by Doneen (1964) classifies water into three classes. Based on this classification, 17% of the samples were unsuitable for irrigation (class III), and 77% and 6% of the samples fall in class II and class I categories, respectively, which are suitable for irrigation.

Health risk assessment for nitrate and fluoride contamination

The groundwater in the study region is used by the local people for irrigation, industrial, and domestic purposes. Many residents in the area use the groundwater for drinking and showering. Since the samples collected had high nitrate and fluoride levels, the estimated concentrations of these pollutants were used for calculating the non-carcinogenic hazard quotient through oral and dermal exposure routes and the total hazard index according to Eqs. (4)–(9). The results obtained for hazard quotients for males, females and children are presented in Table 6.

The risk through dermal contact for nitrate was very low for all 3 categories of people, and the values were less than 1 for all samples. This result was also observed in studies by Zhang et al. (2018) and Gao et al. (2020). The total hazard quotient for nitrate ranged from 0.006 to 9.163 (mean = 1.574) for males, 0.007 to 10.829 (mean = 1.860) for females and 0.010 to 15.917 (mean = 2.734) for children. The $HQ_{nitrate}$ was greater than 1 for 44%, 46% and 52% of the samples for males, females and children respectively. Similar to nitrate, the risk through dermal exposure

Table 5 Irrigation quality of groundwater samples

Parameter	Classification	% of groundwater samples
Electrical conductivity (μS/cm)		
< 250	Low	0
250–750	Medium	13
750–2250	High	33
> 2250	Very high	54
Sodium absorption ratio		
0–10	Low sodic hazard (S1)	100
10–18	Medium sodic hazard (S2)	
18–26	High sodic hazard (S3)	
> 26	Very high sodic hazard (S4)	
Residual sodium carbonate (meq/l)		
< 1.25	Good	83
1.25 – 2.5	Doubtful	2
> 2.5	Unsuitable	15
Soluble sodium percentage (%)		
< 20	Excellent	23
20–40	Good	63
40–60	Permissible	12
60–80	Doubtful	2
> 80	Unsuitable	0
Kelley's ratio		
< 1	Good quality	92
> 1	Unsuitable	8
Magnesium hazard		
< 50	Suitable	2
> 50	Unsuitable	98
Permeability index		
Class I (> 75%)	Suitable	6
Class II (25–75%)	Good	77
Class III (< 25%)	Unsuitable	17

of fluoride was very low, demonstrating that the main health risk is through direct consumption. The total $HQ_{fluoride}$ ranged from 0.031 to 5.078 (mean = 1.433) for males, 0.037 to 6.001 (mean = 1.693) for females and 0.055 to 8.820 (mean = 2.489) for children. $HQ_{fluoride} > 1$ was observed for 58%, 58% and 69% of the samples for males, females, and children respectively. The HQ for nitrate was found to be greater than the $HQ_{fluoride}$ values across all demographics, indicating that nitrate poses a higher health risk to the residents of the study region. However, the percentage of samples with $HQ > 1$ was more for fluoride indicating that the spatial extent of risk was more for fluoride. The spatial distribution of zones with high health risks is presented in Fig. 7a and b.

The total hazard index (HI_{total}) is a summary of the total risks posed by high levels of nitrate and fluoride (Table 7). For the studied groundwater samples, the HI_{total} was found to be greater than 1 for 75%, 79% and 85% for males, females and children respectively. This indicates that the majority of the population in the study area are at some health risk, primarily from consumption of contaminated groundwater. The spatial distribution map presenting the risk zones are given in Fig. 7c. The values of HI_{total} also indicate that the risk is of the order of children > females > males. Owing to their weak resilience and higher consumption per unit of body weight, children are at a greater risk from drinking contaminated water in the study region than adults (Chen et al. 2016; Adimalla 2020; Guo et al. 2022; Xiao et al. 2022a).

Conclusions

This study analysed the groundwater quality and associated health risks in North-West Delhi, India, which is a rapidly urbanizing region. The hydrogeochemical mechanisms influencing the major ion chemistry were

Fig. 6 Groundwater suitability for irrigation according to USSL classification

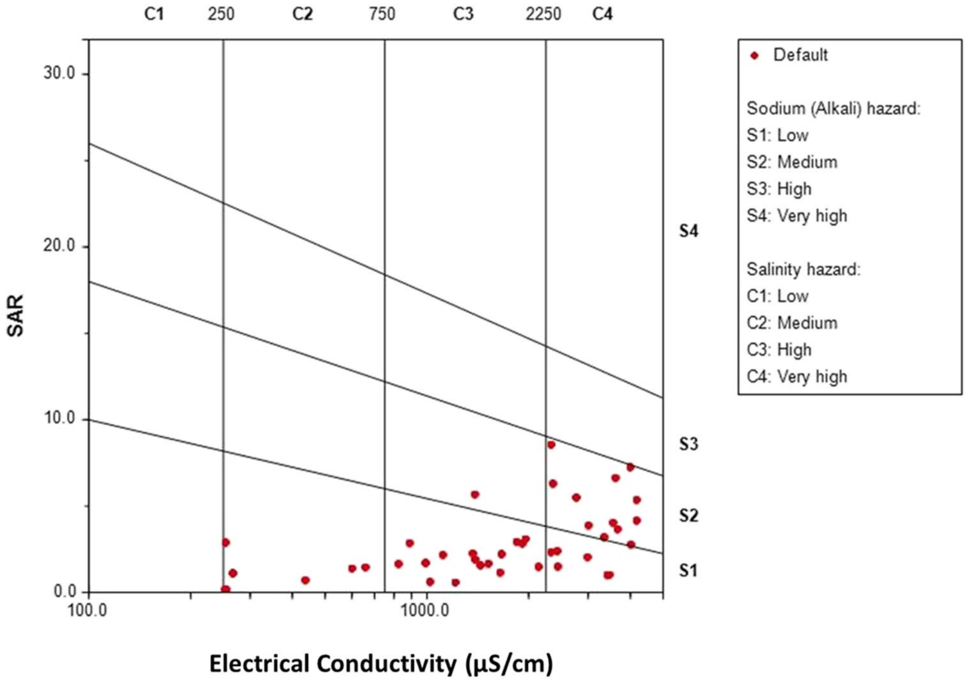


Table 6 Hazard quotient for (a) nitrate and (b) fluoride

(a)	$HQ_{nitrate\ (oral)}$			$HQ_{nitrate\ (dermal)}$			$HQ_{nitrate\ (total)}$		
	Males	Females	Children	Males	Females	Children	Males	Females	Children
Min	0.006	0.007	0.010	0.000	0.000	0.000	0.006	0.007	0.010
Max	9.139	10.801	15.841	0.024	0.029	0.076	9.163	10.829	15.917
Mean	1.570	1.855	2.720	0.004	0.005	0.013	1.574	1.860	2.734
(b)	$HQ_{fluoride\ (oral)}$			$HQ_{fluoride\ (dermal)}$			$HQ_{fluoride\ (total)}$		
	Males	Females	Children	Males	Females	Children	Males	Females	Children
Min	0.031	0.037	0.054	0.000	0.000	0.000	0.031	0.037	0.055
Max	5.064	5.985	8.778	0.013	0.016	0.042	5.078	6.001	8.820
Mean	1.429	1.689	2.477	0.004	0.004	0.012	1.433	1.693	2.489

explored, and the characteristic pollutants were identified. The dominant cations in the groundwater samples were $Mg^{2+} > Na^{+} > Ca^{2+} > K^{+}$, while the dominant anions were $HCO_3^{-} > Cl^{-} > SO_4^{2-} > NO_3^{-} > CO_3^{2-} > F^{-}$. The groundwater is slightly alkaline and TDS, TA, TH, Cl^{-}, SO_4^{2-}, Ca^{2+} and Mg^{2+} exceeded the prescribed drinking water limits in 83%, 100%, 98%, 58%, 58%, 77% and 96% of the analysed samples, respectively. The groundwater in the study region is mostly unsuitable for human consumption.

Piper trilinear diagram showed that maximum samples fell in $CaHCO_3$ type and CaMgCl type categories. The positive value obtained from chloro-alkaline indices showed that Na^{+} and K^{+} from water exchanged with Ca^{2+} and Mg^{2+} from the aquifer. Multivariate analysis using principal component analysis revealed five significant components which account for 83% of the total variance. Pearson correlation matrix indicated that major ion chemistry is influenced by several factors such as mineral dissolution, rock-water interactions and anthropogenic interferences. The water quality index for drinking was calculated for the collected groundwater samples based on the pH, TDS, TH, TA, Cl^{-}, F^{-}, $SO^{2-}{}_4$, $NO^{-}{}_3$, Ca^{2+}, and Mg^{2+} values, and 15% of the samples were found to be unfit for drinking (WQI > 300). The water samples were analysed for irrigation quality, and results showed that all samples had low sodic hazard. However, 54% of the samples had high salinity, which adversely affects crop production.

Nitrate and fluoride were above the recommended limits of 45 mg/l and 1.5 mg/l in 40% and 58% of the samples, respectively. Wastewater infiltration and fertilizer use are the primary sources of NO_3^{-} and F^{-}. High fluoride concentrations in the study region may also be due to geogenic sources. The hazard quotients for nitrate and fluoride suggested that non-carcinogenic health risk is higher

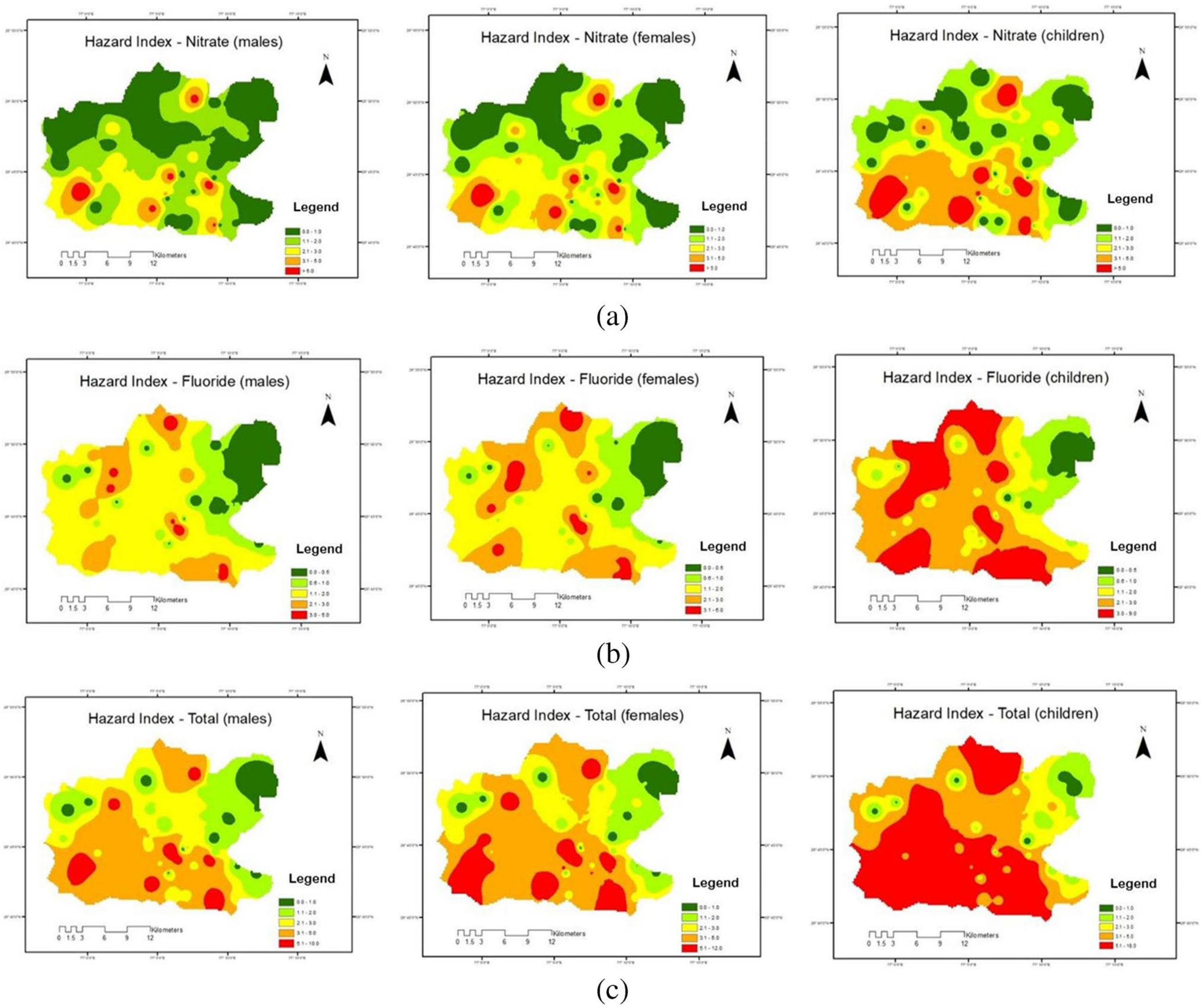

Hazard Index - Nitrate (males)
Hazard Index - Nitrate (females)
Hazard Index - Nitrate (children)
N
Legend
0.0 - 1.0
1.1 - 2.0
2.1 - 3.0
3.1 - 5.0
> 5.0
0 1.5 3 6 9 12 Kilometers
(a)
Hazard Index - Fluoride (males)
Hazard Index - Fluoride (females)
Hazard Index - Fluoride (children)
Legend
0.0 - 0.5
0.6 - 1.0
1.1 - 2.0
2.1 - 3.0
3.0 - 5.0
3.1 - 6.0
3.0 - 9.0
(b)
Hazard Index - Total (males)
Hazard Index - Total (females)
Hazard Index - Total (children)
Legend
0.0 - 1.0
1.1 - 2.0
2.1 - 3.0
3.1 - 5.0
5.1 - 10.0
5.1 - 12.0
(c)

Fig. 7 Spatial distribution of **a** $HI_{nitrate}$, **b** $HI_{fluoride}$, and **c** HI_{total} in the study area

Table 7 Total hazard index for nitrate and fluoride through oral and dermal pathways

	Total hazard index (HI_{total})	Health risk	Number of samples	% of samples
Males	≤ 1	No risk	13	25
	> 1	High risk	39	75
Females	≤ 1	No risk	11	21
	> 1	High risk	41	79
Children	≤ 1	No risk	8	15
	> 1	High risk	44	85

for nitrate contamination. However, the spatial extent of HQ > 1 was more for fluoride, implying that more people are affected by fluoride pollution in the study region. Further, it was observed that the total hazard index was in the order of children > females > males. Due to differences in body weight, children are at a greater health risk than adults. Therefore, groundwater in the study region needs to be continuously monitored and should not be used for direct consumption to avoid adverse health effects. This study is helpful in understanding the chemistry of major contaminants in aquifers of regions that are transitioning from rural to urban areas.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11356-023-26204-0.

Acknowledgements The authors acknowledge the support of the Delhi Technological University, New Delhi, for providing the facilities for this study. The authors thank Ms. Nimisha Singla and Mr. Kulvendra Patel, students of DTU, for their assistance during sampling.

Author contribution Both authors contributed to the conceptualization and design of this study. Selection of appropriate methodology, sample collection and analysis were performed by Riki Sarma. Santosh Kumar Singh provided supervision of the work. Riki Sarma prepared and edited the original draft of the manuscript, and Santosh Kumar Singh helped in editing and reviewing. The authors have read and approved the final manuscript.

Data availability All the data used in this study appear in the manuscript.

Declarations

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication The authors hereby grant all rights of publication of the manuscript to the publisher.

Competing interests The authors declare no competing interests.

References

Adimalla N (2019) Groundwater quality for drinking and irrigation purposes and potential health risks assessment: a case study from semi-arid region of south India. Expo Health 11:109–123. https://doi.org/10.1007/s12403-018-0288-8

Adimalla N, Li P (2019) Occurrence, health risks, and geochemical mechanisms of fluoride and nitrate in groundwater of the rock-dominant semi-arid region, Telangana State. India Hum Ecol Risk Assess 25(1–2):81–103. https://doi.org/10.1080/10807039.2018.1480353

Adimalla N, Li P, Qian H (2019) Evaluation of groundwater contamination for fluoride and nitrate in semi-arid region of Nirmal Province, South India: a special emphasis on human health risk assessment (HHRA). Hum Ecol Risk Assess 25(5):1107–1124. https://doi.org/10.1080/10807039.2018.1460579

Adimalla N (2020) Spatial distribution, exposure, and potential health risk assessment from nitrate in drinking water from semi-arid region of South India. Hum Ecol Risk Assess 26(2):310–334. https://doi.org/10.1080/10807039.2018.1508329

Adimalla N, Qian H (2021) Groundwater chemistry, distribution and potential health risk appraisal of nitrate enriched groundwater: a case study from the semi-urban region of South India. Ecotoxicol Environ Saf 207. https://doi.org/10.1016/j.ecoenv.2020.111277

Adimalla N, Venkatayogi S (2017) Mechanism of fluoride enrichment in groundwater of hard rock aquifers in Medak, Telangana State. South India Environ Earth Sci 76:45. https://doi.org/10.1007/s12665-016-6362-2

Alam M, Rais S, Aslam M (2012) Hydrochemical investigation and quality assessment of ground water in rural areas of Delhi. India Environ Earth Sci 66(1):97–110. https://doi.org/10.1007/s12665-011-1210-x

Ambastha SK, Haritash AK (2021) Prevalence and risk analysis of fluoride in groundwater around sandstone mine in Haryana, India. Rend Lincei 32(3):577–584. https://doi.org/10.1007/s12210-021-00997-z

APHA (2017) Standard methods for the examination of water and wastewater, 23rd edn. American Public Health Association, Washington, DC

Aravinthasamy P, Karunanidhi D, Subramani T, Roy PD (2021) Demarcation of groundwater quality domains using GIS for best agricultural practices in the drought-prone Shanmuganadhi River basin of South India. Environ Sci Pollut Res 28(15):18423–18435. https://doi.org/10.1007/s11356-020-08518-5

Barzegar R, AsghariMoghaddam A, Adamowski J et al (2017) Comparison of machine learning models for predicting fluoride contamination in groundwater. Stoch Environ Res Risk Assess 31:2705–2718. https://doi.org/10.1007/s00477-016-1338-z

Bedi S, Samal A, Ray C, Snow D (2020) Comparative evaluation of machine learning models for groundwater quality assessment. Environ Monit Assess 192:776. https://doi.org/10.1007/s10661-020-08695-3

Bhuvan (2021) Indian Geo-Platform of ISRO, National Remote Sensing Centre, Indian Space Research Organisation, Government of India. https://bhuvan.nrsc.gov.in/home/index.php. Accessed 05 Jul 2022

BIS (2012) Indian Standard Drinking Water - Specification (second revision), IS:10500. Bureau of Indian Standards, New Delhi, India

Ceballos E, Dubny S, Othax N, Zabala ME, Peluso F (2021) Assessment of human health risk of chromium and nitrate pollution in groundwater and soil of the Matanza-Riachuelo River Basin. Argentina Expo Health 13(3):323–336. https://doi.org/10.1007/s12403-021-00386-9

CGWB (2016) Aquifer mapping and ground water management plan of NCT Delhi. Central Ground Water Board, State Unit Office, Delhi

CGWB (2021) Groundwater yearbook national capital territory, Delhi (2019–2020). Central Ground Water Board, State Unit Office, Delhi

CGWB (n.d.) Ground water information booklet of North-west district, NCT, Delhi. Central Ground Water Board, State Unit Office, Delhi http://cgwb.gov.in/District_Profile/Delhi_districtprofile.html. Accessed 30 June 2022

Chaudhuri RR, Sharma P (2020) Addressing uncertainty in extreme rainfall intensity for semi-arid urban regions: case study of Delhi, India. Nat Hazards 104:2307–2324. https://doi.org/10.1007/s11069-020-04273-5

Chen J, Wu H, Qian H (2016) Groundwater nitrate contamination and associated health risk for the rural communities in an agricultural area of Ningxia, northwest China. Expo Health 8:349–359. https://doi.org/10.1007/s12403-016-0208-8

Chen F, Yao L, Mei G, Shang Y, Xiong F, Ding Z (2021) Groundwater quality and potential human health risk assessment for drinking and irrigation purposes: a case study in the semiarid region of north China. Water 13(6). https://doi.org/10.3390/w13060783

Davis SN, DeWiest RJM (1966) Hydrogeology. John Wiley & Sons, New York, pp 463

Dhakate R, More S, Duvva LK et al (2023) Groundwater chemistry and health hazard risk valuation of fluoride and nitrate enhanced groundwater from a semi-urban region of South India. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-023-25287-z

Domenico PA, Schwartz FW (1990) Physical and chemical hydrogeology. Wiley, New York

Doneen LD (1964) Water quality for agriculture. University of California, Davis, pp 48

Duvva LK, Panga KK, Dhakate R, Himabindu V (2022) Health risk assessment of nitrate and fluoride toxicity in groundwater contamination in the semi-arid area of Medchal, South India. Appl Water Sci 12(1). https://doi.org/10.1007/s13201-021-01557-4

Eaton FM (1950) Significance of carbonates in irrigation waters. Soil Sci 39:123–133

Elemile OO, Ibitogbe EM, Folorunso OP, Ejiboye PO, Adewumi JR (2021) Principal component analysis of groundwater sources pollution in Omu-Aran Community, Nigeria. Environ Earth Sci 80:690. https://doi.org/10.1007/s12665-021-09975-y

Freeze RA, Cherry JA (1979) Groundwater. Prentice Hall Inc, New Jersey

Gao S, Li C, Jia C, Zhang H, Guan Q, Wu X, Wang J, Lv M (2020) Health risk assessment of groundwater nitrate contamination: a case study of a typical karst hydrogeological unit in East China. Environ Sci Pollut Res 27(9):9274–9287. https://doi.org/10.1007/s11356-019-07075-w

Giri S, Mahato MK, Singh PK, Singh AK (2021) Non-carcinogenic health risk assessment for fluoride and nitrate in the groundwater of the mica belt of Jharkhand, India. Hum Ecol Risk Assess 27(7):1939–1953. https://doi.org/10.1080/10807039.2021.1934814

Glass C, Silverstein J (1998) Denitrification kinetics of high nitrate concentration water: pH effect on inhibition and nitrite accumulation. Water Res 32(3):831–839. https://doi.org/10.1016/S0043-1354(97)00260-1

Golaki M, Azhdarpoor A, Mohamadpour A, Derakhshan Z, Conti GO (2022) Health risk assessment and spatial distribution of nitrate, nitrite, fluoride, and coliform contaminants in drinking water resources of Kazerun, Iran. Environ Res 203. https://doi.org/10.1016/j.envres.2021.111850

Goyal D, Haritash AK, Singh SK (2021) A comprehensive review of groundwater vulnerability assessment using index-based, modelling and coupling methods. J Environ Manage. https://doi.org/10.1016/j.jenvman.2021.113161

Gugulothu S, Subba Rao N, Das R, Duvva LK, Dhakate R (2022) Judging the sources of inferior groundwater quality and health risk problems through intake of groundwater nitrate and fluoride from a rural part of Telangana. Environ Sci Pollut Res, India. https://doi.org/10.1007/s11356-022-18967-9

Guo Y, Li P, He X, Wang L (2022) Groundwater quality in and around a landfill in Northwest China: characteristic pollutant identification, health risk assessment, and controlling factor analysis. Expo Health 14:885–901. https://doi.org/10.1007/s12403-022-00464-6

He X, Li P, Wu J, Wei M, Ren X, Wang D (2021) Poor groundwater quality and high potential health risks in the Datong Basin, northern China: research from published data. Environ Geochem Health 43(2):791–812. https://doi.org/10.1007/s10653-020-00520-7

Jandu A, Malik A, Dhull SB (2021) Fluoride and nitrate in groundwater of rural habitations of semiarid region of northern Rajasthan, India: a hydrogeochemical, multivariate statistical, and human health risk assessment perspective. Environ Geochem Health 43(10):3997–4026. https://doi.org/10.1007/s10653-021-00882-6

Jiang W, Sheng Y, Liu H, Ma Z, Song Y, Liu F, Chen S (2022) Groundwater quality assessment and hydrogeochemical processes in typical watersheds in Zhangjiakou region, northern China. Environ Sci Pollut Res 29(3):3521–3539. https://doi.org/10.1007/s11356-021-15644-1

Karunanidhi D, Aravinthasamy P, Priyadarsi DR, Praveenkumar RM, Prasanth K, Selvapraveen S (2020) Evaluation of non-carcinogenic risks due to fluoride and nitrate contaminations in a groundwater of an urban part (Coimbatore region) of South India. Environ Monit Assess 192:102. https://doi.org/10.1007/s10661-019-8059-y

Kaur L, Rishi MS, Siddiqui AU (2020) Deterministic and probabilistic health risk assessment techniques to evaluate non-carcinogenic human health risk (NHHR) due to fluoride and nitrate in groundwater of Panipat, Haryana, India. Environ Pollut 259. https://doi.org/10.1016/j.envpol.2019.1137

Kelley WP (1940) Permissible composition and concentration of irrigation waters. Proc Am Soc Civil Eng 66:07–613

Khan AF, Srinivasamoorthy K, Prakash R, Gopinath S, Saravanan K, Vinnarasi F, Babu C, Rabina C (2021) Human health risk assessment for fluoride and nitrate contamination in the groundwater: a case study from the east coast of Tamil Nadu and Puducherry, India. Environ Earth Sci 80(21). https://doi.org/10.1007/s12665-021-10001-4

Kumar PJS, James EJ (2016) Identification of hydrogeochemical processes in the Coimbatore district Tamil Nadu, India. Hydrol Sci J 61(4):719–731. https://doi.org/10.1080/02626667.2015.1022551

Li P, Karunanidhi D, Subramani T, Srinivasamoorthy K (2021a) Sources and consequences of groundwater contamination. Arch Environ Contam Toxicol 80:1–10. https://doi.org/10.1007/s00244-020-00805-z

Li Y, Li P, Cui X, He S (2021b) Groundwater quality, health risk, and major influencing factors in the lower Beiluo River watershed of northwest China. Hum Ecol Risk Assess 27(7):1987–2013. https://doi.org/10.1080/10807039.2021.1940834

Liu J, Ma Y, Gao Z, Zhang Y, Sun Z, Sun T, Fan H, Wu B, Li M, Qian L (2022) Fluoride contamination, spatial variation, and health risk assessment of groundwater using GIS: a high-density survey sampling in Weifang City, North China. Environ Sci Pollut Res 29(23):34302–34313. https://doi.org/10.1007/s11356-021-18443-w

Makubalo SS, Diamond RE (2020) Hydrochemical evolution of high uranium, fluoride and nitrate groundwaters of Namakwaland, South Africa. J African Earth Sci 172. https://doi.org/10.1016/j.jafrearsci.2020.104002

Masood A, Aslam M, Pham QB, Khan W, Masood S (2022) Integrating water quality index, GIS and multivariate statistical techniques towards a better understanding of drinking water quality. Environ Sci Pollut Res 29(18):26860–26876. https://doi.org/10.1007/s11356-021-17594-0

Maurya J, Pradhan SN, Seema GAK (2020) Evaluation of ground water quality and health risk assessment due to nitrate and fluoride in the Middle Indo-Gangetic plains of India. Hum Ecol Risk Assess 27(5):1349–1365. https://doi.org/10.1080/10807039.2020.1844559

Motlagh AM, Yang Z, Saba H (2020) Groundwater quality. Water Environ Res 92(10):1649–1658. https://doi.org/10.1002/wer.1412

Mukherjee I, Singh UK (2018) Groundwater fluoride contamination, probable release, and containment mechanisms: a review on Indian context. Environ Geochem Health 40:2259–2301. https://doi.org/10.1007/s10653-018-0096-x

Nawale VP, Malpe DB, Marghade D, Yenkie R (2021) Non-carcinogenic health risk assessment with source identification of nitrate and fluoride polluted groundwater of Wardha sub-basin, central India. Ecotoxicol Environ Saf 208. https://doi.org/10.1016/j.ecoenv.2020.111548

Orhan O (2021) Monitoring of land subsidence due to excessive groundwater extraction using small baseline subset technique in Konya, Turkey. Environ Monit Assess 193(4). https://doi.org/10.1007/s10661-021-08962-x

Paliwal KV (1972) Irrigation with saline water. Indian Agricultural Research Institute (IARI) Monograph No. 2 (New Series), IARI, New Delhi, pp 198

Panneerselvam B, Muniraj K, Duraisamy K, Pande C, Karuppannan S, Thomas M (2022) An integrated approach to explore the suitability of nitrate-contaminated groundwater for drinking purposes in a semiarid region of India. Environ Geochem Health. https://doi.org/10.1007/s10653-022-01237-5

Panneerselvam B, Muniraj K, Pande C, Ravichandran N, Thomas M, Karuppannan S (2021) Geochemical evaluation and human health risk assessment of nitrate-contaminated groundwater in an industrial area of South India. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-021-17281-0

Piper AM (1944) A graphical interpretation of water analysis. EOS Trans Am Geophys Union 25:914–928. https://doi.org/10.1029/TR025i006p00914

Raghunath HM (1987) Groundwater, 2nd edn. Wiley Eastern Ltd., New Delhi

Rahman A, Mondal NC, Tiwari KK (2021) Anthropogenic nitrate in groundwater and its health risks in the view of background concentration in a semi-arid area of Rajasthan, India. Sci Rep 11(1). https://doi.org/10.1038/s41598-021-88600-1

Rahman MM, Islam MA, Bodrud-Doza M, Muhib MI, Zahid A, Shammi M, Tareq SM, Kurasaki M (2018) Spatio-temporal assessment of groundwater quality and human health risk: a case study in Gopalganj. Bangladesh Expo Health 10(3):167–188. https://doi.org/10.1007/s12403-017-0253-y

Ram A, Tiwari SK, Pandey HK, Chaurasia AK, Singh S, Singh YV (2021) Groundwater quality assessment using water quality index (WQI) under GIS framework. Appl Water Sci 11(2). https://doi.org/10.1007/s13201-021-01376-7

Rashid A, Ayub M, Khan S, Ullah Z, Ali L, Gao X, Li C, El-Serehy HA, Kaushik P, Rasool A (2022) Hydrogeochemical assessment of carcinogenic and non-carcinogenic health risks of potentially toxic elements in aquifers of the Hindukush ranges, Pakistan: insights from groundwater pollution indexing, GIS-based, and multivariate statistical approaches. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-022-21172-3

Reddy CKVC, Golla V, Badapalli PK, Reddy NBY (2022) Evaluation of groundwater contamination for fluoride and nitrate in Nellore Urban Province, Southern India: a special emphasis on human health risk assessment (HHRA). Appl Water Sci 12(3). https://doi.org/10.1007/s13201-021-01537-8

Rezaei M, Nikbakht M, Shakeri A (2017) Geochemistry and sources of fluoride and nitrate contamination of groundwater in Lar area, south Iran. Environ Sci Pollut Res 24(18):15471–15487. https://doi.org/10.1007/s11356-017-9108-0

Richards LA (1954) Diagnosis and improvement of saline and alkali soils. Soil Sci 78(2):154

Rishi MS, Kaur L, Sharma S (2020) Groundwater quality appraisal for non-carcinogenic human health risks and irrigation purposes in a part of Yamuna sub-basin, India. Hum Ecol Risk Assess 26(10):2716–2736. https://doi.org/10.1080/10807039.2019.1682514

Rufino F, Busico G, Cuoco E, Darrah TH, Tedesco D (2019) Evaluating the suitability of urban groundwater resources for drinking water and irrigation purposes: an integrated approach in the Agro-Aversano area of Southern Italy. Environ Monit Assess 191(12). https://doi.org/10.1007/s10661-019-7978-y

Saha S, Reza AHMS, Roy MK (2019) Hydrochemical evaluation of groundwater quality of the Tista floodplain, Rangpur. Bangladesh Appl Water Sci 9:198. https://doi.org/10.1007/s13201-019-1085-7

Sarkar M, Pal SC, Islam ARMT (2022) Groundwater quality assessment for safe drinking water and irrigation purposes in Malda district, Eastern India. Environ Earth Sci 81:52. https://doi.org/10.1007/s12665-022-10188-0

Sarma R, Singh SK (2021) Simulating contaminant transport in unsaturated and saturated groundwater zones. Water Environ Res 93(9):1496–1509. https://doi.org/10.1002/wer.1555

Sarma R, Singh SK (2022) A comparative study of data-driven models for groundwater level forecasting. Water Resour Manage 36:2741–2756. https://doi.org/10.1007/s11269-022-03173-6

Sathe SS, Mahanta C, Subbiah S (2021) Hydrogeochemical evaluation of intermittent alluvial aquifers controlling arsenic and fluoride contamination and corresponding health risk assessment. Expo Health 13(4):661–680. https://doi.org/10.1007/s12403-021-00411-x

Sawyer CN, McCarty PL (1967) Chemistry for sanitary engineers, 2nd edn. McGraw-Hill, New York

Schoeller H (1965) Qualitative evaluation of groundwater resources. In Methods and techniques of groundwater investigation and development. UNESCO, pp 54–83

Sidhu BS, Sharma D, Tuteja T, Gupta S, Kumar A (2015) Human health risk assessment of heavy metals from Bhalaswa Landfill, New Delhi, India. In: Raju N, Gossel W, Sudhakar M (eds) Management of natural resources in a changing environment. Springer, Cham. https://doi.org/10.1007/978-3-319-12559-6_16

Singh CK, Kumar A, Shashtri S, Kumar A, Kumar P, Mallick J (2017) Multivariate statistical analysis and geochemical modeling for geochemical assessment of groundwater of Delhi, India. J Geochem Explor 175:59–71. https://doi.org/10.1016/j.gexplo.2017.01.001

Singh G, Rishi MS, Herojeet R, Kaur L, Sharma K (2020) Evaluation of groundwater quality and human health risks from fluoride and nitrate in semi-arid region of northern India. Environ Geochem Health 42(7):1833–1862. https://doi.org/10.1007/s10653-019-00449-6

Snousy MG, Wu J, Su F, Abdelhalim A, Ismail E (2022) Groundwater quality and its regulating geochemical processes in Assiut Province. Egypt Expo Health 14(2):305–323. https://doi.org/10.1007/s12403-021-00445-1

Stylianoudaki C, Trichakis I, Karatzas GP (2022) Modeling groundwater nitrate contamination using artificial neural networks. Water 14:1173. https://doi.org/10.3390/w14071173

Subba Rao N (2017) Controlling factors of fluoride in groundwater in a part of South India. Arab J Geosci 10:524. https://doi.org/10.1007/s12517-017-3291-7

Subba Rao N, Deepali M, Dinakar A, Chandana I, Sunitha B, Ravindra B, Balaji T (2017) Geochemical characteristics and controlling factors of the chemical composition of groundwater in a part of Guntur district, Andhra Pradesh, India. Environ Earth Sci 76:747. https://doi.org/10.1007/s12665-017-7093-8

Subba Rao N, Dinakar A, Kumari BK (2021) Appraisal of vulnerable zones of non-cancer-causing health risks associated with exposure of nitrate and fluoride in groundwater from a rural part of India. Environ Res 202. https://doi.org/10.1016/j.envres.2021.111674

Swain S, Sahoo S, Taloor AK (2022) Groundwater quality assessment using geospatial and statistical approaches over Faridabad and Gurgaon districts of National Capital Region, India. Appl Water Sci 12:75. https://doi.org/10.1007/s13201-022-01604-8

Todd DK (1980) Groundwater hydrology, 2nd edn. John Wiley, New York

US Salinity Laboratory (USSL) (1954) Diagnosis and improvement of saline and alkaline soils. Government Printing Office, Washington

USEPA (1989) Risk assessment guidance for superfund, Volume I: Human Health Evaluation Manual (Part A). United States Environmental Protection Agency, Washington, DC

USEPA (1991) Risk assessment guidance for superfund, Volume I: Human Health Evaluation Manual (Part B, Development of Risk-based Preliminary Remediation Goals). United States Environmental Protection Agency, Washington, DC

USEPA (1997) Exposure factors handbook, Volume 1: General Factors. United States Environmental Protection Agency, Washington, DC

USEPA (2004) Risk assessment guidance for superfund, Volume I: Human Health Evaluation Manual (Part E). United States Environmental Protection Agency, Washington, DC

USEPA (2014) Human health evaluation manual, supplemental guidance: update of standard default exposure factors, OSWER Directive 9200.1–120. United States Environmental Protection Agency, Washington, DC

Varol S, Şener Ş, Şener E (2021) Assessment of groundwater quality and human health risk related to arsenic using index methods and GIS: a case of Şuhut Plain (Afyonkarahisar/Turkey). Environ Res 202. https://doi.org/10.1016/j.envres.2021.111623

WHO (2011) Guideline for drinking water quality, 4th edn. World Health Organization, Geneva

WHO (2017) Guidelines for drinking water quality, 4th edn. Incorporating the First Addendum. World Health Organization, Geneva

Wilcox LV (1955) Classification and use of irrigation waters. US Department of Agriculture, New York

Wilopo W, Putra DPE, Hendrayana H (2021) Impacts of precipitation, land use change and urban wastewater on groundwater level fluctuation in the Yogyakarta-Sleman Groundwater Basin, Indonesia. Environ Monit Assess 193(2). https://doi.org/10.1007/s10661-021-08863-z

Wu J, Li P, Wang D, Ren X, Wei M (2020) Statistical and multivariate statistical techniques to trace the sources and affecting factors of groundwater pollution in a rapidly growing city on the Chinese Loess Plateau. Hum Ecol Risk Assess 26:1603–1621. https://doi.org/10.1080/10807039.2019.1

Xiao Y, Hao Q, Zhang Y, Zhu Y, Yin S, Qin L, Li X (2022a) Investigating sources, driving forces and potential health risks of nitrate and fluoride in groundwater of a typical alluvial fan plain. Sci Total Environ 802. https://doi.org/10.1016/j.scitotenv.2021.149909

Xiao Y, Liu K, Hao Q et al (2022b) Occurrence, controlling factors and health hazards of fluoride-enriched groundwater in the lower flood plain of Yellow River, Northern China. Expo Health 14:345–358. https://doi.org/10.1007/s12403-021-00452-2

Zhang Y, Wu J, Xu B (2018) Human health risk assessment of groundwater nitrogen pollution in Jinghui canal irrigation area of the loess region, northwest China. Environ Earth Sci 77:273. https://doi.org/10.1007/s12665-018-7456-9

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

A Survey on Smart Parking Management System

Rohit Ramchandani[1], Anjali Bansal[2]

Delhi Technological University, New Delhi, India

Corresponding author: Anjali Bansal, Email: anjalibansal791@gmail.com

With the large increase in population, automated industries and need of vehicles, parking of the vehicles is becoming a critical issue in various cities. Unmanaged parking of vehicles leads to noise pollution, air pollution, traffic congestion. During peak hours, it is difficult task to find vacant parking lot and it becomes the major challenge for driver to park the vehicle. A lot of work is being done in the whole world to manage the efficient parking of vehicles. To give the clear overview about efficient parking system, we go through some existing studies over the period of 2009-2022 which proposed various parking solutions. This survey gives an exhaustive study of available parking solutions and also proposed some recommendations for future research in providing smart parking management system.

Keywords: Smart Parking, IoT, Zigbee, Firefly Algorithm, Edge Computing

2023. In Satyasai Jagannath Nanda & Rajendra Prasad Yadav (eds.), *Data Science and Intelligent Computing Techniques*, 35–46. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.56155/978-81-955020-2-8-4

1. Introduction

IoT (Internet of Things) has changed the life of many peoples as it provides the ease. Due to this, the use of IoT devices is increasing day by day in every area such as smart home automation, smart city, smart parking system etc. With the rapid growth in use of IoT devices and cloud systems, smart cities have great opportunity in changing the people's lives and leads to technological development and ease in accessibility. Smart parking is one of the main concepts in smart cities as the efficient parking system leads to reduction in environment pollution by reducing the fuel and time consumption. Fahim *et al.* [1] presented a review to describe and compare existing smart parking approaches based on sensors used, networking technologies adopted, computational approaches used, and user interfaces used. They have also described pros and cons of existing smart parking approaches and sensors used.

Kalid *et al.* [2] studied all the studies related to smart parking solutions (digitally enhanced parking, empty slot detection, and route planning) and autonomous valet parking solutions (short range autonomous valet parking and long range autonomous valet parking).

Fig 1 [3] shows use case diagram of smart parking management system. This use case diagram includes 8 different actors and 8 different use cases for both administration and operation viewpoints. This use case diagram gives all the details about managing parking data, generating billing receipt, updating parking status and setting the parking rates.

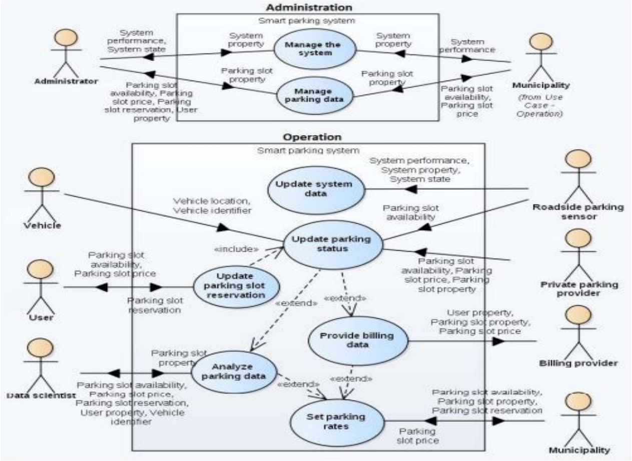

Administration
Smart parking system
Administrator
System performance, System state
System property
Manage the system
System property
System performance
Municipality
(from Use Case - Operation)
Parking slot property
Manage parking data
Parking slot property
Parking slot availability, Parking slot price, Parking slot reservation, User property
Parking slot availability, Parking slot price
Operation
Smart parking system
Update system data
System performance, System property, System state
Roadside parking sensor
Vehicle
Vehicle location, Vehicle identifier
Parking slot availability
Update parking status
«include»
Private parking provider
Parking slot availability, Parking slot price, Parking slot property
User
Parking slot availability, Parking slot price
Update parking slot reservation
Parking slot reservation
«extend»
«extend»
Provide billing data
User property, Parking slot property, Parking slot price
Billing provider
Data scientist
Parking slot property
Analyze parking data
Parking slot availability, Parking slot price, Parking slot reservation, User property, Vehicle identifier
«extend»
«extend»
Set parking rates
Parking slot availability, Parking slot property, Parking slot reservation
Parking slot price
Municipality

Fig. 1: Use Case Diagram of Smart Parking System

Smart parking system is divided into two flows: information flow and traffic flow. Both flows are represented in fig 2 [3]. The traffic flow gives the information about vehicle events and provides the path. The information flow gives all the information related to parking of the vehicles at all the moments such as from the moment when sensors detect the vehicles to the moment when parking reservation displays on the driver's terminal.

Various authors gave many methods and algorithms to manage the efficient parking of vehicles. In this paper, we have given the existing studies over the period 2009-2021 mainly focused on providing various parking solutions. In these studies, the authors have been used various machine learning, deep learning and ensemble techniques for building prediction models, metaheuristic approach such as feed forward back propagation neural network and firefly algorithm and various devices such as Zigbee, BLE beacons etc. Some of the authors have also proposed algorithms to build the efficient parking systems.

The organization of this paper is divided in various sections which are as follows: Section 2 describes the existing methodologies which have already proposed some parking solutions. Section 3 describes the conclusion based on all the studies considered in this survey. Section 4 describes the research gaps found from these studies or in other terms we can say that section 4 describes the future research works which can be done to provide the efficient parking solutions.

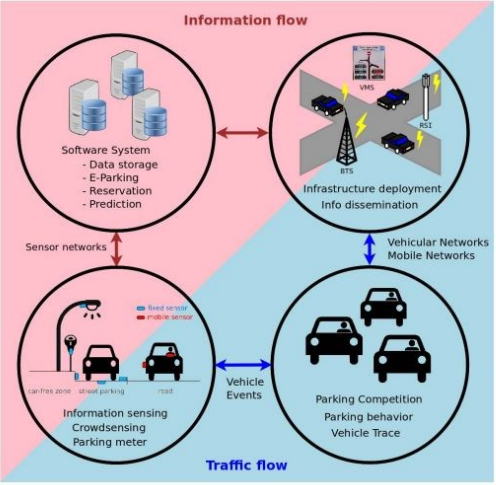


Fig. 2: Information flow and Traffic flow

2 Existing Methodology

Smart parking is the most important part in smart cities which helps in improving the quality-of-life cycle in cities. As production of vehicles are increasing day by day, these vehicles on the streets leads to time and fuel consumption due to which environment gets polluted. So, to overcome this problem, there is a need to build the efficient parking solutions. In this section we will go through some existing studies which helps in introducing efficient parking management system.

There is a need to build an automated, cost-effective, real time and easy-to-use parking management system for car parking. Srikanth *et al.* [4] proposed a Smart PARKing (SPARK) management system which was used to satisfy the requirements of car parking management system. They used wireless technology to provide automatic guidance, effective parking reservation mechanism and remote parking monitoring.

Many approaches have been stated for smart parking but due to covid 19 outbreak there is a need to build a model which also considers social distancing measure because social distancing is the only preventive measure to overcome the virus. Thierry Delota and Sergio Iiarri [5] proposed a method that gives recommendation to drivers about safe vacant parking slots. The main objective of their study was to introduce a methods which gives the suggestions of available parking slots on the basis of social distance measure and also this system maximize the safety of vehicles and the people available in parking lot.

H. Canli and S. Toklu [6] proposed a deep learning-based application. For the experimentation purpose they used ISTPARK dataset. The objective of their study was to build a deep learning and cloud-based application so that the searching time for vacant parking slot get reduced and they also used deep learning with LSTM in the proposed application to predict the parking space.

Tang *et al.* [7] in their study generated a fog computing based smart parking system to improve the real time smart parking system. They performed the experiment on Golden eagle mall parking space. The proposed system combines the benefits of both VANET's and fog computing in order to reduce gasoline waste, average parking cost and vehicle exhaust emission, and this system also improves the parking facilities.

Lin *et al.* [8] introduced a Smart Parking Algorithm (SPA) whose working is based on the behaviour of the driver. This proposed algorithm maximizes the benefits of the parking space owner and improves the service quality. Also, this algorithm predicts the parking time based on the past data of parking records.

Misra *et al.* [9] introduced an intelligent parking scheme to fulfil the vision of parking 4.0. The objective of this study was to fulfil the vision of parking 4.0 as a digital reimagination of the end-to-end parking chains as to provide collaborative ecosystem. This scheme provides right parking at correct place at cheap price.

As vehicles are increasing day by day then finding the empty parking space is a major issue. Conventional methods for parking system are very costly as they use installation of sensors at every parking space. Singh *et al.* [10] build an improved parking system so that vacant parking space searching time gets reduced. The objective of this study was to build a system for park the vehicles using metaheuristic approach such as using feed forward back propagation neural network and firefly algorithm. This approach considers two parameters parking efficiency and parking space search time. Parking efficiency is improved using firefly algorithm and parking space search time is reduced using feed forward back propagation neural network.

Mackey *et al.* [11] build a system which was based on BLE beacons. The objective of this study was to provide smart parking system for both indoor and outdoor spaces. In this system, each parking space is paired with unique BLE beacon to provide guidance and secure payment system and to improve the accuracy, particle filter is used. In this study they used BLE beacons, Google's Eddystone protocol and MATLAB for experimentation work. They introduced Smartphone application which was based on BLE beacon devices and they used Google's Eddystone protocol for secure payments. This study results in more accurate result for checking parking availability.

Provoost *et al.* [12] proposed a prediction system which predicts the parking occupancy via ML techniques such as Random Forest and Neural Network. They did the experimentation on various places such as Gelredome stadium, Dutch metrological institute KNMI, Weerlive API, National Databank Wegverkeersgegevens, Open data service of NDW, Centraal Garage, Municipality portal parking data. The objective of this study was to examine effect of Web of things and artificial intelligence to foresee the vacant parking space. In this study traffic cameras were used as web of things sensors. The result of this study show that ML methods score a MSE (Mean Squared Error) of 7.18 in a time duration of 60 min.

As there is an increase in vehicles in the city, it becomes difficult task to find empty parking space.

Sarang Deshpande [13] proposed M-Parking: algorithm which uses hierarchical wireless sensor networks for vehicle parking guidance system. In this study they used ADC, flash memory, transceiver, Zigbee. M-Parking is more energy efficient as there are 3 passive sensors per parking slots and results are accurate as single active sensor per parking slot.

According to a survey, drivers spend lot of time to search for empty parking space in parking lot. Tekouabou *et al.* [14] proposed a prediction model on the basis of ensemble techniques such as bagging and boosting and IoT. The main objective of this study was to build an integrated model of

IOT and regression algorithms to build prediction model so that drivers can predict available parking space in smart parking. They used Birmingham parking dataset for experimentation. The result of this study shows that Bagging regressor model improves the best existing prediction performance by [166.6% and reduces the system complexity.

Zhang *et al.* [15] proposed a parking system which makes use of Edge Computing. The objective of this study was to generate a P2P based smart parking management system which uses cloud computing, edge computing and P2P network techniques which helps in navigation, enquiries etc. To perform the experiment, they used XD Smart Park protocol and various algorithms such edge computing, cloud computing, P2P algorithm and mobile nodes.

Jong-Ho Shin and Hong-Bae Jun [16] introduced a smart parking guidance algorithm which provides actual time status of availability of parking in smart cities. This algorithm considers many parameters such as driving distance to the guided parking facility, expected parking cost, walking distance from the guided parking facility to destination, and traffic congestion due to parking guidance. The experimentation is performed in Luxembourg city. Proposed algorithm helps in effective usage of parking spaces in the city. It reduces the energy consumption and traffic congestion in the city.

Qadir *et al.* [17] proposed smart parking system which was time and energy efficient and based on Zigbee. The main purpose of their study was to provide communication between various devices with low power consumption and more effective way of parking by sharing the actual time scenario of nearby vacant parking lots. They used Arduino UNO interface and digital transceiver for experimentation purpose. Also, this system can provide location of the vehicle to driver if the vehicle gets theft.

Pampa Sadhukhan [18] proposed an IoT based E parking system. The objective was to build an Internet of things-based E- parking system which uses parking meter to solve various issues such as estimating parking usage by each vehicle, collecting parking charges, detecting improper parking. Parking meter, WLAN or Wifi enabled laptop/workstation, Wifi access points were used to perform the experiment. This system provides reservation based smart parking facility, smart payment to collect payment charges, detects improper parking vehicles within parking lot.

Alharbi *et al.* [19] proposed web application based on OCR algorithm for solving smart parking problem. This application provides the facility of pre-reservation of vacant slots in order to avoid traffic congestion. As wireless sensors play an important role in building smart parking solutions, so Kumar *et al.* [20] proposed an intelligent approach for smart parking solution which is based on wireless sensors. Table 1. shows the summary of selected existing studies related to smart parking system

Table 1: Summary of the selected existing studies

Author Name	**Year**	**Keywords**	**Dataset and Component used**	**Findings**
Srikanth *et al.* [4]	2009	Wireless Sensor Networks (WSN), Automated Guidance, Smart Parking, Lot Reservation, Remote Monitoring.	Sensor node, Sink node, GSM device, LEDs	Proposed a Smart PARKing (SPARK) managen system which provide automatic guidance, effe parking reservation mechanism and remote pa monitoring.
Jong-Ho Shin, Hong-Bae Jun [16]	2014	Parking facility, Smart parking guidance, City transportation management, Parking guidance algorithm.	Luxembourg city five objects: parking lot, central server, personal navigation device, parking management system, driver.	Proposed algorithm which helps in effective usage of parking spaces in the city. It reduces the energy consumption and traffic congestion in the city and also provides actual time status of parking availability in a city.
Sarang Deshpande [13]	2016	Hierarchical Wireless Sensor Networks, Parking Information Subscription, Vehicle Parking Guidance.	ADC, flash memory, transceiver, Zigbee	Proposed an algorithm using hierarchical wireless sensor networks for vehicle parking guidance system. It is more energy efficient as there are 3 passive sensors per parking slots and results are accurate as single active sensor per parking slot.

Pampa Sadhukhan [18]	2017	Smart parking system (SPS), Internet-of-Things (IoT), parking meter (PM), E-parking, parking lot.	Parking meter, WLAN or Wifi enabled laptop/workstation, Wifi access points	Provides reservation based smart parking facility, smart payment to collect payment charges, detects improper parking vehicles within parking lot.
Tang *et al.* [7]	2018	Parking slot, fog computing architecture, Smart VANETs, real time.	Golden eagle mall	Fog computing based smart parking system that combines the benefits of both VANET's and fog computing in order to reduce gasoline waste, average parking cost and vehicle exhaust emission and improvement in parking facilities.
Qadir *et al.* [17]	2018	Internet of Things (IOT), Arduino, Microcontroller, ZigBee, GSM, IR Sensors.	Arduino UNO Interface, Digital Transceiver	Zigbee is both time and energy efficient as compare to bluetooth and Wi fi. Also, this system can provide location of the vehicle to driver if the vehicle gets theft.
Misra *et al.* [9]	2019	Parking 4.0	IPX (Intelligent Parking Scheme)	This provides right parking at correct place at cheap price.
Zhang *et al.* [15]	2020	Smart parking, Cloud computing, P2P network, Edge computing, Intelligent transportation system.	XD Smart Park, cloud computing, edge computing, mobiles, mobile nodes, P2P algorithm	Proposed a P2P based smart parking management system which is more friendly and effective smart parking management system as compared to others which is based on P2P based algorithms. Introduced XD smart park protocol to implement this system.
Tekouabou *et al.* [14]	2020	Smart cities, IoT, Regression,	Birmingham parking dataset	Integrated model of IOT and ensemble methods to predict the available parking space in smart car parks. Bagging regressor

		Parking availability, Ensemble models.		model improves the best existing prediction performance by 6.6% and reduces the system complexity.
Mackey *et al.* [11]	2020	Eddystone, Bluetooth low energy (BLE) beacons, Particle filter, Parking availability estimation, Internet of Things (IoT), Smart cities, Smart parking.	BLE beacon device, Google's eddystone protocol, Matlab	Smartphone application based on BLE beacon devices and for secure payments system, google's eddystone protocol is used. It gives more accurate result for checking parking availability.
Provoost *et al.* [12]	2020	Internet of Things, Neural networks, Machine learning, Parking occupancy, Web of Things.	Gelredome stadium, Dutch metrological institute KNMI, Weerlive API, National Databank Wegverkeersgegevens, Open data service of NDW, Centraal Garage, Municipality portal parking data.	The prediction performance of the Machine Learning models for searching vacant parking space was better than the previous work in the problem under study. ML methods achieve a MSE (Mean Squared Error) of 7.18 in a time period of 60 min. The historical rate of occupied space (i.e., the look-back window) was the most important forecast variable, followed by traffic flows calculated at the orbital highways.
H. Canli and S. Toklu [6]	2021	Smart city, SVM, Deep learning, RF, LSTM, ARIMA model.	ISTPARK Dataset	As compared to Support Vector Machine, Random Forest and ARIMA models, this model improves accuracy. The accuracy rate is 99.57% which is dependent on capacity, time, density, day and holiday.

Singh et al. [21]	2022	Virtualization, Sustainable city, Blockchain, Deep LSTM, Energy efficiency	IoT and sensor devices	Proposed block chain enabled secure approach for smart parking in order to provide energy efficient solution in sustainable environment.
Awaisi et al. [22]	2022	Smart parking, IIoT, Deep reinforcement learning	Smart camera, cloud server, and fog nodes	Proposed deep reinforcement learning based solutions for industrial IoT (IIoT) based smart parking. Result shows that this approach correctly detect vacant places with minimum processing time.

3 Conclusion

With the increase in population and traffic congestion, there is a reduction in land, so smart parking becomes a major topic to work on, not only from research point of view but from economic point of view also. Drivers have to spent more time in finding the available parking space due to which there is more consumption of fuel and increase in environment pollution. In this paper, we have performed a literature survey of existing parking solutions. We have also explained the methodologies used in these studies and these methods solves the parking problems. After analyzing the previous work done, we found that there are some shortcomings in the existing studies. The main objective of this survey is to identify the research gaps that can help in building the new and efficient parking system.

1 Future Scope

We can extend the existing parking solutions to analyze the results in various cities means make the results more generalizable and integrating existing route planning applications to this system so that there is an ease to drivers in finding parking occupancy. Also, for building prediction models, we can use hybrid algorithms instead of machine learning and ensemble techniques. As there can be an increment in traffic density due to waiting in parking, exiting from parking, and searching vacant space in parking, so, we can analyze effect of parking system on traffic density and can introduce some method which helps in reducing traffic density.

References

[1] Fahim, A., Hasan, M., & Chowdhury, M. A. (2021). Smart parking systems: comprehensive review based on various aspects. *Heliyon*, *7*(5), e07050.

[2] Khalid, M., Wang, K., Aslam, N., Cao, Y., Ahmad, N., & Khan, M. K. (2021). From smart parking towards autonomous valet parking: A survey, challenges and future Works. *Journal of Network and Computer Applications*, *175*, 102935.

[3] Lin, T., Rivano, H., & Le Mouël, F. (2017). A survey of smart parking solutions. *IEEE Transactions on Intelligent Transportation Systems*, *18*(12), 3229-3253.

[4] Srikanth, S. V., Pramod, P. J., Dileep, K. P., Tapas, S., Patil, M. U., & Sarat, C. B. N. (2009, May). Design and implementation of a prototype smart parking (spark) system using wireless sensor networks. In *2009 International conference on advanced information networking and applications workshops* (pp. 401-406). IEEE.

[5] Delot, T., & Ilarri, S. (2020). Let my car alone: Parking strategies with social-distance preservation in the age of COVID-19. *Procedia Computer Science*, *177*, 143-150.

[6] Canli, H., & Toklu, S. (2021). Deep learning-based mobile application design for smart parking. *IEEE Access*, *9*, 61171-61183.

[7] Tang, C., Wei, X., Zhu, C., Chen, W., & Rodrigues, J. J. (2018). Towards smart parking based on fog computing. *IEEE access*, *6*, 70172-70185.

[8] Lin, J., Chen, S. Y., Chang, C. Y., & Chen, G. (2019). SPA: Smart parking algorithm based on driver behavior and parking traffic predictions. *IEEE Access*, *7*, 34275-34288.

[9] Misra, P., Vasan, A., Krishnan, B., Raghavan, V., & Sivasubramaniam, A. (2019). The future of smart parking systems with parking 4.0. *GetMobile: Mobile Computing and Communications*, *23*(1), 10-15.

[10] Singh, R., Dutta, C., Singhal, N., & Choudhury, T. (2020). An improved vehicle parking mechanism to reduce parking space searching time using firefly algorithm and feed forward back propagation method. *Procedia Computer Science*, *167*, 952-961.

[11] Mackey, A., Spachos, P., & Plataniotis, K. N. (2020). Smart parking system based on bluetooth low energy beacons with particle filtering. *IEEE Systems Journal*, *14*(3), 3371-3382.

[12] Provoost, J. C., Kamilaris, A., Wismans, L. J., Van Der Drift, S. J., & Van Keulen, M. (2020). Predicting parking occupancy via machine learning in the web of things. *Internet of Things*, *12*, 100301.

[13] Deshpande, S. (2016, January). M-parking: Vehicle parking guidance system using hierarchical Wireless Sensor Networks. In *2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)* (pp. 808-811). IEEE.

[14] Tekouabou, S. C. K., Cherif, W., & Silkan, H. (2020). Improving parking availability prediction in smart cities with IoT and ensemble-based model. *Journal of King Saud University-Computer and Information Sciences*.

[15] Zhang, N., Lu, X., Tian, C., Duan, Z., Sun, Z., & Zhang, T. (2020). P2P Network Based Smart Parking System Using Edge Computing. *Mobile Networks and Applications*, *25*(6), 2226-2239.

[16] Shin, J. H., & Jun, H. B. (2014). A study on smart parking guidance algorithm. *Transportation Research Part C: Emerging Technologies*, *44*, 299-317.

[17] Qadir, Z., Al-Turjman, F., Khan, M. A., & Nesimoglu, T. (2018, October). ZIGBEE based time and energy efficient smart parking system using IOT. In *2018 18th mediterranean microwave symposium (MMS)* (pp. 295-298). IEEE.

[18] Sadhukhan, P. (2017, September). An IoT-based E-parking system for smart cities. In *2017 International conference on advances in computing, communications and informatics (ICACCI)* (pp. 1062-1066). IEEE.

[19] Alharbi, A., Halikias, G., Yamin, M., Sen, A., & Ahmed, A. (2021). Web-based framework for smart parking system. *International Journal of Information Technology*, *13*(4), 1495-1502.

[20] Kumar, S. A., Nair, R. R., Kannan, E., Suresh, A., & Raj Anand, S. (2021). Intelligent Vehicle Parking System (IVPS) Using Wireless Sensor Networks. *Wireless Personal Communications*, 1-16.

[21] Singh, S. K., Pan, Y., & Park, J. H. (2022). Blockchain-enabled Secure Framework for Energy-Efficient Smart Parking in Sustainable City Environment. *Sustainable Cities and Society*, *76*, 103364.

[22] Awaisi, K. S., Abbas, A., Khattak, H. A., Ahmad, A., Ali, M., & Khalid, A. (2022). Deep reinforcement learning approach towards a smart parking architecture. *Cluster Computing*, 1-12.

Attention-based Model for Multi-modal sentiment recognition using Text-Image Pairs

Ananya Pandey
Biometric Research Laboratory, Department of Information Technology, Delhi Technological University
Bawana Road, Delhi-110042, India
ananyaphdit08@gmail.com

Dinesh Kumar Vishwakarma
Biometric Research Laboratory, Department of Information Technology, Delhi Technological University
Bawana Road, Delhi-110042, India
dvishwakarma@gmail.com

2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT) | 978-1-6654-9414-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICITIIT57246.2023.10068626

***Abstract*— Multi-modal sentiment recognition (MSR) is an emerging classification task that aims to categorize sentiment polarities for a given multi-modal dataset. The majority of work done in the past relied heavily on text-based information. However, in many scenarios, text alone is frequently insufficient to predict sentiment accurately; as a result, academics are more motivated to engage in the subject of MSR. In light of this, we proposed an attention-based model for MSR using image-text pairs of tweets. To effectively capture the vital information from both modalities, our approach combines BERT and ConvNet with CBAM (convolution block attention module) attention. The outcomes of our experimentations on the Twitter-17 dataset demonstrate that our method is capable of sentiment classification accuracy that is superior to that of competing approaches.**

Keywords—multi-modal sentiment recognition (MSR), Sentiment recognition (SR), Convolution block attention module (CBAM), BERT, Bidirectional LSTM

I. Introduction

Sentiment recognition (SR) has dramatically expanded with the development in the number of users and acceptance of online community channels due to a user-friendly environment. Due to the internet's global audience platform, everyone can now share their perspectives, thoughts, and opinions through writing, images, audio, emojis, and videos. As a result, all web platforms have become very influential sources of opinionated information. This opinionated information is crucial for evaluating someone's overall sentiment toward something; this is where MSR comes into play. Hence, predicting sentiment using more than one input data type is referred to as MSR Figure *1*. MSR can either be bimodal or trimodal. Bimodal analysis (BMA) is a subset of MSR in which the sentiment can be predicted using any of the two modalities (text + audio or image + text or audio + images). Contrarily, trimodal analysis (TMA) is a type of MSR in which the sentiment is forecast using three modalities (text, audio, and video).

In latest years, the MSR issue has garnered a growing amount of attention, and a number of research have been proposed to assist with this tough endeavour. To anticipate the polarities of sentiment, some research utilizes images and captions [1]–[3], while others use videos with subtitles [4]–[6]. The concept of aspect-based MSR is also gaining popularity these days since, in addition to predicting sentiment, it also provides some extra information about the category type, including whether it is positive, negative, or neutral [7]–[9]. Even though these research studies offer many solutions, more work is still needed to improve the solutions in terms of many metrics like accuracy, precision, recall, F1-score, etc. Hence, an efficient model is necessary to predict sentiment using multi-modal content.

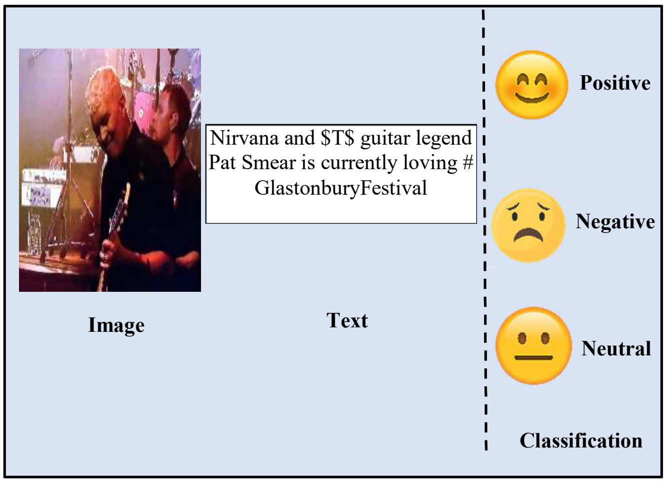

Nirvana and T guitar legend Pat Smear is currently loving # GlastonburyFestival
Image
Text
Positive
Negative
Neutral
Classification

Figure 1 Example of MSR

Even though the previous research works succeeded in aligning images and texts to some extent, much more research needs to be done in this area to improve the results. This inspired us to develop an attention-based model that combines BERT and ConvNet with CBAM [10] attention for extracting the essential data from both modalities.

To demonstrate the viability of our strategy, we empirically assess the model on one of the benchmark datasets, Twitter-17 [11]. Results indicate that our method produces superior outcomes. More crucially, our model can figure out how the text and image information relate to one another.

As a whole, the following are our key contributions:

- We propose an attention-based model for MSR, which is the combination of BERT and ConvNet with CBAM attention module, to efficiently combine information from text and image in a single task,
- We conduct in-depth analyses and experiments using one of the benchmark datasets, Twitter-17, to show how our model can successfully and robustly simulate the multi-modal representations of descriptive texts and images and produce ground-breaking outcomes for MSR.

Following is how the remaining manuscript is organized. Section II highlights the most current cutting-edge research in sentiment recognition, with an emphasis on multi-modal sentiment identification. Section III describes the proposed methodology framework. Section IV presents the research observations, and Section V concludes the research.

978-1-6654-9414-4/23/$31.00 ©2023 IEEE

II. Related Work

A. Text-based Sentiment recognition(TBSR)

TBSR is the process of predicting the polarity of sentiment for any script-based information. SR can be classified into three categories based on script-based data: document, sentence, and aspect-based. Document-based approach [12], [13] determine whether the entire assessment report is positive or negative. On the other hand, [14], the sentence-based analysis identifies if a particular sentence demonstrates a positive, negative, or neutral stance. Aspect-based study [15]–[17], as its name suggests, primarily concentrates on the type of aspect, which is a text-based assessment that refers to the traits or features. The term "fine-grained" refers to this kind of SR system, which indicates a thorough examination of text-based data. It tries to identify the sort of aspect first before categorizing feelings for that specific aspect. Transformers-based pre-trained language models have recently gained more popularity because of their ease of building, effective language representation, and comprehension abilities. For example,[18] used RoBERTa for the prediction of sentiment.

B. Visual Sentiment recognition(VSR)

There are several social media sites accessible nowadays where users may publish acoustic, image, textual, and video data to express their opinions. Not only have researchers examined textual data, but they have also created visuals for SR. As a result, a model that analyses emotions may also be given input in images. VSR is the term used for this kind of SR. Deep learning models have established their superiority for visual data throughout the history of computer vision [19], and they have been the main focus of all visual sentiment recognition research [20]–[22]. To do VSR, images are input into deep learning models. These models learn visual characteristics and calculate the sentiment polarity based on those data.

C. MSR

Discovering the emotion reflected in the multi-modal dataset is the goal of MSR. [11] utilized a BERT model to identify features in text samples, a pre-trained ResNet50 model to identify features in image samples, and then combined the two to determine sentiment. [7] also utilized the BERT model and a variant of ConvNet. [9], [23], [24] are some other recent publications in the field of MSR using image-text pairs.

III. Proposed Methodology

The task formulation, architecture for processing text and image-based data, and the attention mechanism have all been covered in the following subsections. The proposed model is described in Figure 2.

A. Task Formulation

For a given set of multi-modal dataset D, each sample d ∈ D consists of a sentence t, an image i, and a label l, that is d = {t, i, l}. t = (v1, v2, v3, ..., vn), in which n is the word count. There are three kinds of labels that l∈ {positive, negative, and neutral}. Given D, the objective of MSR is to train a sentiment classifier to correctly predict the sentiment labels for multi-modal samples that have not yet been observed.

B. Model to Process Text

Motivated by the recent emergence of a pre-trained transformer-based language model, we have utilized BERT [11] model to process text-based data for our dataset. The textual data from the dataset Twitter-17 is first pre-processed and cleaned to eliminate stop words and extraneous words from the sentences. The pre-trained BERT tokenizer transforms the cleaned text into a collection of tokens, which are nothing more than a set of numeric values. Tokenized text is then sent to a layer of Conv1D, followed by three layers of Bidirectional LSTM, for feature extraction. The text model uses the LeakyRelu activation function.

C. Model to Process Image

To extract image-based features, ConvNet with a CBAM attention module is utilized. The ConvNet consists of five layers of Conv2D followed by max-pooling, batch normalization, dropout, and an attention layer. The CBAM attention module has been used to concentrate on the crucial details while ignoring the unimportant ones.

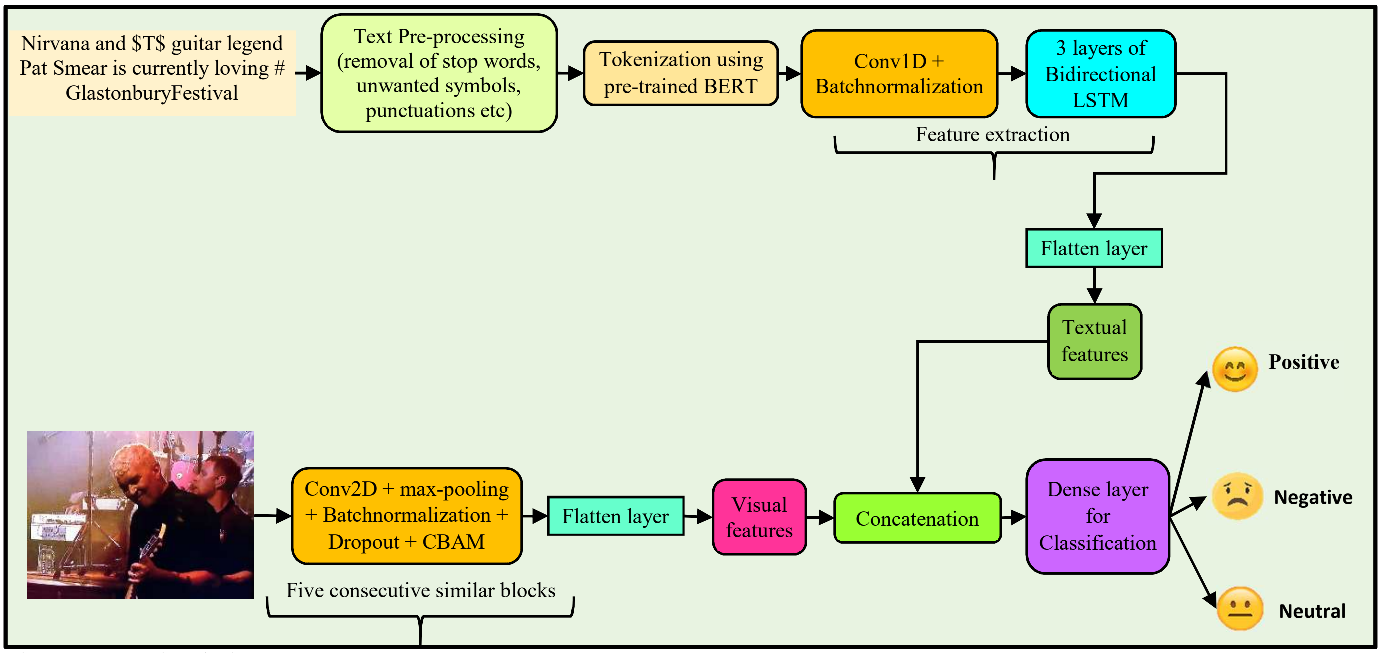

Nirvana and T guitar legend Pat Smear is currently loving # GlastonburyFestival
Text Pre-processing (removal of stop words, unwanted symbols, punctuations etc)
Tokenization using pre-trained BERT
Conv1D + Batchnormalization
3 layers of Bidirectional LSTM
Feature extraction
Flatten layer
Textual features
Conv2D + max-pooling + Batchnormalization + Dropout + CBAM
Five consecutive similar blocks
Flatten layer
Visual features
Concatenation
Dense layer for Classification
Positive
Negative
Neutral

Figure 2 Proposed method

D. CBAM Attention

It is better to employ attention modules since they concentrate on relevant details while discarding extraneous data. Therefore, for our model to effectively extract useful information from the visual data, we have integrated the CBAM Figure 3 [10] attention module into our ConvNet. The channel attention (CA) and spatial attention (SA) modules are two sub-modules that make up the CBAM module. CA concentrates on "what," which is significant when providing visual input. On the other hand, SA emphasizes "where" is that informative component. CA incorporates maximum and mean pooling outcomes in conjunction with a network connection. In contrast, SA's max and average pooling outputs are forwarded to the convolution layer after being pooled along the channel axis.

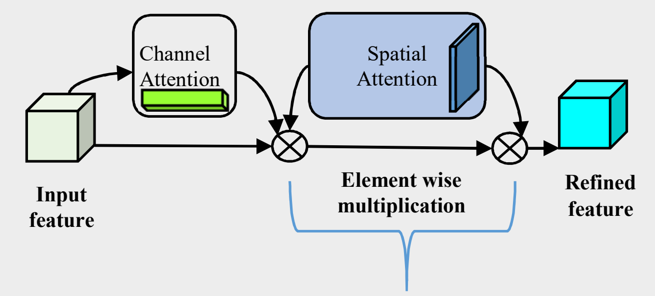


Figure 3 CBAM attention module [10]

E. Fusion Strategy

Textual features from the bidirectional LSTM are combined with features from the image model produced by ConvNet with the CBAM attention module. The sentiment polarity categorization as positive, negative, and neutral is then performed using the final dense layer on this combined feature vector.

IV. Experimental Results

A. Dataset

We use a benchmark dataset, Twitter-17, to assess our model for the MSR problem. Twitter-17 consists of tweets, including a caption and a picture. The goal is to determine if the sentiment is good, negative, or neutral. 20% of the 3562 total samples in the Twitter-17 dataset are chosen for testing.

B. Implementation Details

We use ConvNet with a CBAM attention module for image data and a BERT-based model for text data as our framework. We adjust the downstream tasks for 50 epochs and lock in all the hyperparameters after optimizing them on the training data. During this phase, the batch size is set to 32, and the default learning rate of the Adam optimizer is set to 0.001. Categorical cross-entropy loss function and Adam, as an optimizer, were both employed. All the models were implemented using Keras and TensorFlow framework. Experiments are carried out on a Persistence-M GPU.

C. Evaluation Metrics

Accuracy (A), Precision (P), Recall (R), and F1-score (F1) are the assessment metrics we use to assess the performance of our model for MSR tasks.

D. Main Results

Table 1 displays the results of our MSR approach on the Twitter-17 dataset. The experimental findings validate the efficacy of our framework and indicate that our model outperforms several multi-modal methods. It performs considerably better on Twitter 2017 than prior techniques, outperforming them on accuracy by 2.19 total percentage points. Figure 5 depicts the training accuracy, training loss, testing accuracy, and testing loss curves.

Table 1 Experimental results

		TWITTER-17 (FOR TEXT + IMAGE)				
REF.	YEAR	A	P	R	MACRO -F1	F1
[11]	2019	70.50	-	-	68.04	-
[25]	2021	67.77	-	-	-	65.32
[23]	2021	72.36	-	-	69.19	-
[7]	2022	71.14	-	-		69.16
[9]	2022	-	-	-	-	68.05
OURS		**73.33**	**87.6**	**88.4**	**-**	**88.05**

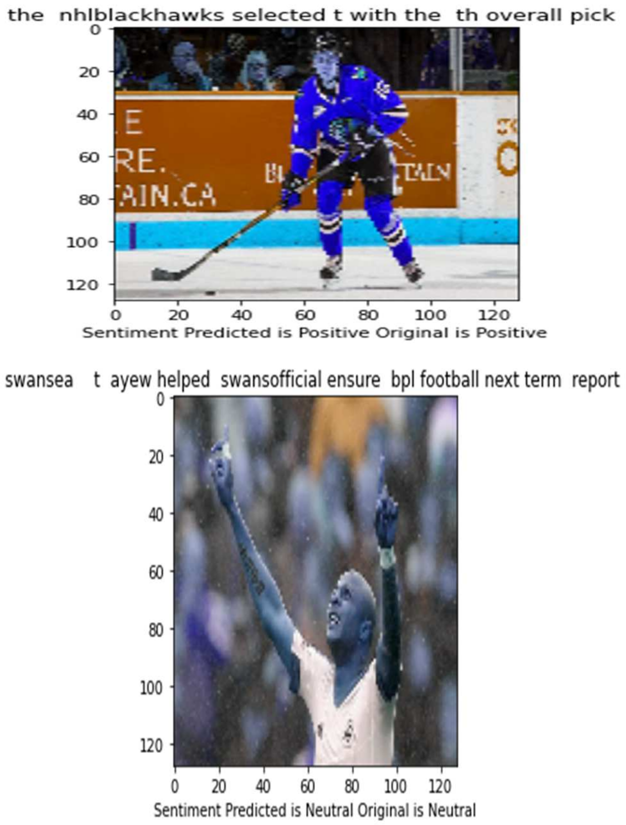

the nhlblackhawks selected t with the th overall pick
0
20
40
60
80
100
120
0 20 40 60 80 100 120
Sentiment Predicted is Positive Original is Positive
swansea t ayew helped swansofficial ensure bpl football next term report
0
20
40
60
80
100
120
0 20 40 60 80 100 120
Sentiment Predicted is Neutral Original is Neutral

Figure 4 Few samples with the original labels and the predicted labels

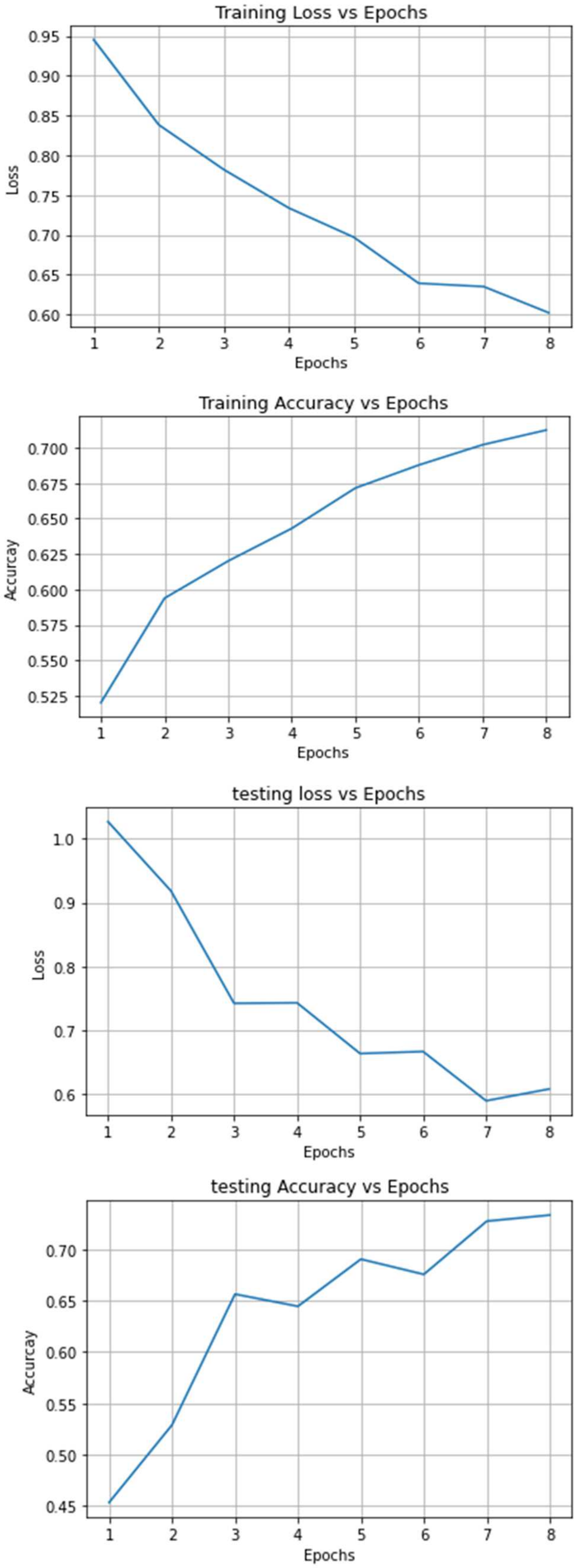


Figure 5 Curves shows training loss, training accuracy, testing loss and testing accuracy

V. CONCLUSION

We suggest an MSR approach in this work. Experimental findings demonstrate that, on the common benchmark Twitter-17, our suggested strategy typically performs better than the cutting-edge approaches. The suggested paradigm unifies two architectures: one for text and the other for images. In the future, we hope to: (1) apply our model to a few other benchmark datasets; and (2) expand this work to include aspect-based MSR, which not only predicts sentiment but also identifies the category to which that specific sentiment belongs.

REFERENCES

[1] J. Yu, J. Jiang, and R. Xia, "Entity-Sensitive Attention and Fusion Network for Entity-Level Multi-modal Sentiment Classification," *IEEE/ACM Trans Audio Speech Lang Process*, vol. 28, pp. 429–439, 2020, doi: 10.1109/TASLP.2019.2957872.

[2] N. Xu and W. Mao, "A residual merged neutral network for multi-modal sentiment recognition," in *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)(*, Mar. 2017, pp. 6–10. doi: 10.1109/ICBDA.2017.8078794.

[3] F. Huang, X. Zhang, Z. Zhao, J. Xu, and Z. Li, "Image–text sentiment analysis via deep multi-modal attentive fusion," *Knowl Based Syst*, vol. 167, 2019, doi: 10.1016/j.knosys.2019.01.019.

[4] K. Yang, H. Xu, and K. Gao, "CM-BERT: Cross-Modal BERT for Text-Audio Sentiment Analysis," in *Proceedings of the 28th ACM International Conference on Multimedia*, Oct. 2020, pp. 521–528. doi: 10.1145/3394171.3413690.

[5] M. Arjmand, M. J. Dousti, and H. Moradi, "TEASEL: A Transformer-Based Speech-Prefixed Language Model," Sep. 2021, [Online]. Available: http://arxiv.org/abs/2109.05522

[6] J.-B. Delbrouck, N. Tits, M. Brousmiche, and S. Dupont, "A Transformer-based joint-encoding for Emotion Recognition and Sentiment Analysis," in *Second Grand-Challenge and Workshop on Multimodal Language (Challenge-HML)*, 2020, pp. 1–7. doi: 10.18653/v1/2020.challengehml-1.1.

[7] J. Yu, K. Chen, and R. Xia, "Hierarchical Interactive Multimodal Transformer for Aspect-Based Multi-modal sentiment recognition," *IEEE Trans Affect Comput*, 2022, doi: 10.1109/TAFFC.2022.3171091.

[8] J. Zhou, J. Zhao, J. X. Huang, Q. V. Hu, and L. He, "MASAD: A large-scale dataset for multi-modal aspect-based sentiment analysis," *Neurocomputing*, vol. 455, pp. 47–58, Sep. 2021, doi: 10.1016/j.neucom.2021.05.040.

[9] Y. Ling, J. Yu, and R. Xia, "Vision-Language Pre-Training for Multi-modal Aspect-Based Sentiment Analysis," 2022, vol. 1, pp. 2149–2159. [Online]. Available: https://github.com/NUSTM/

[10] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11211 LNCS. doi: 10.1007/978-3-030-01234-2_1.

[11] J. Yu and J. Jiang, "Adapting BERT for target-oriented multi-modal sentiment classification," in *IJCAI International Joint Conference on Artificial Intelligence*, 2019, vol. 2019-August. doi: 10.24963/ijcai.2019/751.

[12] Y. Zhang, J. Wang, and X. Zhang, "Conciseness is better: Recurrent attention LSTM model for document-level sentiment analysis," *Neurocomputing*, vol. 462, pp. 101–112, Oct. 2021, doi: 10.1016/j.neucom.2021.07.072.

[13] S. Liu and I. Lee, "Sequence encoding incorporated CNN model for Email document sentiment classification," *Appl Soft Comput*, vol. 102, p. 107104, Apr. 2021, doi: 10.1016/j.asoc.2021.107104.

[14] C. Yang, X. Chen, L. Liu, and P. Sweetser, "Leveraging semantic features for recommendation: Sentence-level emotion analysis," *Inf Process Manag*, vol. 58, no. 3, p. 102543, May 2021, doi: 10.1016/j.ipm.2021.102543.

[15] H. Wu, Z. Zhang, S. Shi, Q. Wu, and H. Song, "Phrase dependency relational graph attention network for Aspect-based Sentiment Analysis," *Knowl Based Syst*, vol. 236, p. 107736, Jan. 2022, doi: 10.1016/j.knosys.2021.107736.

[16] B. Liang, H. Su, L. Gui, E. Cambria, and R. Xu, "Aspect-based sentiment analysis via affective knowledge enhanced graph

convolutional networks," *Knowl Based Syst*, vol. 235, p. 107643, Jan. 2022, doi: 10.1016/j.knosys.2021.107643.

[17] Y.-C. Chang, C.-H. Ku, and D.-D. le Nguyen, "Predicting aspect-based sentiment using deep learning and information visualization: The impact of COVID-19 on the airline industry," *Information & Management*, vol. 59, no. 2, p. 103587, Mar. 2022, doi: 10.1016/j.im.2021.103587.

[18] J. Dai, H. Yan, T. Sun, P. Liu, and X. Qiu, "Does syntax matter? A strong baseline for Aspect-based Sentiment Analysis with RoBERTa," in *NAACL-HLT 2021 - 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference*, 2021. doi: 10.18653/v1/2021.naacl-main.146.

[19] J. Joo, W. Li, F. F. Steen, and S.-C. Zhu, "Visual Persuasion: Inferring Communicative Intents of Images," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 216–223. doi: 10.1109/CVPR.2014.35.

[20] S. Ruan, K. Zhang, L. Wu, T. Xu, Q. Liu, and E. Chen, "Color Enhanced Cross Correlation Net for Image Sentiment Analysis," *IEEE Trans Multimedia*, pp. 1–1, 2021, doi: 10.1109/TMM.2021.3118208.

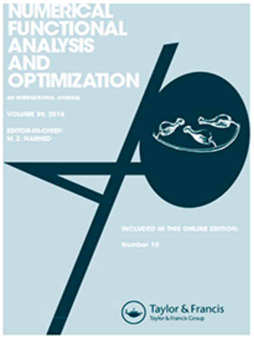
[21] S. Jindal and S. Singh, "Image sentiment analysis using deep convolutional neural networks with domain specific fine tuning," in *2015 International Conference on Information Processing (ICIP)*, Dec. 2015, pp. 447–451. doi: 10.1109/INFOP.2015.7489424.

[22] L. Wu, S. Liu, M. Jian, J. Luo, X. Zhang, and M. Qi, "Reducing noisy labels in weakly labeled data for visual sentiment analysis," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sep. 2017, pp. 1322–1326. doi: 10.1109/ICIP.2017.8296496.

[23] Z. Zhang, Z. Wang, X. Li, N. Liu, B. Guo, and Z. Yu, "ModalNet: an aspect-level sentiment classification model by exploring multimodal data with fusion discriminant attentional network," *World Wide Web*, vol. 24, no. 6, 2021, doi: 10.1007/s11280-021-00955-7.

[24] Z. Khan and Y. Fu, "Exploiting BERT for Multi-modal Target Sentiment Classification through Input Space Translation," in *MM 2021 - Proceedings of the 29th ACM International Conference on Multimedia*, Oct. 2021, pp. 3034–3042. doi: 10.1145/3474085.3475692.

[25] D. Gu *et al.*, "Targeted Aspect-Based Multi-modal sentiment recognition: An Attention Capsule Extraction and Multi-Head Fusion Network," *IEEE Access*, vol. 9, pp. 157329–157336, 2021, doi: 10.1109/ACCESS.2021.3126782.



Taylor & Francis
Taylor & Francis Group

Numerical Functional Analysis and Optimization

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lnfa20

λ-Bernstein Operators Based on Pólya Distribution

Km. Lipi & Naokant Deo

NUMERICAL FUNCTIONAL ANALYSIS AND OPTIMIZATION
https://doi.org/10.1080/01630563.2023.2185896

Taylor & Francis
Taylor & Francis Group

λ-Bernstein Operators Based on Pólya Distribution

Km. Lipi and Naokant Deo

Department of Applied Mathematics, Delhi Technological University, Delhi, India

ABSTRACT

In this manuscript, we propose a Pólya distribution-based generalization of λ-Bernstein operators. We establish some fundamental results for convergence as well as order of approximation of the proposed operators. We present theoretical result and graph to demonstrate the proposed operator's intriguing ability to interpolate at the interval's end points. In order to illustrate the convergence of proposed operators as well as the effect of changing the parameter "μ," we provide a variety of results and graphs as our paper's conclusion.

ARTICLE HISTORY
Received 9 December 2022
Revised 25 February 2023
Accepted 25 February 2023

KEYWORDS
λ-Bernstein operators; Lipschitz continuous functions; Pólya-Eggenberger distribution

MATHEMATICS SUBJECT CLASSIFICATION
MSC 41A25; MSC 41A36

1. Introduction

The original Pólya-Eggenberger urn model, often known as the Pólya urn, was created in 1923 by Eggenberger and Pólya [1] to explore phenomena like the transmission of infectious diseases. The Pólya-Eggenberger urn model consists of M white balls and N black balls in one of its most basic forms. A ball is picked at random and then replaced with O other balls of the same color. This process is carried out n times, and then the probability of drawing s $(s = 1, 2, \ldots, n)$ white ball is

$$\Pr[X = s] = \binom{n}{s} \frac{M(M+O)\ldots[M+(s-1)O]N(N+O)\ldots[N+(n-s-1)O]}{(M+N)(M+N+O)\ldots[M+N+(n-1)O]}. \tag{1.1}$$

The distribution described above is referred to as the Pólya-Eggenberger distribution with parameters $(n; M; N; O)$ and includes hypergeometric and binomial distribution as special cases.

Stancu [2] constructed a sequence of linear positive operators using the Pólya-Eggenberger distribution as

CONTACT Km. Lipi ✉ chaudhary.lipi123@gmail.com Department of Applied Mathematics, Delhi Technological University, Bawana Road, Delhi 110042, India.

$$S_n^{\langle\mu\rangle}(f;x)=\sum_{k=0}^{n}p_{n,k}^{\langle\mu\rangle}(x)f\left(\frac{k}{n}\right), \tag{1.2}$$

where

$$p_{n,k}^{\langle\mu\rangle}(x)=\binom{n}{k}\frac{\prod\limits_{i=0}^{k-1}(x+i\mu)\prod\limits_{i=0}^{n-k-1}(1-x+i\mu)}{\prod\limits_{i=0}^{n-1}(1+i\mu)}$$

and μ is a non-negative parameter that may only be dependent on the natural number n. When $\mu=0,$ operators (1.2) reduce into the classical Bernstein operators [3].

The distribution of the number P of drawings required to obtain n white balls from an urn containing M white balls and N black balls is known as the inverse Pólya-Eggenberger distribution, and it is defined as

$$\Pr(P=n+s)=\binom{n+s-1}{s}\frac{M(M+O)...[M+(n-1)O]N(N+O)...[N+(s-1)O]}{(M+N)(M+N+O)...[M+N+(n+s-1)O]}, \tag{1.3}$$

for $s\in\mathbb{N}\cup\{0\}$. We direct the readers to [4] in order to provide additional information regarding distributions (1.1) and (1.3).

For a real valued bounded function on $[0,\infty)$ with $0\leqslant\mu=\mu(n)\to 0$ as $n\to\infty,$ Stancu [5] provided the generalization of the Baskakov operators using the inverse Pólya-Eggenberger distribution. For the case $\mu=0,$ these operators reduce into the classical Baskakov operators [6].

Razi [7] developed Bernstein Kantorovich operators based on the Pólya-Eggenberger distribution in 1989 and investigated the rate of convergence and degree of approximation for these operators. Büyükyazici [8] introduced Chlodowsky type generalization of Stancu polynomials (also known as Stancu Chlodowsky polynomials) and presented theorems on weighted approximation of functions on the interval $[0,\infty)$. Agrawal et al. [9] introduced the Pólya and Bernstein basis function-based Bézier variant of summation integral type operators. Deo et al. [10] introduced inverse Pólya based Baskakov Kantorovich operators along with its asymptotic formula. The reader is directed to [11–16] for additional research in this area.

Depending on the parameter λ, Cai et al. [17] proposed and took into consideration a new generalization of Bernstein polynomials known as λ-Bernstein operators. When $\lambda=0,$ these λ-Bernstein operators reduce into the well-known Bernstein operators [3]. Acu et al. [18] defined a Kantorovich form of λ-Bernstein operators and demonstrated how this generalization enhances convergence rate over the classical Kantorovich operators. In order to approximate a function on [0, 1] as well as on its subinterval, Rahman et al. [19] introduced the Kantorovich form of λ-Bernstein operators with shifted knots and demonstrated that these operators approximate the function more accurately than classical Bernstein Kantorovich operators and λ-Bernstein Kantorovich operators. Cai [20] provided the Bézier form of λ-Bernstein Kantorovich operators and derived asymptotic estimate for absolutely continuous function by combining the Bojanic-Cheng decomposition method with a few analysis techniques. Cai and Zhou [21] considered the GBS of the bivariate tensor product of λ -Bernstein Kantorovich operators and established approximation properties of these operators for both B-continuous and B-differentiable functions. Acu et al. [22] considered and investigated a generalization of U_n^{ρ} operators based on λ-Bernstein operators. The reader is instructed to read [23–26] for further information on this topic.

In this paper, the generalization of λ-Bernstein operators [17] based on Pólya distribution is presented in the following manner:

$$\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x)=\sum_{k=0}^{n}\hat{p}_{n,k}^{\langle\lambda,\mu\rangle}(x)f\left(\frac{k}{n}\right), \tag{1.4}$$

where $f\in C[0,1]$, $\lambda\in[-1,1]$, $\mu=\mu(n)\to 0$ as $n\to\infty$ and $\hat{p}_{n,k}^{\langle\lambda,\mu\rangle}(x), k=0,1,\ldots,n$ are defined below:

$$\begin{cases}\hat{p}_{n,0}^{\langle\lambda,\mu\rangle}(x)=p_{n,0}^{\langle\mu\rangle}(x)-\dfrac{\lambda}{n+1}p_{n+1,1}^{\langle\mu\rangle}(x),\\ \hat{p}_{n,k}^{\langle\lambda,\mu\rangle}(x)=p_{n,k}^{\langle\mu\rangle}(x)+\lambda\left(\dfrac{n-2k+1}{n^2-1}p_{n+1,k}^{\langle\mu\rangle}(x)-\dfrac{n-2k-1}{n^2-1}p_{n+1,k+1}^{\langle\mu\rangle}(x)\right), 1\leqslant k\leqslant n-1,\\ \hat{p}_{n,n}^{\langle\lambda,\mu\rangle}(x)=p_{n,n}^{\langle\mu\rangle}(x)-\dfrac{\lambda}{n+1}p_{n+1,n}^{\langle\mu\rangle}(x).\end{cases}$$

Special cases:

1. For $\lambda=0$ and $\mu=0$, proposed operators $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ transform into well known Bernstein operators [3].
2. For $\lambda=0$ and $\mu\neq 0$, these operators $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ reduces to operators (1.2).
3. For $\lambda\neq 0$ and $\mu=0$, operators $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ includes λ-Bernstein operators [17].

This paper is divided into three key sections. Basic results that are relevant for establishing key theorems are covered in the first section. In this section, we also give a theorem and graphical illustrations in support of the proposed operator's interpolation behavior. We have demonstrated a few results in the second section for convergence rate of the proposed operators. The final section makes use of several Mathematica-derived graphs to validate the approximation behavior of the operators $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$.

2. Preliminaries

Lemma 2.1. *The following equalities hold for the proposed operator* $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ *described by Equation (1.4):*

$$\mathscr{P}_n^{\langle\lambda,\mu\rangle}(1;x)=1,$$

$$\mathscr{P}_n^{\langle\lambda,\mu\rangle}(t;x)=x+\lambda\left(\frac{1-2x}{n(n-1)}+\frac{\prod_{i=0}^{n}(x+i\mu)-\prod_{i=0}^{n}(1-x+i\mu)}{n(n-1)\prod_{i=0}^{n}(1+i\mu)}\right),$$

$$\begin{aligned}\mathscr{P}_n^{\langle\lambda,\mu\rangle}(t^2;x)&=\frac{x^2}{\mu+1}+\frac{x(1+\mu n-x)}{(\mu+1)n}\\&+\lambda\left(\frac{2(1-\mu)x-4x^2}{(\mu+1)n(n-1)}-\frac{1}{n^2(n-1)}+\frac{(1+2n)\prod_{i=0}^{n}(x+i\mu)+\prod_{i=0}^{n}(1-x+i\mu)}{n^2(n-1)\prod_{i=0}^{n}(1+i\mu)}\right).\end{aligned}$$

Proof. Form (1.4), it is easy to prove $\mathscr{P}_n^{\langle\lambda,\mu\rangle}(1;x)=1$. Next,

$$\begin{aligned}&\mathscr{P}_n^{\langle\lambda,\mu\rangle}(t;x)\\&=\sum_{k=0}^{n}\hat{p}_{n,k}^{\langle\lambda,\mu\rangle}(x)\frac{k}{n}\\&=\sum_{k=1}^{n-1}\left\{p_{n,k}^{\langle\mu\rangle}(x)+\lambda\left[\frac{n-2k+1}{n^2-1}p_{n+1,k}^{\langle\mu\rangle}(x)-\frac{n-2k-1}{n^2-1}p_{n+1,k+1}^{\langle\mu\rangle}(x)\right]\right\}\frac{k}{n}\\&\quad+p_{n,n}^{\langle\mu\rangle}(x)-\frac{\lambda}{n+1}p_{n+1,n}^{\langle\mu\rangle}(x)\\&=\sum_{k=0}^{n}p_{n,k}^{\langle\mu\rangle}(x)\frac{k}{n}+\lambda\left[\sum_{k=0}^{n}p_{n+1,k}^{\langle\mu\rangle}(x)\frac{n-2k+1}{n^2-1}\frac{k}{n}-\sum_{k=1}^{n-1}p_{n+1,k+1}^{\langle\mu\rangle}(x)\frac{n-2k-1}{n^2-1}\frac{k}{n}\right]\\&=\sum_{k=0}^{n}p_{n,k}^{\langle\mu\rangle}(x)\frac{k}{n}+\lambda\left[\frac{1}{n-1}\sum_{k=0}^{n}p_{n+1,k}^{\langle\mu\rangle}(x)\frac{k}{n}-\frac{2}{n^2-1}\sum_{k=0}^{n}p_{n+1,k}^{\langle\mu\rangle}(x)\frac{k^2}{n}\right.\\&\quad\left.-\frac{1}{n+1}\sum_{k=1}^{n-1}p_{n+1,k+1}^{\langle\mu\rangle}(x)\frac{k}{n}+\frac{2}{n^2-1}\sum_{k=1}^{n-1}p_{n+1,k+1}^{\langle\mu\rangle}(x)\frac{k^2}{n}\right]\end{aligned}$$

$$\begin{aligned}
&= \sum_{k=0}^{n} p_{n,k}^{\langle\mu\rangle}(x)\frac{k}{n} + \lambda\Bigg[\frac{1}{n(n+1)}\sum_{k=0}^{n} p_{n+1,k}^{\langle\mu\rangle}(x)k - \frac{2}{n(n^2-1)}\sum_{k=0}^{n} p_{n+1,k}^{\langle\mu\rangle}(x)k(k-1)\\
&- \frac{1}{n(n-1)}\left(\sum_{k=1}^{n-1} p_{n+1,k+1}^{\langle\mu\rangle}(x)(k+1) - \sum_{k=1}^{n-1} p_{n+1,k+1}^{\langle\mu\rangle}(x)\right)\\
&\quad + \frac{2}{n(n^2-1)}\sum_{k=1}^{n-1} p_{n+1,k+1}^{\langle\mu\rangle}(x)k(k+1)\Bigg]
\end{aligned} \tag{2.1}$$

It is easy to derive the following equalities:

$$\sum_{k=0}^{n} p_{n+1,k}^{\langle\mu\rangle}(x)k = (n+1)\left(x - \frac{\prod\limits_{i=0}^{n}(x+\mu i)}{\prod\limits_{i=0}^{n}(1+\mu i)}\right),$$

$$\sum_{k=0}^{n} p_{n+1,k}^{\langle\mu\rangle}(x)k(k-1) = n(n+1)\left(\frac{x(x+\mu)}{1+\mu} - \frac{\prod\limits_{i=0}^{n}(x+\mu i)}{\prod\limits_{i=0}^{n}(1+\mu i)}\right),$$

$$\sum_{k=1}^{n-1} p_{n+1,k+1}^{\langle\mu\rangle}(x) = 1 - \frac{\prod\limits_{i=0}^{n}(1-x+\mu i)}{\prod\limits_{i=0}^{n}(1+\mu i)} - \frac{(n+1)x\prod\limits_{i=0}^{n-1}(1-x+\mu i)}{\prod\limits_{i=0}^{n}(1+\mu i)} - \frac{\prod\limits_{i=0}^{n}(x+\mu i)}{\prod\limits_{i=0}^{n}(1+\mu i)},$$

$$\sum_{k=1}^{n-1} p_{n+1,k+1}^{\langle\mu\rangle}(x)(k+1) = (n+1)\left(x - \frac{x\prod\limits_{i=0}^{n-1}(1-x+\mu i)}{\prod\limits_{i=0}^{n}(1+\mu i)} - \frac{\prod\limits_{i=0}^{n}(x+\mu i)}{\prod\limits_{i=0}^{n}(1+\mu i)}\right),$$

$$\sum_{k=1}^{n-1} p_{n+1,k+1}^{\langle\mu\rangle}(x)(k+1)k = n(n+1)\left(\frac{x(x+\mu)}{1+\mu} - \frac{\prod\limits_{i=0}^{n}(x+\mu i)}{\prod\limits_{i=0}^{n}(1+\mu i)}\right).$$

Using these equalities in Equation (2.1), we get the value of $\mathscr{P}_n^{\langle\lambda,\mu\rangle}(t;x)$. We can also determine the value of $\mathscr{P}_n^{\langle\lambda,\mu\rangle}(t^2;x)$ in a similar manner. ■

Lemma 2.2. *For $x \in [0,1]$, $\lambda \in [-1,1]$, $\mu = \mu(n) \to 0$ as $n \to \infty$ and $\lim\limits_{n\to\infty} n\mu(n) = l \in \mathbb{R}$, we have*

$$\mathscr{P}_n^{\langle\lambda,\mu\rangle}((t-x);x)=\lambda\left(\frac{1-2x}{n(n-1)}+\frac{\prod\limits_{i=0}^{n}(x+i\mu)-\prod\limits_{i=0}^{n}(1-x+i\mu)}{n(n-1)\prod\limits_{i=0}^{n}(1+i\mu)}\right),$$

$$\mathscr{P}_n^{\langle\lambda,\mu\rangle}\left((t-x)^2;x\right)=\frac{(1+\mu n)(1-x)x}{(\mu+1)n}$$
$$+\lambda\left(\frac{4\mu x(x-1)}{(\mu+1)n(n-1)}-\frac{1}{n^2(n-1)}+\frac{(1+2n(1-x))\prod\limits_{i=0}^{n}(x+i\mu)+(1+2nx)\prod\limits_{i=0}^{n}(1-x+i\mu)}{n^2(n-1)\prod\limits_{i=0}^{n}(1+i\mu)}\right).$$

Furthermore,

$$\lim_{n\to\infty}\mathscr{P}_n^{\langle\lambda,\mu\rangle}((t-x);x)=0,$$

$$\lim_{n\to\infty}n\mathscr{P}_n^{\langle\lambda,\mu\rangle}\left((t-x)^2;x\right)=(l+1)(1-x)x.$$

Proof. By substituting the values from Lemma 2.1, we can easily prove this Lemma.

Throughout the paper, let us define $\varphi_{n,1}^{\langle\lambda,\mu\rangle}(x)=\mathscr{P}_n^{\langle\lambda,\mu\rangle}(t;x)$, $\varphi_{n,2}^{\langle\lambda,\mu\rangle}(x)=\mathscr{P}_n^{\langle\lambda,\mu\rangle}(t^2;x)$, $\delta_{n,1}^{\langle\lambda,\mu\rangle}(x)=\mathscr{P}_n^{\langle\lambda,\mu\rangle}((t-x);x)$, and $\delta_{n,2}^{\langle\lambda,\mu\rangle}(x)=\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ $((t-x)^2;\,x)$.

Remark 2.1. For $\lambda\in[-1,1]$, $\mu=\mu(n)\to0$ as $n\to\infty$ and $x\in[0,1]$, the proposed operators $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ possess the endpoint interpolation property, that is,

$$\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;0)=f(0),\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;1)=f(1).$$

We can establish the proof using the definition of $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ and the fact that

$$\hat{p}_{n,k}^{\langle\lambda,\mu\rangle}(x)=\begin{cases}0, & (k\neq0)\\ 1, & (k=0)\end{cases}\ \hat{p}_{n,k}^{\langle\lambda,\mu\rangle}(x)=\begin{cases}0, & (k\neq n)\\ 1, & (k=n).\end{cases}$$

Example 2.1. Figure 1 displays the graphs of $\hat{p}_{3,k}^{\langle\lambda,\mu\rangle}(x)$ for the values of $\lambda=1,0,$ and -1. Figure 2 displays the corresponding $\mathscr{P}_3^{\langle\lambda,\mu\rangle}$ when $f(x)=\left(x-\frac{1}{4}\right)\sin\left(\frac{5\pi x}{2}\right)+\frac{2}{5}$ with $\mu=\mu(n)=\frac{1}{\sqrt{2\pi n}}\left(\frac{e}{n}\right)^n$. The graphs make it evident

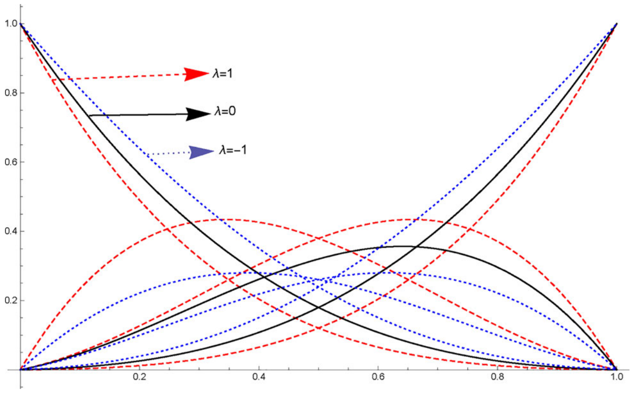


Figure 1. The graph $\hat{p}_{n,k}^{\langle\lambda,\mu\rangle}(x)$ of with different value of λ.

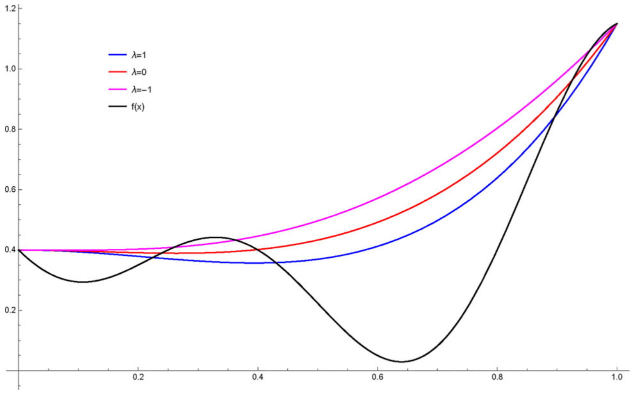


Figure 2. Convergence of $\mathscr{P}_3^{\langle -1,\mu\rangle}$ (magenta), $\mathscr{P}_3^{\langle 0,\mu\rangle}$ (red) and $\mathscr{P}_3^{\langle 1,\mu\rangle}$ (blue) with $\mu=\mu(n)=\frac{1}{\sqrt{2\pi n}}\left(\frac{e}{n}\right)^n$ to $f(x)=\left(x-\frac{1}{4}\right)\sin\left(\frac{5\pi x}{2}\right)+\frac{2}{5}$ (black).

that $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ interpolates the end points of the interval $[0, 1]$, which is based on the interpolation property of $\hat{p}_{n,k}^{\langle\lambda,\mu\rangle}(x)$.

3. Main results

Let

$$C^2[0,1]=\left\{f\in C[0,1]:f''\in C[0,1]\right\},$$

and the norm of the space $C^2[0,1]$ is defined by

$$\|f\|_{C^2[0,1]}=\|f\|+\|f'\|+\|f''\|.$$

According to Ditzian and Totik [27], for absolute constant $C>0,$ the relationship stated below is valid for $K(f;\delta)=\inf_{g\in C^2[0,1]}\left\{\|f-g\|+\delta\|g''\|\right\}$ and

$$\omega_2(f;\delta) = \sup_{0<t\leqslant\delta}\left\{|f(x+t)-2f(x)+f(x-t)| : x, x\pm t \in [0,1]\right\},$$
$$K(f;\delta) \leqslant C\omega_2\left(f;\sqrt{\delta}\right) \tag{3.1}$$

for any function $f \in C[0,1]$ and $\delta > 0$. Where $\omega_2(f;\delta)$ and $K(f;\delta)$ is referred to as second order modulus of continuity and Peetre's K-functional [27] respectively and

$$\omega(f;\delta) = \sup_{0<t\leqslant\delta}\left\{|f(x+t)-f(x)| : x, x+t \in [0,1]\right\}$$

is called first order or usual modulus of continuity. Additionally, $\omega(f;\delta)$ meets the following characteristics:

1. $|f(y)-f(x)| \leqslant \omega(f;|y-x|)$ for any $x \neq y \in [0,1]$,
2. f is uniformly continuous $\iff \lim_{\delta\to 0}\omega(f;\delta) = 0$,
3. $\omega(f;\delta)$ is monotonically increasing function,
4. $\omega(f;\lambda\delta) \leqslant (1+\lambda)\omega(f;\delta)$, for any $\lambda > 0$.

The smoothness characteristics of the function determine the degree of approximation of positive linear operators, and suitable tools for determining the smoothness of functions are represented by the moduli of continuity of various types. Our subsequent theorems determine the degree of approximation for our proposed operators $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ in terms of usual and second order modulus of continuity.

Theorem 3.1. *Let $\lambda \in [-1,1]$ and $\mu = \mu(n) \to 0$ as $n \to \infty$, then the inequality*

$$\left|\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x) - f(x)\right| \leqslant \delta_{n,1}^{\langle\lambda,\mu\rangle}(x)|f'(x)| + 2\sqrt{\delta_{n,2}^{\langle\lambda,\mu\rangle}(x)}\,\omega\left(f';\sqrt{\delta_{n,2}^{\langle\lambda,\mu\rangle}(x)}\right)$$

holds for $f \in C^1[0,1]$.

Proof. For $f \in C^1[0,1]$ and $x, t \in [0,1]$, we have

$$f(t) - f(x) = (t-x)f'(x) + \int_x^t (f'(y) - f'(x))dy.$$

Applying $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ on both sides of above mentioned relation, we get

$$\mathscr{P}_n^{\langle\lambda,\mu\rangle}\left(f(t)-f(x);x\right) = \mathscr{P}_n^{\langle\lambda,\mu\rangle}((t-x);x)f'(x) + \mathscr{P}_n^{\langle\lambda,\mu\rangle}\left(\int_x^t (f'(y)-f'(x))dy;x\right).$$

Using the property (1) and (4) of modulus of continuity, with a few manipulations, we have the relation

$$|f(t)-f(x)| \leqslant \left(1+\frac{|t-x|}{\delta}\right)\omega(f;\delta), \delta > 0,$$

this implies

$$\left|\int_x^t (f'(y)-f'(x))dy\right| \leqslant \left[|t-x| + \frac{\left|(t-x)^2\right|}{\delta}\right]\omega\big(f';\delta\big).$$

Therefore,

$$\begin{aligned}\left|\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x)-f(x)\right| &\leqslant \left|\mathscr{P}_n^{\langle\lambda,\mu\rangle}((t-x);x)\right||f'(x)|\\ &+\left\{\frac{1}{\delta}\mathscr{P}_n^{\langle\lambda,\mu\rangle}\big((t-x)^2;x\big)+\mathscr{P}_n^{\langle\lambda,\mu\rangle}(|t-x|;x)\right\}\omega\big(f';\delta\big).\end{aligned}$$

Using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}|\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x)-f(x)| &\leqslant |\mathscr{P}_n^{\langle\lambda,\mu\rangle}((t-x);x)||f'(x)|\\ &+\sqrt{\mathscr{P}_n^{\langle\lambda,\mu\rangle}((t-x)^2;x)}\left\{\frac{1}{\delta}\sqrt{\mathscr{P}_n^{\langle\lambda,\mu\rangle}((t-x)^2;x)}+1\right\}\omega(f';\delta)\\ &\leqslant \delta_{n,1}^{\langle\lambda,\mu\rangle}(x)|f'(x)| + \sqrt{\delta_{n,2}^{\langle\lambda,\mu\rangle}(x)}\left\{\frac{1}{\delta}\sqrt{\delta_{n,2}^{\langle\lambda,\mu\rangle}(x)}+1\right\}\omega(f';\delta).\end{aligned}$$

Choosing $\delta = \sqrt{\delta_{n,2}^{\langle\lambda,\mu\rangle}(x)}$, we find the desired inequality. ■

Theorem 3.2. *Let $\lambda \in [-1,1]$ and $\mu = \mu(n) \to 0$ as $n \to \infty$, then the inequality*

$$\left|\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x)-f(x)\right| \leqslant 2\omega\left(f;\sqrt{\delta_{n,2}^{\langle\lambda,\mu\rangle}(x)}\right)$$

holds for $f \in C[0,1]$.

Proof. For any $t, x \in [a,b]$, using the following property of modulus of continuity, we get

$$|f(t)-f(x)| \leqslant \left(1+\frac{(t-x)^2}{\delta^2}\right)\omega(f;\delta).$$

Applying $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ on both sides of above relation, we get

$$\begin{aligned}\left|\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x)-f(x)\right| &\leqslant \mathscr{P}_n^{\langle\lambda,\mu\rangle}(|f(t)-f(x)|;x)\\ &\leqslant \left(1+\frac{\mathscr{P}_n^{\langle\lambda,\mu\rangle}\left((t-x)^2;x\right)}{\delta^2}\right)\omega(f;\delta).\end{aligned}$$

Choosing $\delta^2=\delta_{n,2}^{\langle\lambda,\mu\rangle}(x)=\mathscr{P}_n^{\langle\lambda,\mu\rangle}((t-x)^2;x)$, we obtain the desired result. ■

Theorem 3.3. *For $\lambda\in[-1,1]$ and $\mu=\mu(n)\to 0$ as $n\to\infty$, then the inequality*

$$\left|\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x)-f(x)\right|\le C\omega_2\left(f;\frac{1}{2}\sqrt{\Upsilon_n^{\langle\lambda,\mu\rangle}(x)}\right)+\omega\left(f;\delta_{n,1}^{\langle\lambda,\mu\rangle}(x)\right)$$

holds for $f\in C[0,1]$, $\Upsilon_n^{\langle\lambda,\mu\rangle}(x)=\delta_{n,2}^{\langle\lambda,\mu\rangle}(x)+(\delta_{n,1}^{\langle\lambda,\mu\rangle}(x))^2$ and absolute constant C.

Proof. Consider the operators $\mathbb{P}_n^{\langle\lambda,\mu\rangle}$ defined by

$$\mathbb{P}_n^{\langle\lambda,\mu\rangle}(f;x)=\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x)-f\left(\varphi_{n,1}^{\langle\lambda,\mu\rangle}(x)\right)+f(x). \tag{3.2}$$

Due to Lemma 2.1 and the fact that these operators are linear in nature, it is obvious that

$$\begin{aligned}\mathbb{P}_n^{\langle\lambda,\mu\rangle}(1;x)&=\mathscr{P}_n^{\langle\lambda,\mu\rangle}(1;x)=1,\\ \mathbb{P}_n^{\langle\lambda,\mu\rangle}(t;x)&=\varphi_{n,1}^{\langle\lambda,\mu\rangle}(x)+x-\varphi_{n,1}^{\langle\lambda,\mu\rangle}(x)=x.\end{aligned}$$

For $g\in C^2[0,1]$, consider the Taylor's formula

$$g(t)=g(x)+(t-x)g'(x)+\int_x^t(t-w)g''(w)dw.$$

Applying $\mathbb{P}_n^{\langle\lambda,\mu\rangle}$ on both sides of above equality and using $\mathbb{P}_n^{\langle\lambda,\mu\rangle}(1;x)=1$, we get

$$\begin{aligned}\mathbb{P}_n^{\langle\lambda,\mu\rangle}(g;x)&=g(x)+\mathbb{P}_n^{\langle\lambda,\mu\rangle}((t-x);x)g'(x)+\mathbb{P}_n^{\langle\lambda,\mu\rangle}\left(\int_x^t(t-w)g''(w)dw;x\right)\\ &=g(x)+\mathscr{P}_n^{\langle\lambda,\mu\rangle}\left(\int_x^t(t-w)g''(w)dw;x\right)-\int_x^{\varphi_{n,1}^{\langle\lambda,\mu\rangle}(x)}\left(\varphi_{n,1}^{\langle\lambda,\mu\rangle}(x)-w\right)g''(w)dw\end{aligned}$$

and hence

$$\begin{aligned}\left|\mathscr{P}_n^{\langle\lambda,\mu\rangle}(g;x)-g(x)\right| &\leqslant \mathscr{P}_n^{\langle\lambda,\mu\rangle}\left(\left|\int_x^t(t-w)g''(w)dw;x\right|\right)+\left|\int_x^{\varphi_{n,1}^{\langle\lambda,\mu\rangle}(x)}\left|\left(\varphi_{n,1}^{\langle\lambda,\mu\rangle}(x)-w\right)\right||g''(w)|dw\right|\\ &\leqslant \delta_{n,2}^{\langle\lambda,\mu\rangle}(x)\|g''\|+\left(\varphi_{n,1}^{\langle\lambda,\mu\rangle}(x)-x\right)^2\|g''\|\\ &=\left\{\delta_{n,2}^{\langle\lambda,\mu\rangle}(x)+\left(\delta_{n,1}^{\langle\lambda,\mu\rangle}(x)\right)^2\right\}\|g''\|\\ &=\Upsilon_n^{\langle\lambda,\mu\rangle}(x)\|g''\|. \end{aligned}\tag{3.3}$$

From relation (3.2), we have

$$\begin{aligned}\left|\mathbb{P}_n^{\langle\lambda,\mu\rangle}(f;x)\right| &\leqslant \left|\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x)\right|+|f(x)|+\left|f\left(\varphi_{n,1}^{\langle\lambda,\mu\rangle}(x)\right)\right|\\ &\leqslant \|f\|\mathscr{P}_n^{\langle\lambda,\mu\rangle}(1;x)+2\|f\|=3\|f\|.\end{aligned}\tag{3.4}$$

Now,

$$\begin{aligned}&\left|\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x)-f(x)\right|=\left|\mathbb{P}_n^{\langle\lambda,\mu\rangle}(f;x)-f(x)+f\left(\varphi_{n,1}^{\langle\lambda,\mu\rangle}(x)\right)-f(x)\right|\\ &\quad\leq\left|\mathbb{P}_n^{\langle\lambda,\mu\rangle}(f-g;x)\right|+\left|\mathbb{P}_n^{\langle\lambda,\mu\rangle}(g;x)-g(x)\right|+|f(x)-g(x)|+\left|f\left(\varphi_{n,1}^{\langle\lambda,\mu\rangle}(x)\right)-f(x)\right|,\end{aligned}$$

using relation (3.3) and (3.4) and definition of modulus of continuity, we have

$$\left|\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x)-f(x)\right|\leqslant 4\|f-g\|+\Upsilon_n^{\langle\lambda,\mu\rangle}(x)\|g''\|+\omega\left(f;\varphi_{n,1}^{\langle\lambda,\mu\rangle}(x)-x\right).$$

Applying infimum to all of $g\in C^2[0,1]$, we get

$$\left|\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x)-f(x)\right|\leqslant 4K\left(f;\frac{1}{4}\Upsilon_n^{\langle\lambda,\mu\rangle}(x)\right)+\omega\left(f;\delta_{n,1}^{\langle\lambda,\mu\rangle}(x)\right).$$

This concludes the proof in view of relation (3.1). ∎

Our following theorem determines the rate of convergence of the operators $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ for functions belonging to Lipschitz class $Lip_C(\gamma)$.

For $C>0$ and $0<\gamma\leqslant 1$ the class $Lip_C(\gamma)$ is defined by

$$Lip_C(\gamma)=\left\{f:|f(y)-f(x)|\leqslant C|y-x|^{\gamma}\ where\,x,y\in[0,1]\right\}.$$

Theorem 3.4. *Let $\lambda \in [-1,1], \mu = \mu(n) \to 0$ as $n \to \infty$ and $x \in [0,1]$, then the inequality*

$$\left|\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x) - f(x)\right| \le C\left[\delta_{n,2}^{\langle\lambda,\mu\rangle}\right]^{\frac{\gamma}{2}},$$

holds for $f \in Lip_C(\gamma)$.

Proof. Since $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ are linear and positive in nature and $f \in Lip_C(\gamma)$, we have

$$\begin{aligned}\left|\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x) - f(x)\right| &\leqslant \mathscr{P}_n^{\langle\lambda,\mu\rangle}(|f(t)-f(x)|;x)\\ &= \sum_{k=0}^{n}\hat{p}_{n,k}^{\langle\lambda,\mu\rangle}(x)\left|f\left(\frac{k}{n}\right)-f(x)\right|\\ &\leqslant C\sum_{k=0}^{n}\hat{p}_{n,k}^{\langle\lambda,\mu\rangle}(x)\left|\frac{k}{n}-x\right|^{\gamma}\\ &\leqslant C\sum_{k=0}^{n}\left[\hat{p}_{n,k}^{\langle\lambda,\mu\rangle}(x)\left(\frac{k}{n}-x\right)^2\right]^{\frac{\gamma}{2}}\left[\hat{p}_{n,k}^{\langle\lambda,\mu\rangle}(x)\right]^{\frac{2-\gamma}{2}}.\end{aligned}$$

Applying Hölder's inequality for sums, we obtain

$$\begin{aligned}\left|\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x) - f(x)\right| &\leqslant C\left[\sum_{k=0}^{n}\hat{p}_{n,k}^{\langle\lambda,\mu\rangle}(x)\left(\frac{k}{n}-x\right)^2\right]^{\frac{\gamma}{2}}\left[\sum_{k=0}^{n}\hat{p}_{n,k}^{\langle\lambda,\mu\rangle}(x)\right]^{\frac{2-\gamma}{2}}\\ &= C\left[\mathscr{P}_n^{\langle\lambda,\mu\rangle}((t-x)^2;x)\right]^{\frac{\gamma}{2}}.\end{aligned}$$

This proves theorem 4. ∎

Finally, we give a Voronovskaja asymptotic formula for $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$.

Theorem 3.5. *Let $\lambda \in [-1,1]$, $\mu = \mu(n) \to 0$ as $n \to \infty$ and $f(x)$ be bounded on $[0,1]$. Then, for any $x \in (0,1)$ at which $f''(x)$ exists, we have*

$$\lim_{n\to\infty} n\left[\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x) - f(x)\right] = \frac{1}{2}(l+1)(1-x)xf''(x).$$

Proof. By the Taylor formula, we may write

$$f(t) = f(x) + (t-x)f'(x) + \frac{1}{2}(t-x)^2 f''(x) + (t-x)^2 r(t,x), \tag{3.5}$$

where $r(t,x) \in C[0,1]$ is the Peano form of the remainder. Using L'Hopital's rule, we have

$$\lim_{t\to x} r(t,x) = 0.$$

Applying $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ to (3.5), we obtain

$$\lim_{n\to\infty} n\left[\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x)-f(x)\right] = \lim_{n\to\infty} n\mathscr{P}_n^{\langle\lambda,\mu\rangle}((t-x);x)f'(x)$$
$$+\frac{1}{2}\lim_{n\to\infty} n\mathscr{P}_n^{\langle\lambda,\mu\rangle}\left((t-x)^2;x\right)f''(x)+\lim_{n\to\infty} n\mathscr{P}_n^{\langle\lambda,\mu\rangle}\left((t-x)^2 r(t,x);x\right). \tag{3.6}$$

By the Cauchy-Schwarz inequality, we have

$$\mathscr{P}_n^{\langle\lambda,\mu\rangle}\left((t-x)^2 r(t,x);x\right)\leqslant\sqrt{\mathscr{P}_n^{\langle\lambda,\mu\rangle}\left((t-x)^4;x\right)}\sqrt{\mathscr{P}_n^{\langle\lambda,\mu\rangle}(r^2(t,x);x)}. \tag{3.7}$$

Since $r^2(x,x)=0$ then using (3.7), we can obtain

$$\lim_{n\to\infty} n\mathscr{P}_n^{\langle\lambda,\mu\rangle}\left((t-x)^2 r(t,x);x\right)=0. \tag{3.8}$$

Finally, using (3.6), (3.8) and Lemma 2.2, we get

$$\lim_{n\to\infty} n\left[\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x)-f(x)\right]=\frac{1}{2}(l+1)(1-x)xf''(x).$$

Hence, we get the proof. ■

4. Numerical results

Example 4.1. The convergence of $\mathscr{P}_{15}^{\langle\lambda,\mu\rangle}$ (magenta), $\mathscr{P}_{25}^{\langle\lambda,\mu\rangle}$ (red) and $\mathscr{P}_{45}^{\langle\lambda,\mu\rangle}$ (blue) to $f(x)=\sin(3\sin(3x))$ (black) is illustrated in Figure 3 for fixed $\lambda=-0.5$ and $\mu=\mu(n)=\frac{1}{n^5+2^{\log(n)}}$. Table 1 computes the absolute error $\varepsilon_n^{\langle\lambda,\mu\rangle}(x)=|\mathscr{P}_n^{\langle\lambda,\mu\rangle}(f;x)-f(x)|$ of the function f for various values of x in the interval [0, 1], and Figure 4 displays this error graphically. When n rises from 15 to 45, we notice that the approximation of f by $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ gets better and error also continues to decrease.

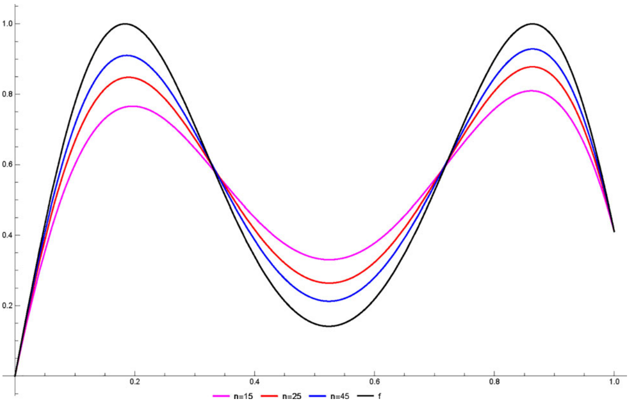


Figure 3. Convergence of $\mathscr{P}_{15}^{\langle\lambda,\mu\rangle}$ (magenta), $\mathscr{P}_{25}^{\langle\lambda,\mu\rangle}$ (red) and $\mathscr{P}_{45}^{\langle\lambda,\mu\rangle}$ (blue) for fixed $\lambda=-0.5$ and $\mu=\mu(n)=\frac{1}{n^5+2^{\log(n)}}$ to $f(x)=\sin(3\sin(3x))$ (black).

Table 1. Estimation of error for various value of x in the interval [0, 1].

x	$\mathcal{E}_{15}^{\langle\lambda,\mu\rangle}$	$\mathcal{E}_{25}^{\langle\lambda,\mu\rangle}$	$\mathcal{E}_{45}^{\langle\lambda,\mu\rangle}$	$\mathcal{E}_{75}^{\langle\lambda,\mu\rangle}$
0.1	0.175061	0.108151	0.0610693	0.0369243
0.2	0.226676	0.146409	0.0861958	0.0534319
0.3	0.0650047	0.0373058	0.0196547	0.0113319
0.4	0.111161	0.0768502	0.0470771	0.0297085
0.5	0.187009	0.121607	0.0707124	0.0432592
0.6	0.159335	0.104776	0.0616742	0.038047
0.7	0.03193	0.0248162	0.0164783	0.0108634
0.8	0.140273	0.0895243	0.0521938	0.0321723
0.9	0.17437	0.110194	0.0634311	0.0387883

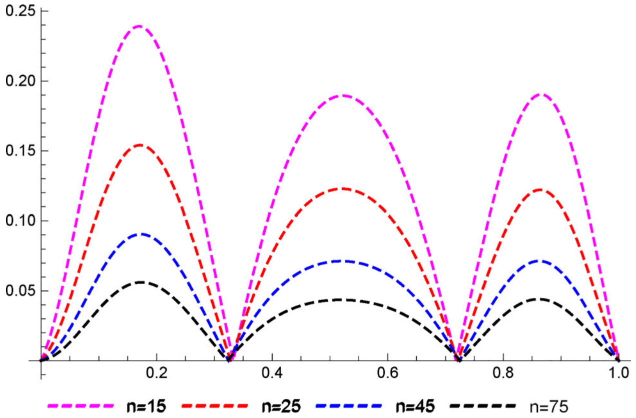


Figure 4. Graph of $\mathcal{E}_{15}^{\langle\lambda,\mu\rangle}(x)$ (magenta), $\mathcal{E}_{25}^{\langle\lambda,\mu\rangle}(x)$ (red), $\mathcal{E}_{45}^{\langle\lambda,\mu\rangle}(x)$ (blue) and $\mathcal{E}_{75}^{\langle\lambda,\mu\rangle}(x)$ (black) with $\lambda=-0.5$ and $\mu=\mu(n)=\frac{1}{n^5+2^{\log(n)}}$ for $f(x)=\sin\Big(3\sin(3x)\Big)$.

Example 4.2. Figure 5 shows the graph for the operators $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ for two different sequences $\mu=\mu(n)=\frac{1}{n!}$ (red) and $\mu=\mu(n)=\frac{1}{n\log(n)}$, (magenta) while keeping $n=30$ and $\lambda=0.5$ fixed for the function $f(x)=x^4-\frac{12x^3}{5}+\frac{193x^2}{100}-\frac{57x}{100}+\frac{3}{50}$ (black). Figure 6 shows the graph for the operators $\mathscr{P}_n^{\langle\lambda,\mu\rangle}$ for the function $f(x)=10x+2\cos(10x)$ (black) with fixed $n=20$ and $\lambda=0.1$

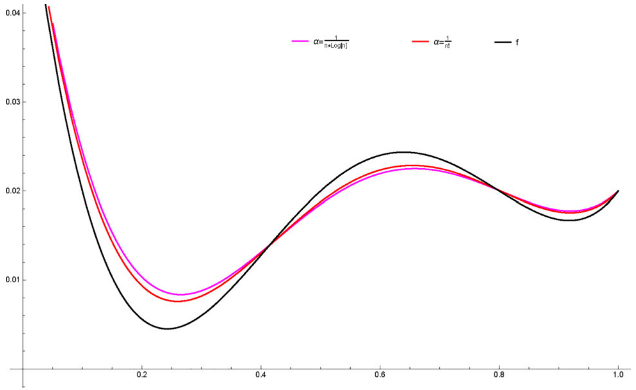


Figure 5. Convergence of $\mathscr{P}_{30}^{\langle 0.5,\frac{1}{n\log(n)}\rangle}$ (magenta) and $\mathscr{P}_{30}^{\langle 0.5,\frac{1}{n!}\rangle}$ (red) to $f(x)=x^4-\frac{12x^3}{5}+\frac{193x^2}{100}-\frac{57x}{100}+\frac{3}{50}$ (black).

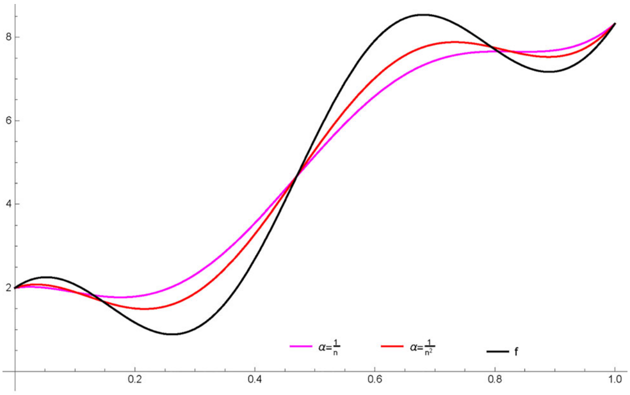


Figure 6. Convergence of $\mathscr{P}_{20}^{\langle 0.1,\frac{1}{n}\rangle}$ (magenta) and $\mathscr{P}_{20}^{\langle 0.1,\frac{1}{n^2}\rangle}$ (red) to $f(x) = 10x + 2\cos(10x)$ (black).

for two different sequences $\mu = \mu(n) = \frac{1}{n^2}$ (red) and $\mu = \mu(n) = \frac{1}{n}$, (magenta). For these two cases, the graphs make it evident that convergence of the operators toward the function occurs best for the sequence with higher rate of convergence.

Acknowledgments

The authors take this opportunity to express their deep gratitude to the learned referees for several valuable suggestions and constructive comments, which resulted in the subsequent improvement of this research paper.

Disclosure statement

The authors declare that they have no conflict of interest.

Funding

This work is supported by the Council of Scientific and Industrial Research (CSIR) with reference no.: 08/133(0029)/2019-EMR-I for the first author.

References

[1] Eggenberger, F., Pólya, G. (1923). Uber die Statistik verkerter Vorgänge. *Z. Angew. Math. Mech.* 1:279–289. DOI: 10.1002/zamm.19230030407.

[2] Stancu, D. D. (1968). Approximation of functions by a new class of linear polynomial operators. *Rev. Roumaine Math. Pures Appl.* 13:1173–1194.

[3] Bernstein, S. N. (1912). Démonstration du théoréme de weierstrass fondée sur le calcul de probabilités. *Commun. Sco. Math. Charkov.* 13(2):1–2.

[4] Johnson, N. L., Kotz, S. (1969). *Discrete Distributions.* Boston: Houghton-Mifflin.

[5] Stancu, D. D. (1970). Two classes of positive linear operators. *Anal. Univ. Timişoara, Ser. Matem.* 8:213–220.

[6] Baskakov, V. A. (1957). An instance of a sequence of linear positive operators in the space of continuous functions. *Dokl. Akad. Nauk.* 113:249–251.

[7] Razi, Q. (1989). Approximation of a function by Kantorovich type operators. *Mat. Vesnik*. 41(3):183–192.

[8] Büyükyazici, İ. (2010). Approximation by Stancu-Chlodowsky polynomials. *Comput. Math. Appl.* 59(1):274–282. DOI: 10.1016/j.camwa.2009.07.054.

[9] Agrawal, P. N., Ispir, N., Kajla, A. (2015). Approximation properties of Bézier-summation-integral type operators based on Polya-Bernstein functions. *Appl. Math. Comput.* 259:533–539. DOI: 10.1016/j.amc.2015.03.014.

[10] Deo, N., Dhamija, M., Miclăuş, D. (2016). Stancu-Kantorovich operators based on inverse Pólya-Eggenberger distribution. *Appl. Math. Comput.* 273(1):281–289. DOI: 10.1016/j.amc.2015.10.008.

[11] Aral, A., Gupta, V. (2016). Direct estimates for Lupas-Durrmeyer operators. *Filomat*. 30(1):191–199. DOI: 10.2298/FIL1601191A.

[12] Cárdenas-Morales, D., Gupta, V. (2014). Two families of Bernstein-Durrmeyer type operators. *Appl. Math. Comput.* 248:342–353. DOI: 10.1016/j.amc.2014.09.094.

[13] Dhamija, M., Deo, N. (2016). Jain-Durrmeyer operators associated with the inverse Pólya-Eggenberger distribution. *Appl. Math. Comput.* 286:15–22. DOI: 10.1016/j.amc.2016.03.015.

[14] Dhamija, M., Deo, N. (2017). Approximation by generalized positive linear Kantorovich operators. *Filomat*. 31(14):4353–4368. DOI: 10.2298/FIL1714353D.

[15] Dhamija, M., Deo, N., Pratap, R., Acu, A. M. (2022). Generalized Durrmeyer operators based on inverse Pólya-Eggenberger distribution. *Afr. Mat.* 33(1):1–13. DOI: 10.1007/s13370-021-00949-8.

[16] Deshwal, S., Agrawal, P. N., Araci, S. (2017). Modified Stancu operators based on inverse Pólya-Eggenberger distribution. *J. Inequal. Appl.* 2017(1):57. DOI: 10.1186/s13660-017-1328-9.

[17] Cai, Q. B., Lian, B. Y., Zhou, G. (2018). Approximation properties of. λ Bernstein operators. *J. Inequal. Appl.* 2018(1):61. DOI: 10.1186/s13660-018-1653-7.

[18] Acu, A. M., Manav, N., Sofonea, D. F. (2018). Approximation properties of λ-Kantorovich operators. *J. Inequal. Appl.* 2018(1):202. DOI: 10.1186/s13660-018-1795-7.

[19] Rahman, S., Mursaleen, M., Acu, A. M. (2019). Approximation properties of λ-Bernstein-Kantorovich operators with shifted knots. *Math. Methods Appl. Sci.* 42(11):4042–4053. DOI: 10.1002/mma.5632.

[20] Cai, Q. B. (2018). The Bézier variant of Kantorovich type λ-Bernstein operators. *J. Inequal. Appl.* 2018(1):90. DOI: 10.1186/s13660-018-1688-9.

[21] Cai, Q. B., Zhou, G. (2018). Blending type approximation by GBS operators of bivariate tensor product of λ-Bernstein-Kantorovich type. *J. Inequal. Appl.* 2018(1): 268. DOI: 10.1186/s13660-018-1862-0.

[22] Acu, A. M., Acar, T., Radu, V. A. (2019). Approximation by modified U_n^ρ operators. *RACSAM*. 113:2715–2729. DOI: 10.1007/s13398-019-00655-y.

[23] Braha, N. L., Mansour, T., Mursaleen, M., Acar, T. (2021). Convergence of λ-Bernstein operators via power series summability method. *Appl. Math. Comput.* 65:125–146.

[24] Kajla, A., Acar, T. (2018). Blending type approximation by generalized Bernstein-Durrmeyer type operators. *Miskolc Math. Notes*. 19(1):319–336. DOI: 10.18514/MMN.2018.2216.

[25] Kajla, A., Acar, T. (2018). A new modification of Durrmeyer type mixed hybrid operators. *Carpathian J. Math.* 34(1):47–56. DOI: 10.37193/CJM.2018.01.05.

[26] Kajla, A., Mursaleen, M., Acar, T. (2020). Durrmeyer-type generalization of parametric Bernstein operators. *Symmetry*. 12(7):1141. DOI: 10.3390/sym12071141.

[27] Ditzian, Z., Totik, V. (1987). *Moduli of Smoothness*. New York: Springer-Verlag.

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-14881-5

Blockchain DrivenAccess control architecture for the internet of things

Rajiv K. Mishra[1] · Rajesh K. Yadav[1] · Prem Nath[2]

Received: 4 June 2022 / Revised: 26 January 2023 / Accepted: 6 February 2023

Abstract

In the last few years, Internet of Things (IoT) and Blockchain (BC) technology have been ruling their respective research area. The integration of IoT and Blockchain enables delivering many effective and prominent services by incorporating in-built features like scalability, flexibility, and resilience along with availability and integrity. However, taking into account the constrained nature of IoT devices, it's quite hard to implement BC peers on top of IoT devices. Additionally, the rate at which transactions are produced by a huge number of constrained devices, BC could not handle effectively. The proposed work presented a solution to cater to these challenges. It incorporates the Interplanetary File System (IPFS) for the distribution of resources generated by IoT devices. The proposed system is based on the Hyperledger Fabric BC framework and comprises smart contracts that are accountable for policy definition, policy enforcement, user identity management, and data retrieval. The experimental results illustrate that the running time taken by smart contract methods of the proposed solution is fairly less than the prominent work in the same domain. The performance evaluation clearly depicts how effectively the presented model achieves Confidentiality, Availability, Integrity, and prevents DoS and DDoS attacks.

Keywords Internet of Things (IoT) · Blockchain · Hyperledger fabric · IPFS · Smart contract

✉ Rajiv K. Mishra
mishrarajiv99@gmail.com

[1] Department of Computer Science & Engineering, Delhi Technological University, Delhi, India

[2] Department of Computer Science & Engineering, HNB Garhwal University, Garhwal, India

Published online: 03 March 2023

1 Introduction

The recent growth in the field of computer hardware and the internet has made it much easier to interconnect a large number of devices through wireless networks resulting in an exponential scaling of the Internet of Things (IoT). The entire process of interconnecting devices and transforming them into the smarter device has made everyone's life more convenient [21]. Now a day the IoT has become an integral part of our everyday lives by facilitating routine tasks through their smart services. The convenience and ease offered by IoT to improvise our living standard results in continuous tracking and intruding in our private space which in turn makes our privacy and safety more vulnerable. The data resource generated through these smart devices mostly encompasses sensitive information and thus imposes a serious threat if accessed illicitly. Multiple compnies are performing a series of operations like storing, processing, sharing, and analyzing on the data produced by many smart devices for offering several useful and innovative services to the society. Data security and privacy is one of the key concern that has not been addressed as much as it is required especially in the IoT context.So the major challenge IoT is confronted with is Privacy and Security, and to overcome this concern access control techniques are required.The access control mechanisms are vital to guard resources, which have been commonly utilized in a variety of systems [33]. Moreover, the security mechanism must consider the storage and processing capabilities of these constrained devices.

The sensitive data generated by the IoT may suffer from serious security breaches if the IoT system is not integrated with sufficient protection measures. Thus, the IoT strategy and access policy need to be aligned. In the proposed resource access control scheme, context plays a key role while designing the policies, during resource access, and post access. The proposed mechanism enables resource owners to control who is accessing their resource, which of the resource is being accessed, and when it happens.

An inclusive access control scheme necessitates three components: authentication, authorization, and auditing [45]. The authentication recognizes the approved identity of the requestor. The authorization verifies if the requestor has the appropriate right to perform an operation or access any specific resource. Lastly, the auditing allows the posterior analysis of the realized activities in the system.

Some of the existing literature works have approached the problem of access control employing centralized schemes where a central entity is accountable for running the authorization mechanisms. These conventional access control methods do not meet the constraints enforced by the IoT environment and lack to deliver features like scalability, flexibility, and resilience. These problems can be resolved through a distributed scheme, in which participating entities are indulging themselves in authorization decisions and decision making is not delegated to any central entity.

Security and privacy concerns raised in the age of IoTs necessitate that access control mechanisms should possess some additional non-functional requirements like scalability, flexibility, resilience, and lightweight other than integrity, confidentiality, and availability. Considering the limitations of commonly used centralized and decentralized access control models, Blockchain-based solutions can offer more desirable solutions in the IoT environment [12, 24]. Blockchain can be described as a technology that facilitates the immutability and integrity of information through a peer-to-peer network that comprises several distributed nodes where records of transactions are maintained [25].

The conventional scheme of data sharing normally records IoT data on the third-party agency which sometimes results in the revelation of sensitive data. Transmission and storage of data should be shifted from a centralized storage system to a decentralized one to guard the availability and privacy of data in a much better way. This shifting to distributed storage technique has many advantages over centralized techniques like large data throughput, more cost-effective, and resilient. This work proposes a framework for effective access control over data stored in distributed storage. The proposed work adopted a distributed, peer-to-peer storage system, IPFS (InterPlanetary File System).

This article proposes a novel framework for decentralized resource access control. In the proposed solution, access control schemes are imposed through Blockchain technology. Since, public Blockchain has its own challenges such as scalability issues and higher cost of transaction, in the proposed solution, access control schemes are imposed through Hyperledger fabric.

The key contributions of this paper are as follows:

- A blockchain-based, access behavior-driven access control mechanism is proposed for efficient resource sharing among IoT devices.
- IPFS-based decentralized data storage scheme resulting in high availability of data.
- Multiple permission levels are defined to offer permissioned access privileges to resource consumers.
- Either positive or negative value is assigned to IoT devices depending on their access behavior which eventually facilitates a dynamic resource access scheme.
- Provides two-phase authorization for resource access- static authorization (based on predefined access policies) and dynamic authorization(based on resource access behavior)

The remaining part of this work is structured as follows: Section II presents an overview of Hyperledger fabric and its constituent elements. Section III analyzes the related work found in the literature offering access control. Section IV describes the proposed access management framework along with the corresponding algorithm. Section V presents the performance of the proposed approach. Finally, we concluded our work in section VI.

2 Preliminaries and basic definitions

2.1 InterPlanetary file system (IPFS)

IPFS [31] is a new internet protocol and P2P distributed file system that links computational devices with a common system of files. High throughput is achieved by IPFS by enabling content-addressed hyperlinks. IPFS employs content-addressing to distinctively recognize every file from the global space. Distributed hash tables (DHT), incentivized block exchange, and self-certifying namespaces are a few technologies that bring together on a common platform by IPFS. It enjoys an advantage over cloud storage in that data is distributed and uploaded at various parts of the world and no central host concept is there hence it does not suffer from a single point of failure. Whenever a file is uploaded on the IPFS system, a distinct fingerprint termed the cryptographic hash is created which is later on used to retrieve the file. This contenthash can be thought of as a URL on the web.

2.2 Hyperledger fabric

In the Hyperledger Fabric, every program runs in the docker containers. This container offers an environment that isolates physical resources and application programs and to make sure the safety measures of the application, containers are separated from each other. Fabric is a sort of coalition chain, in which each of the nodes is required to be authorized to connect with the blockchain system. Kafka message queue-based consensus is employed by Fabric which results in faster consensus even in the scenario of large-scale application. Hyperledger Fabric is capable to address various limitations of the public chain discussed earlier.

Membership Service Provider (MSP) and Nodes in Hyperledger Fabric {CA, Client Node, Peer Node, Orderer Node}.

The MSP is implemented as a Certificate Authority (CA) which offers integrated management for digital certificates of associate nodes and produces or revokes identity certificates of associates.

Nodes in Hyperledger Fabric are communication entities and require to have a valid certificate to interact with the network. Nodes can be classified as a Client node, Peer node, and Orderer node.

Client node, mostly an application based on SDK is accountable to run the blockchain system while interacting with the peer node. The operation performed by the client is bifurcated into two classes. First belongs to the management class, which is primarily responsible to administer the nodes through the start, stops, and configures operation. The second is chaincode class which is primarily managing the life cycle of chaincode through the install, upgrade, and execution operation of chaincode.

Peers are an essential component of the Blockchain and are accountable to host ledgers and chaincodes. To make query or update ledger, applications connect to peers through the invocation of chaincode, Peers are classified into two categories which are endorser and committer. Verification, simulation, and endorsement of transactions are the liability of the endorser node. Updation of Blockchain and ledger status are the accountability of the committer node by validating the genuine transaction. Peers can be classified as committing peer, endorsing peer, leader peer, and anchor peer.

Orderer is accountable for many operations. It accepts the transactions, arranges the transaction as per specific policy, creates a block and wrapping up transactions within it, and finally facilitates its distribution.

Channel The Hyperledger Fabric devises a channel scheme to segregate the Blockchain data of different groups to maintain data privacy and confidentiality. Every channel comprises an autonomous personal ledger and a Blockchain. As a result, Hyperledger Fabric is considered as a system that comprises multiple channel, multiple ledger, and multiple Blockchain.

Ledger The ledger is an ordered, tamper-proof record of all state transitions where state transitions are an outcome of chaincode invocation. The data of Fabric is accumulated as a distributed ledger in the key-value pair form and the state of the ledger is comprised of these key-value pairs. One ledger is required for each channel and its copy is maintained by each peer.

CHAINCODE In Fabric, chaincode is the terminology used for smart contracts. Chaincode is a programmable code written in golang (supports other languages like Java) to implements interfaces and contains all the business logic. The interface offered by chaincode is exploited by applications to interact with a blockchain ledger. In Fabric, chaincode generates transactions while running on the peers. Therefore, it is required to be installed on each peer that wishes to endorse a transaction. Assets are generated and renewed by a particular chaincode, and thus inaccessible through a different chaincode. Since it encompasses business logic, developers need to write different chaincode to implement different applications.

3 Related work

While the distribution of access control schemes in diverse layers is not straightforward for all time we have categorized this standardization from two perspectives: Architecture layer and Authorization model layer. Architecture layer classification of access management consists of technologies like XACML [32], OAuth [41], UMA [47], and Blockchain [49]. Access schemes offered through XACML, OAuth, and UMA are centralized schemes while Blockchain offers distributed means to achieve the same. Similarly, the Authorization model layer classification of access management is comprised of three approaches: centralized, decentralized, and hybrid. Some significant works exploiting these approaches are emphasized in this section.

3.1 Architecture layer access control

Atlam et al. [44] presented a mechanism for access control Adaptive Risk-Based Access Control(AdRBAC) having four input elements which are user context, resource sensitivity, action severity, and risk history. These components assess the risk on all access queries and under the risk evaluation, an access decision is made. The risk policy defined by this scheme is based on the principle of XACML.

Sciancalepore et al. [3] presented an OAuth based (token-based) access control framework OAuth-IoT. In this scheme a Gateway, being the vital component handles major tasks like collecting data from the smart devices, receiving access queries through third-party applications, taking care of applications authentication, and authorization. However, OAuth-IoT is a token-based scheme where tokens need to be validated from the authorization server (AS), its presumption that resources are always connected with the internet is not possible all the time in a constrained environment.

Cirani et al. [9] proposed “IoT-OAS” an OAuth based framework for IoT environment which offers authorization delegation capability to HTTP/CoAP based service providers. However, larger radio transmission took place while having bulky size packets at the application level.

Cruz-Piris et al. [14] proposed an access control scheme based on UMA for IoT environments having web-based services. UMA is an improvisation of OAuth2.0 with variations in application layer protocol, token format, and transport layer security which enables it more suitable for IoT. However, this scheme works exclusively for MQTT protocol and is not generalized for other IoT protocols like CoAP, AMQP, and REST. Additionally, the MQTT communication system is not enabled with the encryption system that makes it protected even in insecure environments.

3.2 Authorization model-based access control

Authorization model-based access control is mainly categorized as a centralized model, distributed model, and hybrid model.

Role-Based Access Control (RBAC) [37] and Organization-Based Access Control (OBAC) [16] are two major centralized-based schemes for access control. Since both of these models, RBAC and OBAC are based on a centralized architecture, they are easy to implement and manage but also confronted with few inbuilt limitations. The primary concerns regarding these schemes are that they are not suitable for IoT devices as implementation is too complex to implement without any lightweight tool or mechanism. Additionally, single-point failure, large-scale implementation, and flexibility are other concern that prevails.

The distributed architecture comprises several models: Attribute-Based Access Control (ABAC) [15, 50], Usage Control-Based Access Control (UCON) [29, 51], Trust-Based Access Control (TBAC) [35], and Capability-Based Access Control (CBAC) [11]. Attributes being the core concepts in ABAC models provide more scalable and fine-grained means to gain access to resources. However, it also exhibits a few limitations and the most crucial one is its complex deployment, apart from that sensor data and attribute values are required to map together also. UCON encompasses a collection of new perceptions in contrast to prevalent conventional models but yet it's not enough to take the context of IoT into account for several reasons: broad elucidation of the access method is missing and the availability of only conceptual model as of now. TBAC introduces some dynamic elements in the decision process of the access scheme in terms of trust value which is associated with every constrained device. But so far this model is only implemented for the cloud environment and it is not fit for a constrained environment. CBAC is based on the notion of capability which is nothing but a privilege and entities possessing the privilege are granted to access the specified resource. Despite providing better flexibility and distribution than the previous mechanisms, this model has to cope with various limitations. One of the major concerns of this model is its usability on mobile devices and not considering the context during the evaluation of the access permission process. Capability propagation and revocation are also an issue that needs to be tackled [2, 5, 6, 18, 34, 38, 46, 48].

Hybrid architecture-based models are Smart Organization Based Access Control (SmartOBAC) [4] and Pervasive Based Access Control (PBAC) [8]. Although these hybrid approaches tender better flexibility and scalability but are susceptible to DoS attack in certain scenarios (if overflow at node surpasses threshold) and security strategy descriptions are complicated.

Solutions based on Blockchain are mainly bifurcated into Transaction-Based Access Control (TransBAC) [7, 23, 28] and Smart Contract-Based Access Control (SCBAC) [1, 22, 26, 27, 30, 42, 52]. Blockchain-based solutions have multiple inbuilt advantages like decentralization, immutability, resilience, and data integrity. Consequently, many researchers have presented an array of access control solutions by incorporating existing models with Blockchain (mainly smart contracts).

Novo et al. [26] proposed a blockchain-based architecture for scalable access control that encompasses manager and management hub nodes in addition to IoT and Blockchain networks. However, it suffers from various limitations like failure of the management hub, all the IoT devices connected through it get disappear from the network, and in the case of a malicious manager node, the system becomes insecure.

Zhang et al. [52] presented a smart contract-based framework for access control which consists of three different types of smart contracts. However, the major challenge with the solution is that it requires defining an access policy for each pair of subjects and object which eventually makes it static and specific.

Liu et al. [22] utilize the concept of attribute-based access control and Hyperledger fabric network. However, this scheme does not ensure the trustfulness and security of the participating nodes.

Siris et al. [42] proposed four models for the authorization of entities involved in the data-sharing process. However, it is confronted with two limitations: depends on tokens for authorization, and the incorporation of multiple blockchain networks and many authorization servers incur an inter-ledger delay.

Oktian et al. [27] proposed “BorderChain” a Blockchain-based mechanism to achieve access control. In this, only authorized nodes are allowed to communicate with IoT gateways. However, it is also a token-based approach and is not compatible with real-time use cases.

Alphand et al. [1] and Pinno et al. [30] combines existing work with blockchain and exploit smart contract to enable access control. Moreover, these models are better equipped to provide access control in the IoT environment by still they suffer from several limitations like additional overhead incurred to implement and manage smart contracts, demonstration of these schemes in real-world scenarios, etc.

Rizzardi et al. [36] presented the integration of permissioned blockchain along with IoTmiddleware considering fog computing perspective. This mechanism utilizes the consensus feature of blockchain to prevent altering predefined access rules within the IoT network. However, this work was restricted to very few data sources and did not simulate malicious behaviors in the IoT context.

Han, Dezhi, et al. [13] proposed attribute-based access control model for Internet of Things. The proposed solution utilized Hyperledger fabric to protect against unauthorized access to sensitive data. However, the scalability of this solution is a big concern, as there is a big mismatch in the speed of IoT data generation and Blockchain block creation and validation. Additionally, this model has not been implemented or tested in real-world scenarios.

Shi et al. [40] proposes a private Ethereum-based access control mechanism (BacS) for distributed IoT system. This model ensures a single identity applicable to all domains within a distributed IoT network and thus simplifies the complexity of the identity management process. However, the proposed model is not suitable for small IoT networks, and even the conventional access control approach performs better in this scenario. Additionally, this approach is not appropriate for privacy protection as the concerning algorithm works quite slowly.

Sisi Zuhu et al. [43] proposed a blockchain-based solution for energy-aware mobile crowd sensing in the Internet of Things (IoT). However, the model was designed to be energy-aware, but the exact energy consumption of the system was not evaluated.

Kamal et al. [17] presented a confidentiality-preserving architecture for the distributed cloud storage systems. The proposed architecture even beats popular techniques such as AES and ICA in terms of memory consumption and time taken to perform encryption and decryption. However, it is restricted to genetic algorithms only while can have a better scope if extended for deep learning or fuzzy logic.

The work done in [10, 19, 20] presents some interesting Blockchain-based solutions for access control in IoT environments. Although, the proposed models offer a secure and auditable way to manage access to sensitive data in an IoT environment but they did not take scalability and energy consumption into consideration while designing the architecture.

Additionally, since the model was based on a centralized trusted authority, it compromises the privacy of data Tables 1, 2, 3 and 4.

Further, to illustrate the novelty of our solution, the prominent work in the same domain are analyzed with our proposed architecture as shown in table [21].

In today's time, multiple services are being offered which rely on smart devices to share their data with each other securely. To accomplish this, a blockchain-based and access behavior-driven access control technique is proposed in the subsequent section which employs both static and dynamic authorization of the communicating entities.

4 Proposed framework

In this section, we have presented a blockchain -based access control mechanism for the IoT environment. The entire process of secure data sharing is elaborated further through a series of sub-sections: Policy Model, Storage Model, System Architecture, System Interaction & Workflow, Smart Copntract, and Algorithm & Implementation.

4.1 Policy model

Mostly the data resources generated by constrained devices are unstructured [26]. Considering some real-world IoT scenarios like smart cameras taking real-world images and generating pictures or video resources, the microphone receives sound and generates audio resources. Physical signals (humidity, temperature, pressure, light, etc.) are perceived by sensors and translated into digital signal resources. Since the majority of this data is unstructured, it's infeasible to store it in a relational database directly. Moreover, these data resources are real-time data, they are required to be distributed to authorized entities in time. The resource data generated by constrained devices are distributed over IPFS and cryptographic hash (hyperlinks) is generated in return. These resource hyperlinks are uploaded to Blockchain through an application gateway Fig. 1.

4.2 Storage model

In this work, we designed a novel data storage model that is based on IPFS and Blockchain. The proposed storage model comprises decentralized storage (IPFS) where actual IoT-generated data is recorded in encrypted form while its corresponding cryptographic hash is uploaded onto the Blockchain network. The cryptographic hash of the actual content is a fixed-size data that requires significantly less space and is thus very much suited for integrating IoT and Blockchain. The storage scheme in Fig. 2 illustrates that all IoT generated data (RO data) is uploaded on off-chain storage and its corresponding fixed-size hash is recorded on the blockchain. Additionally, whenever an authorized entity (RC) wishes to access some data, it firstly fetches the hash data from Blockchain and subsequently gets the actual data from the off-chain storage.

4.3 System architecture

The proposed IoT Blockchain platformencompasses a huge number of IoT devices (RC & RO), distributed data storage (IPFS), user devices (RC), servers (RO), and Gateways that are

Table 1 Comparative analysis with state of the art techniques

Paper	Implementation	Blockchain Platform	Consensus	Data Censorship	Data Confidentiality	Storage System	Scalability	Behavior-driven	Auditability
[26]	Yes	Ethereum (Private)	PoC	Yes	No	Blockchain	No	No	No
[52]	No	Ethereum	PoW	Yes	No	Blockchain	No	No	Yes
[22]	Yes	Hyperledger Fabric	Kafka	Yes	Yes	CouchDB	Yes	No	No
[42]	No	Ethereum & Hyperledger Fabric	PoW & PoA	Yes	No	Blockchain	No	No	No
[39]	Yes	Hyperledger Fabric	Raft	Yes	Yes	CouchDB	Yes	No	No
Our	No	Hyperledger Fabric	Kafka	No	Yes	IPFS	Yes	Yes	Yes

Table 2 Device Information Table(DIT)

RC	RO	R	Action	LRtime	SLRtime	TS
RC A	RO X	File 1	R	2022/01/09 15:21:18	2022/01/08 10:20:33	0.19
RC B	RO Y	File 2	R,w	2022/01/01 03:47:55	2022/01/01 03:44:00	−0.51
RC C	RO Z	Program 3	W	2022/01/05 18:42:12	2022/01/03 14:11:48	0.39

Table 3 Device Penalty Table

RC	RO	Access Behavior	Penalty
RC A	RO X	Request canceled after approval	Request blocked for 15 minutes
RC B	RO Y	Too frequent access request	Request blocked for 1 hour
RC C	RO Z	Multiple requests within a fixed period	Request blocked for 2 hours
RC D	RO W	Requesting higher authorization level resource	Request blocked for 20 minutes

coupled together with a blockchain network. Both RC & RO who seeks access to the resource and hold the requested resource respectively are linked to the Blockchain through a Gateway. The IPFS is connected with IoT devices for data storage & data retrieval. All the RC’s and RO’s are required to register themselves under at least one Device Managers (DM) who subsequently register them within the blockchain network. The complete structure is represented in Fig. 3.

4.4 System interaction and workflow

The sequence diagram depicted in Fig. 4 captures the sequence of interactions among various components of the proposed scheme which are explained as follows:

- Network setup
- Deployment of chaincode by Endorsing peer
- Device Managers register themselves onto the Blockchain network.

Table 4 Terminologies used in Algorithm

Term	Description
RC	Resource Consumer
RO	Resource Owner
R	Resource
Action	read/write/execute.
UnblockTime	time until which request is blocked
StaticCheck	predefined access policies
DynamicCheck	regulates consumers behavior dynamically
LRtime	Last Request time
allowedInterval	the minimum allowable time between successive requests.
NoRR	Number of Recent Requests (request made in a fixed time interval)
fine]X]	Request canceled after approval
fine]Y]	Too frequent access request
fine]Z]	Multiple requests within a fixed period

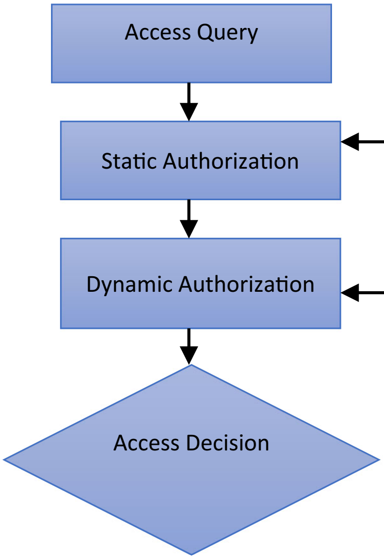


Fig. 1 Proposed policy model

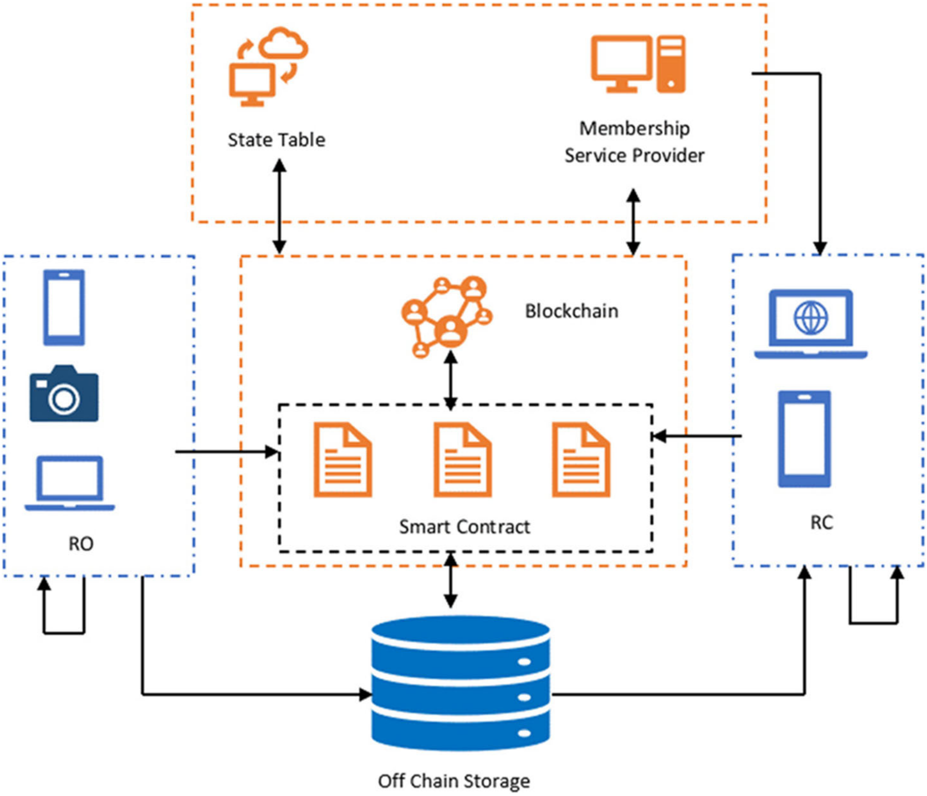


Fig. 2 Storage Model

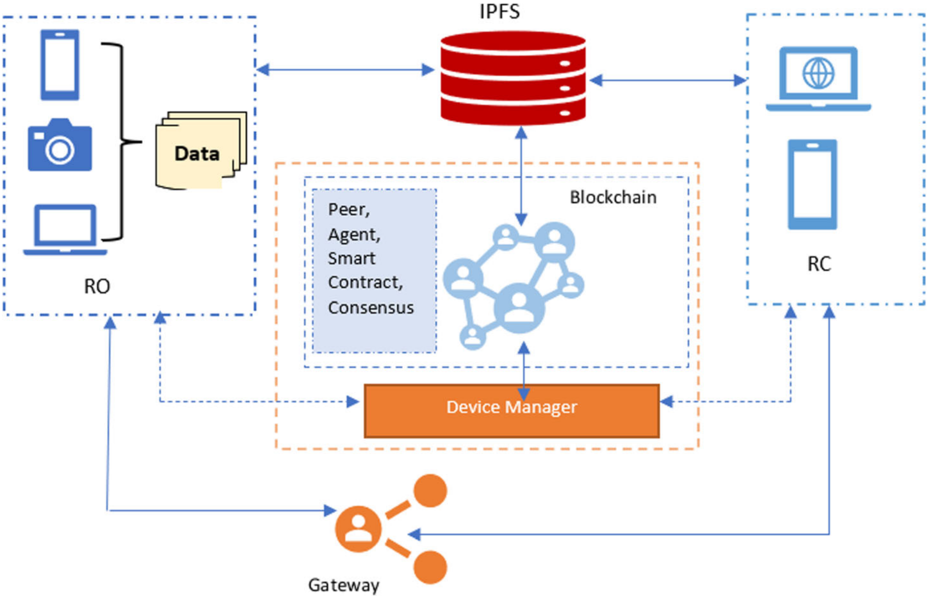


Fig. 3 Proposed system architecture

- Device Manager registers IoT devices (RO's and RC's) under them. (However, an IoT device can de-register itself from any Device Manger)
- Resource access policies are defined by the device manager and forwarded to the corresponding RO seeking its approval.
- The RO approves the received access policy and notifies it to the nearest Gateway.
- Access policy is recorded on the Blockchain through a transaction.
- The RO sends resource (data) upload requests to the IPFS along with its authentication credentials.
- After successful authentication, the IPFS uploads the resource and returns the content hash of the uploaded data to the RO.
- The RO forwards the signed content hash along with its retrieval context to the Gateway.
- The content hash and the context are recorded onto the Blockchain through a transaction.
- RC sends a request for a resource to a Gateway.
- The Gateway translate request to a Blockchain action.
- The Blockchain action is run by peers.
- If RC is not an authorized entity access request is denied and RC is notified through the Gateway.
- If RC is an authorized entity then encrypted (by the RC public key) content hash is returned to the Gateway.
- The Gateway forwards the message into the CoAP format.
- Upon receiving the message, RC decrypts the message by its private key and request to the IPFS by sending a signed content hash.
- The IPFS returns the requested resource.
- Upon receiving resources along with its content hash, The RO verify the content hash with the one it received from the Blockchain.
- Meanwhile, The IPFS report to the corresponding RO about its resource access by this RC.

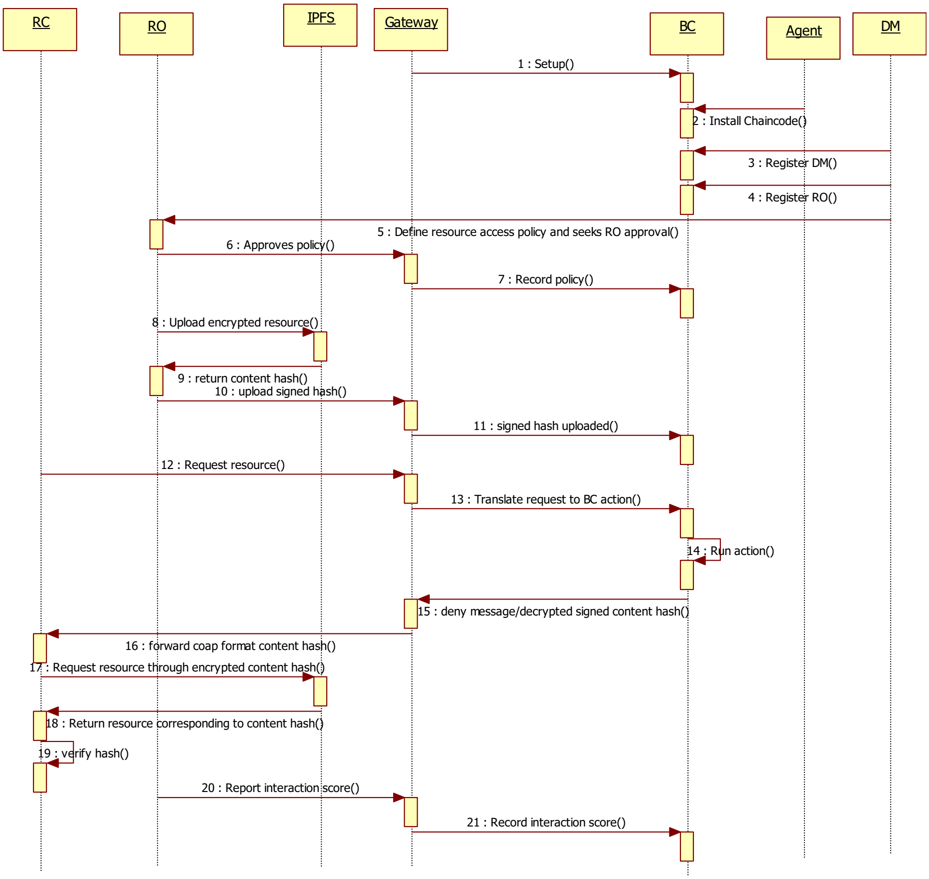

RC
RO
IPFS
Gateway
BC
Agent
DM
1 : Setup()
2 : Install Chaincode()
3 : Register DM()
4 : Register RO()
5 : Define resource access policy and seeks RO approval()
6 : Approves policy()
7 : Record policy()
8 : Upload encrypted resource()
9 : return content hash()
10 : upload signed hash()
11 : signed hash uploaded()
12 : Request resource()
13 : Translate request to BC action()
14 : Run action()
15 : deny message/decrypted signed content hash()
16 : forward coap format content hash()
17 : Request resource through encrypted content hash()
18 : Return resource corresponding to content hash()
19 : verify hash()
20 : Report interaction score()
21 : Record interaction score()

Fig. 4 Control message flow of the proposed model

- The RO sends an interaction score (trust score) to the Gateway.
- The interaction score is recorded on the Blockchain.

4.5 Smart contract

The proposed system employs three smart contracts (chaincode) to manage resource sharing among the IoT devices. These smart contracts are Access Policy Contract (APC), Device Contract (DC), and Trust Contract (TC). APC is being the core of the model and implements an access management scheme. DC encompasses the procedure to upload the URL(content hash) of data generated by IoT devices and a procedure to query it. It also comprises information about the IoT devices for their identification and authentication. TC consists of a method that associates a trust value to each IoT device depending on their access behavior and a method to retrieve it. On every resource access request by a resource consumer (RC), both permission level and past access behavior of the RC are evaluated and access is allowed only if a positive response is evaluated in the process. Furthermore, a complete depiction of APC, DC, and TC is given below.

Dc This smart contract includes methods to upload the content hash of data generated by IoT devices and a procedure to query it. A Device Information Table (DIT) is maintained by this smart contract holding all relevant information of the device which are eventually utilized for their identification and authentication during device registration (under Device Manager) and resource access requests. It comprises of methods like registerThing, getThing, and getAuthenticity. The DIT includes the following information:

- RC: the entity that sends access requests.
- RO: the entity that holds (owner) requested resources.
- Resource(R): specific resource (data or file) requested by RC.
- Action: read (r), write (w), or execute(x) operation.
- LRtime: last request time for a resource by the RC.
- SLRtime: second last request time for a resource by the RC.
- TrustScore(TS): final trust score of the RC according to its access behavior.

APC Being the main smart contract of the proposed model, it manages the access control among the IoT devices. Every time an RC requires accessing a resource of an RO, it sends an access request to the system (through the gateway). Subsequently, APC is executed and takes care of maintaining the access management of the RC. This smart contract consists of multiple methods to serves the purpose: addPolicy, deletePolicy, updatePolicy, verifyPolicy, and verifyAccess which are employed for the inclusion of new access policy, removal of existing policies, modifications of access rules, verification of newly defined policy, and verification of authorization against current access request respectively.

Tc This smart contract primarily focuses on assessing the access behavior of various registered IoT devices (RC and RO) by implementing three methods: setTrust, getTrust, and setFine. Depending on the access behavior either positive or negative value(fine) is associated with registered devices. Multiple reasons cause negative fine assignment which is as follows:

- RC sends access request before the allowed interval (too frequent access request)
- multiple access requests by the RC within a fixed period.
- access request of a resource having a higher authorization level
- not accessing resources (canceled request) after approval.

If none of these situations occurs then a positive value is assigned to the corresponding device. TC also maintains a penalty table to record access behavior and the corresponding penalty of the IoT devices.

4.6 Algorithm & implementation

Trust-based authorization With the trust-based authorization scheme, both participating entities (RC & RO) are assigned a Trust Score (TS) depending on their past access interactions. There are two types of TS values:

- Local: - Corresponding to RC for the requested resource.
- Global: - Overall trust score for RC.

$$\boldsymbol{TS} = \boldsymbol{TS_{local}} + \boldsymbol{TS_{global}} \tag{1}$$

$$\boldsymbol{TS} = \boldsymbol{trust.setTrust(\ r, fine)} \tag{2}$$

Local TS is computed corresponding to the current access request, and it depends on multiple access violations. On each such violation, a predefined fine is imposed on the requested entity.

$$\mathbf{TS_{local}} = \left(\sum_{i=0}^{2} \boldsymbol{fine[i]}^{i+1} \right) \Big)/\mathbf{3} \tag{3}$$

The second argument of the above method "fine" is a vector that comprises multiple components representing different access violation scenarios.

fine[0] = X (RC tries to access a resource that requires a higher authorization level).

fine [1] = Y (RC tries to access resource before allowed time interval).

fine [2] = Z (RC made frequent access request).

{one more form of fine, fine [45] can be added if RC got access approval but did not access requested resource}

$$\mathbf{TS_{global}} = \left(\sum_{i=1}^{n} \boldsymbol{W}^{n-i} * \mathbf{TS_{locali}} \right)/\mathbf{n} \tag{4}$$

Where Wi is a weight, higher weight is assigned to most recent interactions and lower weight for past interactions.

W(n-i) is an aging parameter.

Now, assigning expression for $\mathrm{TS_{local}}$&$\mathrm{TS_{global}}$ from eq. 3 & 4 into eq. 1.

$$\mathbf{TS} = \left(\sum_{i=0}^{2} \boldsymbol{fine[i]}^{i+1} \right) \Big)/\mathbf{3} + \left(\sum_{i=1}^{n} \boldsymbol{W}^{n-i} * \mathbf{TS_{locali}} \right)/\mathbf{n} \tag{5}$$

Algorithm 1 accessControl ()

```
Input: RC, RO, R, time, Action
Output: result, penalty
Requirement: StaticCheck←false, DynamicCheck←true,
penalty ← 0
r ← Policy [(RC, RO, R)] [Action]
if time >=UnblockTime then
    r. UnblockTime ← 0
    if (r. permission) then
        StaticCheck ← true
    else if r. permissionLevel! = Action
            fine[0]←X
        Endif
    Endif
    if time - r. LRtime < r. allowedInterval then
        fine[1]←Y
        NoRR← time - r. SLRtime
        if NoRR >= r. threshold   then
            fine[2]←Z
            Add implicit behavior IB to the behavior list of
            TC
        Endif
    DynamicCheck← false
    Endif
Endif
r. SLRtime ← r. LRtime
r.LRtime←time
TS←trust.setTrust(r, fine)
if (fine[2]) then
    UnblockTime←time+mod(TS)*multiplier
Endif
Check← StaticCheck AND DynamicCheck
if (Check = = TRUE) AND (TS>=$TS_{th}$)then
    Trigger getApproval()
Else
    Notify penalty
Endif
```

Algorithm 2 Uploading data to IPFS

```
Input: data resource, Old resourceTree hash (ORT-hash)
Output: content hash of data resource, new ResourceTree hash (NRT-hash)
ORT_object← get the old version of ResourceTree object by ORT_hash
if data resource is text data,then
    DataPackage← get data package according to type property of data resource
    If DataPackage is Null then
        Create new DataPackage
        Insert ResourceData to the DataPackage
    Else
        Append DataResource to DataPackage
    Endif
    ifDataPackage reached storage limit then
        Store DataPackage to IPFS and get content hash
        NRT_hash← attach DataResource to DataBlock and obtain new resource tree hash
    Else
        DataPackage is temporarily stored
    Endif
Else
    Content_hash← store dataResource to IPFS and obtain the content hash
    NRT_hash← attach content hash with ORT hash
    Create a new resource tree object and get NRT_hash
Endif
```

5 Simulations and result

5.1 Experimental setup

The development environment is setup by installing a few prerequisites on the system with the given configuration: IntelCore i5, CPU 2.25 GHz, 8 GB Primary memory, running on Ubuntu 20.04. The prerequisites are Git client, Docker & Docker compose, Go, and Node.js & NPM. On successful installation of all the prerequisites, Hyperledger fabric (a permissioned blockchain platform) version 2.2 LTS is deployed alongside kafka to attain the consensus among the nodes within the blockchain. The smart contract (chaincodes) is implemented in the Go language. In the Hyperledger fabric, chaincodes are used by the client nodes to propose transactions and peer nodes are accountable for executing the chaincodes and achieving consensus with the remaining nodes within the network. To facilitate the interaction web3 JavaScript is used while caliper-benchmarks test tool is employed for simulating the experiment works.

5.2 Results

We evaluated the running time of DC, APC, and TC methods against multiple queries, where concurrent requests are taken as 50, 100, 200, 500, and 1000. The time taken against every request is documented for further analysis and illustrated in Figs. 5, 6, 7, 8, 9, 10, 11 and 12. Registering new consumers incurs a higher time as compared to fetching information about the registered consumers as depicted in Fig. 5. Since registering a new consumer is a part of a transaction while fetching information of a registered consumer does not require any transaction execution. Similarly, computation of the trust score of consumers and the fine imposed on undesired consumers took more time than the fetching trust score and imposed fine as shown in Fig. 6. Additionally, we have assessed the performance of our solution along with work in [22, 39] in terms of the running time of access control contract methods. The running cost of addPolicy, updatePolicy, deletePolicy, and verifyAccess methods of stated models are shown

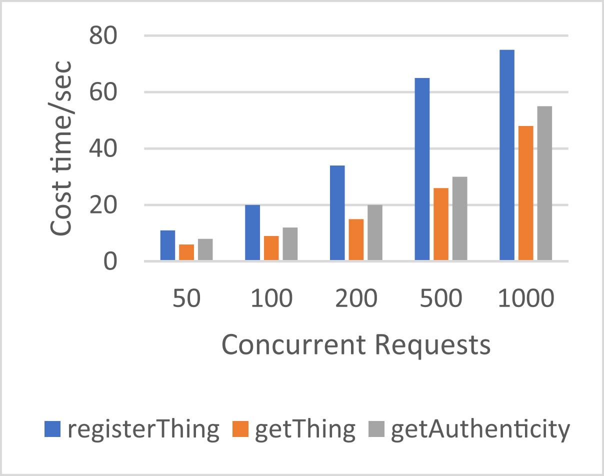


Fig. 5 Running time of DC methods

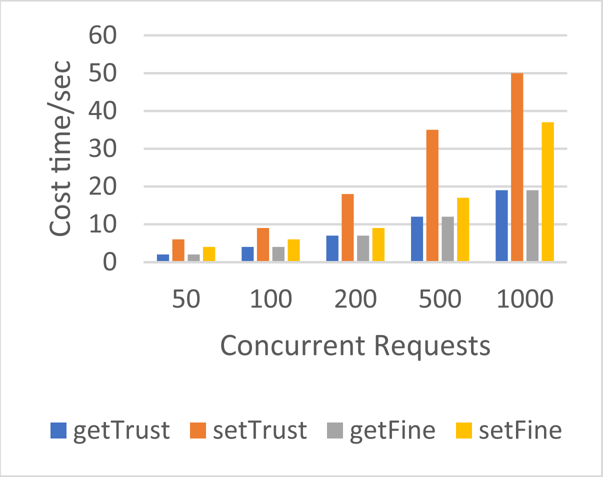


Fig. 6 Running time of TC methods

in Figs. 7, 8, 9 and 10 respectively. The time cost of adding a new policy, updating an existing policy, and removing a policy of our approach is at par with that of [39] while [22] has a higher time cost. However, the cost of verifying the access policy takes more time than in [39] but offers better availability and auditability.

Additionally, the data generation rate of IoT is much higher than the data validation and storage at the Blockchain. Therefore, to address this mismatch IoT data is not directly uploaded on the blockchain rather it is uploaded on IPFS, and a fixed-size hash corresponding to data is recorded on Blockchain only. Figure 12 illustrates that uploading IoT data at IPFS is much faster than uploading it directly on Blockchain. The encryption, decryption, and key generation time is observed against multiple test cases such as 1, 2, 4, and 8 and the relation is depicted in Fig. 11. The key generation time is significantly high while the encryption time is relatively higher than the decryption time.

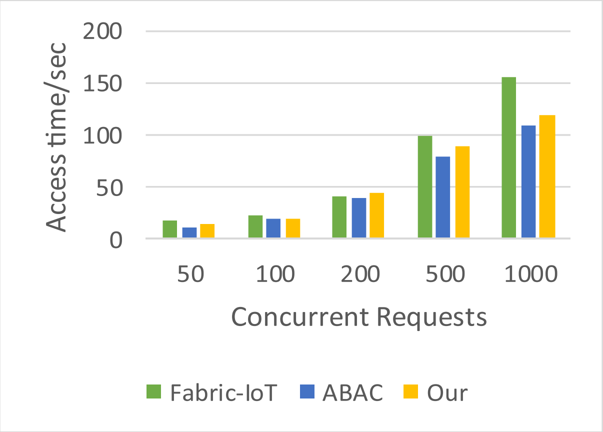


Fig. 7 Running time of addPolicy methods

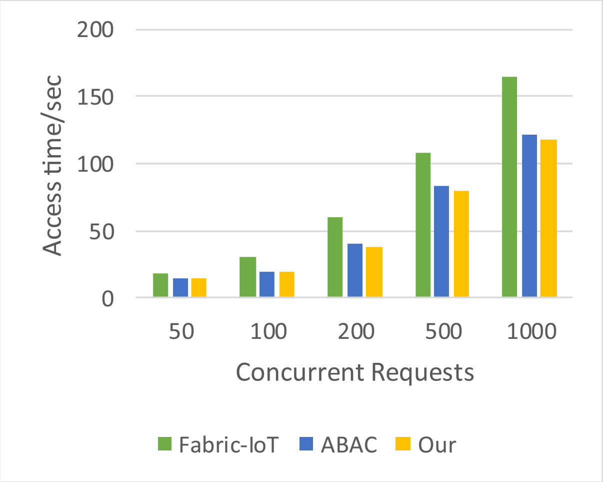


Fig. 8 Running time of updatePolicy methods

5.3 Security analysis

In the proposed architecture, the security parameters emphasized majorly are Availability, Confidentiality, and Integrity.

The Availability is achieved by preventing DoS & DDoS attacks and shifting from centralized storage to distributed storage which in turn prevents censorship of data and is free from the single point of failure. The DoS & DDoS attacks are prevented by enforcing a rule where each communicating entity needs to register itself within the blockchain network via at least one manager. Additionally, the dynamic authorization of the proposed architecture analyzes the past interaction behavior of the resource consumers (RC) and imposes a strict constraint on frequent access requests by those RC's having poor trust scores.

Integrity is attained by preventing the forgery of IoT-generated data. Since cryptographic hash content is uploaded along with the signature of the resource owner (RO) instead of

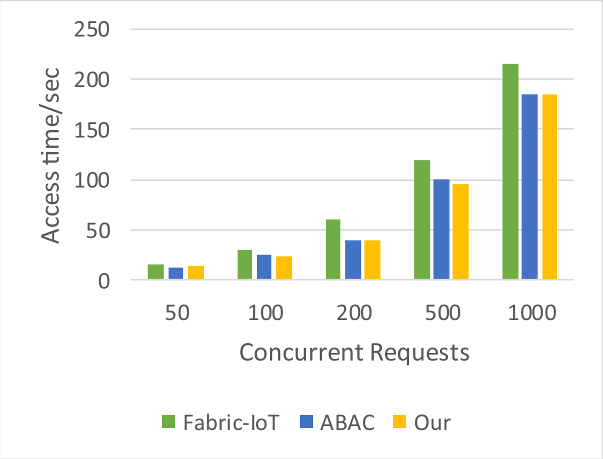


Fig. 9 Running time of deletePolicy methods

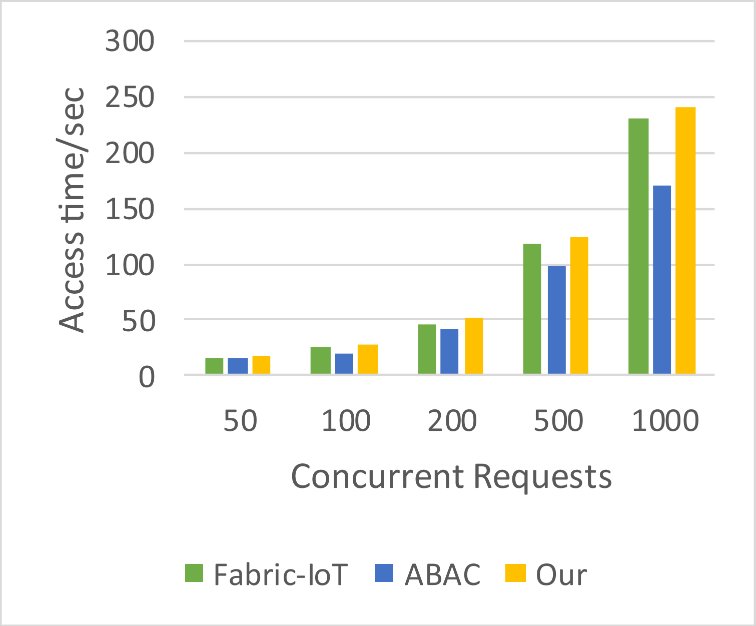


Fig. 10 Running time of verifyAccess methods

original data. Therefore, any RC can use hash and signature to easily validate the origin and integrity of the data.

Confidentiality is achieved by implementing static & dynamic authorization and fine-grained access control by the proposed architecture. In the proposed design, both on-chain & off-chain storage gets only encrypted data and thus realization of confidentiality is easily done.

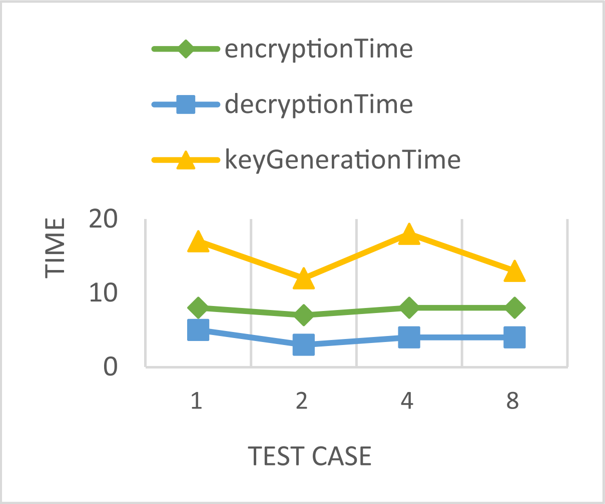


Fig. 11 Pre-processing & Post-processing

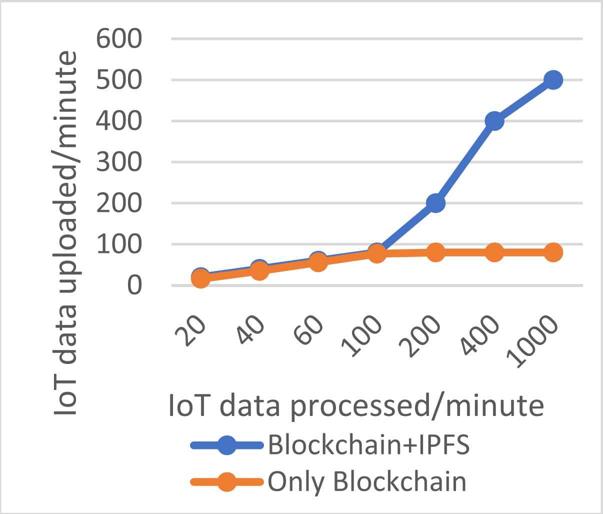


Fig. 12 Storage rate comparison

5.4 Limitations

However, this work has a few limitations that could not be addressed in the current version. The current solution is restricted on a single Blockchain platform (Hyperledger Fabric) which needs to be integrated with hybrid Blockhain networks. The distributed performance of e presented solution is not verified in this work. Additionally, the throughput and reliability of the system were tested on very limited physical devices, therefore, in the next phase; we shall perform more extensive tests.

6 Conclusion

In this paper, we integrated IoT & Blockchain technologies which offer a secure IoT data sharing scheme and supports scalability, flexibility, and resilience along with availability and integrity of IoT data. This blockchain-based and IPFS-enabled scheme is comprised of two phases of authorization (static authorization & dynamic authorization). The static authorization verifies the predefined access policy whereas dynamic authorization computes a trust score of each participating entity (IoT devices and user devices) and compares it against a predefined threshold value. Only on the successful authorization from both phases, a requesting entity is approved to obtain the stated resource. Moreover, the IPFS is incorporated to record the actual IoT data and thus enhancing the availability of the IoT-generated data as against the centralized storage scheme. The dynamic authorization part enforces the DoS and DDoS prevention mechanism while the integration of architecture with IPFS makes it possible to achieve integrity and confidentiality. Experimental results illustrate that uploading IoT data at IPFS is much faster than uploading it directly on Blockchain and the running time of smart contract methods of the proposed solution is fairly less than the prominent work in the same domain. However, the proposed architecture works only on a single Blockchain platform which is not suited to a real IoT ecosystem. In future work, we are planning to improvise this approach to work on hybrid Blockchain networks. Moreover, in future, we intend to test the discussed scenario in a more extensive environment with some malicious behavior simulations. In the upcoming work, we plan to explore how the presented work can be extended for different use cases such as healthcare, supply chain, and drug counterfeiting. Additionally, consensus protocols and energy consumption by the network will be further investigated.

Code availability Code is available.

Authors contribution All authors read and approved the final manuscript.

Funding The authors declare that they have competing interests and funding.

Data availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflict of interests On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Alphand O, Amoretti M, Claeys T, Dall'Asta S, Duda A, Ferrari G, Rousseau F, Tourancheau B, Veltri L Zanichelli F (2018) IoTChain: a blockchain security architecture for the internet of things. In: 2018 IEEE wireless communications and networking conference (WCNC) IEEE pp 1-6.
2. Atlam HF, Alassafi MO, Alenezi A, Walters RJ, Wills GB (2018) XACML for building access control policies in internet of things. In IoTBDS. pp. 253-260
3. Benet J. (2014) "IPFS-content addressed, versioned, P2P file system."[Online]. Available: https://arxiv.org/abs/1407.3561
4. Bouij-Pasquier I, Ouahman AA, El Kalam AA, de Montfort MO (2015) SmartOrBAC security and privacy in the internet of things. In 2015 IEEE/ACS 12th international conference of computer systems and applications (AICCSA). pp. 1-8
5. Cirani S, Picone M, Gonizzi P, Veltri L, Ferrari G (2014) Iot-oas: an oauth-based authorization service architecture for secure services in iot scenarios. IEEE Sensors J 15(2):1224–1234
6. Cruz-Piris L, Rivera D, Marsa-Maestre I, De La Hoz E, Velasco JR (2018) Access control mechanism for IoT environments based on modeling communication procedures as resources. Sensors. 18(3):917
7. Ding S, Cao J, Li C, Fan K, Li H (2019) A novel attribute-based access control scheme using Blockchain for IoT. IEEE Access 7:38431–38441
8. El Bouanani S, El Kiram MA, Achbarou O, Outchakoucht A (2019) Pervasive-based access control model for IoT environments. IEEE Access 7:54575–54585
9. El-Aziz AA, Kannan A (2013) A comprehensive presentation to xacml. In: Third International Conference on Computational Intelligenceand Information Technology (CIIT 2013). pp. 155–161
10. Gaba P, Raw RS, Mohammed MA, Nedoma J, Martinek R (2022) Impact of block data components on the performance of blockchain-based VANET implemented on hyperledger fabric. IEEE Access 10:71003–71018
11. Gusmeroli S, Piccione S, Rotondi D (2013) A capability-based security approach to manage access control in the internet of things. Math Comput Model 58(5–6):1189–1205
12. Han D, Zhu Y, Li D, Liang W, Souri A, Li KC (2021) A Blockchain-based auditable access control system for private data in service-centric IoT environments. IEEE Transactions on Industrial Informatics
13. Han D et al (2021) A blockchain-based auditable access control system for private data in service-centric IoT environments. IEEE Transac Indust Inform 18(5):3530–3540

14. Hardt D (2012) “The oauth 2.0 authorization framework”, Internet Requests for Comments, RFC Editor, RFC 6749
15. Kaiwen S, Lihua Y (2014) Attribute-role-based hybrid access control in the internet of things. In: Asia-Pacific web conference springer, Cham pp 333-343.
16. Kalam AAE, Baida RE, Balbiani P, Benferhat S, Cuppens F, Deswarte Y, Miege A, Saurel C, Trouessin G (2003) Organization based access control. In: Proceedings POLICY 2003. IEEE 4th international workshop on policies for distributed systems and networks. (pp. 120-131)
17. Kamal M, et al. (2022) Privacy-aware genetic algorithm based data security framework for distributed cloud storage. Microprocessors and Microsystems 94, 104673
18. Kantara Initiative, Inc (2017) “User-managed access (uma),” https://kantarainitiative.org/confluence/display/uma/Home, visited on 5 Apr 2017.
19. Lakhan A (2022) Federated-Learning Based Privacy Preservation and Fraud-Enabled Blockchain IoMT System for Healthcare. IEEE J Biomed Health Inform
20. LakhanA, et al. (2022) Blockchain-Enabled Cybersecurity Efficient IIOHT Cyber-Physical System for Medical Applications. IEEE Transac Netw Sci Eng
21. Li Z, Hao J, Liu J, Wang H, Xian M (2020) An IoT-applicable access control model under double-layer Blockchain. IEEE Transact Circuits Syst II: Expr Briefs 68(6):2102–2106
22. Liu H, Han D, Li D (202
Jan 21) Fabric-iot: a Blockchain-based access control system in IoT. IEEE Access 8:18207–18218
23. Maesa DDF, Mori P, Ricci L (2017) Blockchain based access control. In IFIP international conference on distributed applications and interoperable systems, springer, Cham. pp. 206-220
24. Mishra R, Yadav R (2020) Access control in IoT networks: analysis and open challenges. Available at SSRN 3563077
25. Mishra R, Yadav R, Nath P (2021) Blockchain-Based Decentralized Authorization Technique for Data Sharing in the Internet of Things. 5th International Conference on Information Systems and Computer Networks (ISCON). pp. 1–6 https://doi.org/10.1109/ISCON52037.2021.9702297
26. Novo O (2018) Blockchain meets IoT: an architecture for scalable access management in IoT. IEEE Internet Things J 5(2):1184–1195
27. Oktian YE., Lee SG (2020) BorderChain: Blockchain-based access control framework for the internet of things endpoint. IEEE Access, 9. 3592–3615
28. Ouaddah A, Abou Elkalam A, AitOuahman A (2016) FairAccess: a new Blockchain-based access control framework for the internet of things. Sec Commun Netw 9(18):5943–5964
29. Park J Sandhu R (2002) Towards usage control models: beyond traditional access control. In: Proceedings of the seventh ACM symposium on access control models and technologies, ACM. pp. 57-64
30. Pinno OJ, Gregio AR, De Bona LC (2017) Controlchain: Blockchain as a central enabler for access control authorizations in the iot. InGLOBECOM 2017 E global communications conference (pp. 1-6). IEEE.
31. Pradhan NR et al (2022) A blockchain based lightweight peer-to-peer energy trading framework for secured high throughput micro-transactions. Sci Rep 12(1):14523
32. Pradhan NR et al (2022) A Novel Blockchain-Based Healthcare System Design and Performance Benchmarking on a Multi-Hosted Testbed. Sensors 22(9):3449
33. Putra GD, Dedeoglu V, Kanhere SS, Jurdak R, Ignjatovic A (2021) Trust-based Blockchain authorization for IoT. arXiv preprint arXiv:2104.00832.
34. Razzaq A (2022) Blockchain-based Secure Data Transmission for Internet of Underwater Things, Available at SSRN 4127827
35. Riad K, Yan Z (2017) Multi-factor synthesis decision-making for trust-based access control on cloud. Int J Coopera Inform Syst 26(04):1750003
36. Rizzardi A, et al. (2022) Securing the access control policies to the Internet of Things resources through permissioned blockchain. Concurrency and Computation: Practice and Experience, e6934.
37. Sandhu RS (1998) Role-based access control. In: Advances in computers Elsevier, Vol. 46. pp. 237–286
38. Sciancalepore S, Piro G, Caldarola D, Boggia G, Bianchi G (2017) OAuth-IoT: an access control framework for the internet of things based on open standards. In 2017 IEEE symposium on computers and communications (ISCC) (pp. 676-681). IEEE
39. Shammar EA, Zahary AT, Al-Shargabi AA (2022) An attribute-based access control model for internet of things using hyperledger fabric blockchain. Wirel Commun Mob Comput 2022:1–25
40. Shi N, Tan L, Yang C, He C, Xu J, Lu Y, Xu H (2021) BacS: a blockchain-based access control scheme in distributed internet of things. Peer-to-peer Netw Appl 14(5):2585–2599
41. Singh J, Thakur D, Gera T, Shah B, Abuhmed T, Ali F (2021) Classification and analysis of android malware images using feature fusion technique. IEEE Access 9:90102–90117
42. Siris VA, Dimopoulos D, Fotiou N, Voulgaris S, Polyzos GC (2020 Feb 15) Decentralized authorization in constrained IoT environments exploiting interledger mechanisms. Comput Commun 152:243–251

43. Sisi Z, Souri A (2021) Blockchain technology for energy-aware mobile crowd sensing approaches in Internet of Things. Transac Emerg Telecommun Technol, e4217.
44. Srinivasu PN et al (2021) An AW-HARIS based automated segmentation of human liver using CT images. Comput Mater Contin 69(3):3303–3319
45. Sun S, Du R, Chen S, Li W (2021) Blockchain-based IoT access control system: towards security, lightweight, and cross-domain. IEEE Access 9:36868–36878
46. Sun S, Du R, Chen S, Li W (2021) Blockchain-based IoT access control system: towards security, lightweight, and cross-domain. IEEE Access 9:36868–36878
47. Tamang J, Dieu Nkapkop JD, Ijaz MF, Prasad PK, Tsafack N, Saha A, Kengne J, Son Y (2021) Dynamical properties of ion-acoustic waves in space plasma and its application to image encryption. IEEE Access 9: 18762–18782
48. Viriyasitavat W, Hoonsopon D (2018) Blockchain characteristics and consensus in modern business processes. J Ind Inf Integr 13:32–39
49. Vulli A et al (2022) Fine-Tuned DenseNet-169 for Breast Cancer Metastasis Prediction Using FastAI and 1-Cycle Policy. Sensors 22(8):2988
50. Ye N, Zhu Y, Wang RC, Malekian R, Qiao-Min L (2014) An efficient authentication and access control scheme for perception layer of internet of things. Appl Mathem Inform Sci 8(4):1617
51. Zhang X, Parisi-Presicce F, Sandhu R, Park J (2005) Formal model and policy specification of usage control. ACM Transac Inform Syst Sec (TISSEC) 8(4):351–387
52. Zhang Y, Kasahara S, Shen Y, Jiang X, Wan J (2018) Smart contract-based access control for the internet of things. IEEE Internet Things J 6(2):1594–1605

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

2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS) | 978-1-6654-6216-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAIS56108.2023.10073881

Community Detection using Unsupervised Learning Approach

Akansha Mittal
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
akanshamittal2207@gmail.com

Anurag Goel
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
anurag@dtu.ac.in

***Abstract*—A community is referred to as a set of nodes in a network that has a high degree of connectivity with each other and a low degree of connectivity with other nodes in the same network. Community Detection is a renowned research problem for the past many years. The applications of Community Detection is spread across several domains like social networks, transportation networks, genetic networks, citation networks, web networks etc. In this work, several unsupervised learning techniques namely Louvain Algorithm, K-means clustering Algorithm and Gaussian Mixture Model have been examined to identify communities in social networks. The results demonstrated that the Louvain Algorithm outperforms the other two unsupervised learning techniques.**

***Index Terms*—Community Detection, Louvain Algorithm, K-means Clustering, Gaussian Mixture Model**

I. Introduction

Similar opinions, functions, purposes etc., are common interests or preferences shared among the same group of people or persons called communities. To find similarities or dissimilarities between the communities, community detection is used which has been instrumental in the field of data analytics and marketing. Community detection is an important tool that helps to analyze complicated networks in various domains like computational biology, computational social sciences etc. For example, in the network representing interaction among proteins, community detection identifies the group of proteins that have similar biological functions. Community detection in citation networks explores the interconnections, significance and evolution of various research topics. In social networks like Facebook and Twitter, community detection identifies mutual friends and people with common interests through which the e-commerce companies identify the potential customers for their products and services and plan their marketing strategy accordingly.

Networks can be represented by graphs which are a set of vertices(nodes) linked with edges. Most of the real-life networks are inhomogeneous that consist of various distinct groups. These groups have large number of edges within the group but very few edges exist between different groups. These locally dense connected sub-graphs are known as communities. Figure 1 shows the various communities in a sample network of nodes.

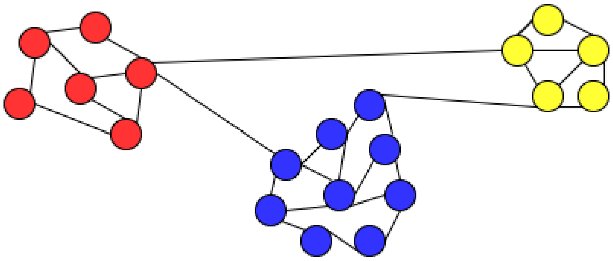


Fig. 1. Communities in a Sample Network

Several community detection techniques have been proposed in past years based on supervised learning and unsupervised learning approaches. In this work, several unsupervised learning approaches are explored to identify the various communities in complex unlabeled networks.

The related work is explained in Section 2 while Section 3 explains the Unsupervised Learning Approaches which are utilized in this work. The experimental setup results are presented in Section 4 and Section 5 concludes the work.

II. Related Work

In [1], the authors proposed an approach to detect communities that are based on the identification of k-plex associated with a particular node. K-plex is a group of n nodes in which each node connects with at least n-k other nodes of the same group. In [2], the authors used N-cliques and K-cores based approaches to identify the communities of telecom customers. N-cliques are the group of nodes in which the maximal distance between each pair of nodes is N and in K-cores each node within the set is connected to k other nodes of the same set.

Over the last two decades, many machine learning and deep learning-based algorithms are proposed for community detection tasks. While some are simple heuristics, like hierarchical clustering or the Girvan-Newman algorithm [4], most of them are optimization techniques in which the maximization of various objective functions is performed. In the Girvan-Newman algorithm [5], initially, the whole network is considered as a single community and further, the network is disjoined into communities in the form of a hierarchy by removing the links which have the largest link betweenness (centrality measure). This process is repeated till each node is a community of its own. Link betweenness is the total count of briefest routes between all the node pairs that include that link. In the Ravasz algorithm [3], initially, each node is taken as a community of its own and then they are merged one by one in the form of a hierarchy by optimizing the modularity and using group similarity. This process is repeated till the whole network is merged into a single community.

In [6], the authors proposed a CLARE model which consists of two modules: Community Locator and Community Rewriter. The potential communities are located by Community Locator while the Community Rewriter refines them. The authors have explored Community detection in complex networks in [17-19].

In [7], Contrastive Clustering (CC) is proposed which performs the clustering at two levels i.e. instance-level clustering and cluster-level clustering. In contrastive clustering, positive and negative instance pairs are made using data augmentation and then the generated data instance pairs are projected in the feature space. The basic idea of Contrastive Clustering is the similarity of positive pairs and negative pairs are maximized and minimized respectively by implementing instance-level clustering and cluster-level clustering in a row and column space respectively.

III. Unsupervised Learning Approaches

Unsupervised learning is used to analyze and cluster the unlabeled data. It analyzes data in search of hidden patterns, makes the group of the most similar data and divides the different data into different groups. In the real world, we don't have labeled data many times so, we need unsupervised learning. In this work, three unsupervised learning approaches are explored namely Louvain Algorithm, Gaussian Mixture Model and K-Means Clustering Algorithm.

A. Louvain Algorithm

Louvain Algorithm is an unsupervised greedy algorithm for detecting communities. It maximizes the modularity of a network. Connection patterns between the nodes should be uniform in any random wired network and they should not depend on the degree of network distribution. For any community of a network, the total count of links within the network should be greater than the expected total count of links in any arbitary network. Modularity measures the quality of each partition and identifies which community partition is best [9].

Consider a network comprising N number of vertices and L links. Consider a partition having total n_c communities. Each community has N_c nodes which are linked to each other by L_c edges where c=1,2,....,n_c. Modularity is interpreted as:

$$M(C_c) = \frac{1}{2L}\sum_{i,j=1}^{N}(A_{ij} - P_{ij})\delta(C_i - C_j) \tag{1}$$

where A_{ij} is the adjacency matrix entry containing the weight of the edge connecting nodes i and j.

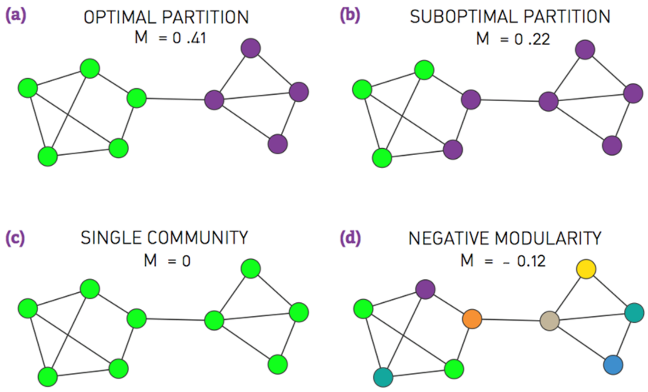


Fig. 2. Modularity for different partitions of a network

If x = 0, $\delta(x)$ = 1, else $\delta(x)$ = 0. It ensures that only the value for nodes that belongs to the same community is added to the value of modularity.

P_{ij} is the total count of links expected between i and j in any arbitrarily connected network.

$$P_{i,j} = \frac{k_i k_j}{2L} \tag{2}$$

where, k_i , k_j are the degree of nodes i and j respectively.

Using the above equations, we can simplify the modularity as follows:

$$M = \sum_{c=1}^{n_c}\left[\frac{L_c}{L} - \left(\frac{k_c}{2L}\right)^2\right] \tag{3}$$

The partition which gives the higher value of modularity is considered a better partition. Figure 2 shows the modularity in different partitions of a sample network. Based on the values of modularity for the various partitions, the partitions are categorized as follows:

1) Optimal partition: The partition which has the largest value of modularity is known as the optimal partition.
2) Suboptimal partition: The partition which has some positive value of modularity is known as the suboptimal partition.
3) Single community: If the whole network is considered as a single partition then the value of modularity comes out to be 0.
4) Negative modularity: Negative modularity is reported when all nodes are assigned to different communities or very dissimilar nodes are assigned to the same community.

Louvain is a hierarchical clustering algorithm. It is divided into two phases: Optimization of Modularity and Aggregation of Community.

1) Modularity Optimization:
In this phase, the nodes are first ordered in a random fashion. After that, each node is removed and inserted in a different community until the modularity value becomes larger than the threshold value [10]. The change

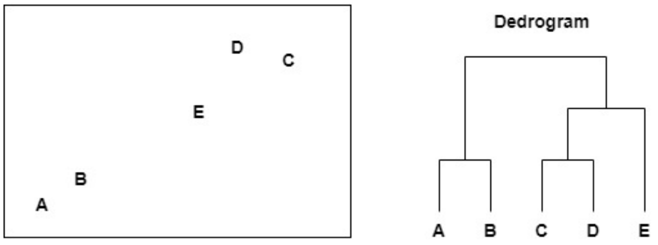


Fig. 3. Dendrogram for partitions of a network

in modularity value when the i^{th} node changes its community can be computed as follows:

$$\triangle M = \left[\frac{\Sigma_{in} + 2w_{i,in}}{2W} - \left(\frac{\Sigma_{tot} + w_i}{2W}\right)^2\right] - \left[\frac{\Sigma_{in}}{2W} - \left(\frac{\Sigma_{tot}}{2W}\right)^2 - \left(\frac{w_i}{2W}\right)^2\right] \tag{4}$$

where, Σ_{tot} is the aggregate weight of the edges connected to the nodes in C.
Σ_{in} is the aggregate weight of the links inside the community C.
W is the aggregate weight of the links in the network.
$w_{i,in}$ is the aggregate weight of the edges from node i to the nodes in community C.
w_i is the aggregate weight of the links from node i.
After simplification,

$$\triangle M = \left[\frac{w_{i,in}}{m} - \left(\frac{2\Sigma_{tot} w_i}{2W}\right)^2\right] \tag{5}$$

2) Community Aggregation:
In the second phase, all the nodes which belong to the same community are combined into a single super node. The weight of the links connecting the supernodes in the network is the aggregate of all the links that were previously attached between the two communities. The supernodes also have self-loops which are the aggregate of all the links inside a community [8].

Figure 3 shows the dendrogram that represents the hierarchy of the communities of a sample network.

Louvain algorithm supports weighted graphs also. The time complexity of the Louvain algorithm is O(nlogn).

B. K-means Clustering Algorithm

In machine learning, It is a very popular clustering algorithm based on unsupervised learning. The unlabeled data is grouped into different clusters using this algorithm. Here, K represents the number of communities identified [11]. For example, if the value of K is 2, it denotes that the data is grouped into two communities.

This algorithm mainly has two steps:

1) Calculation of best K centroids.
2) Each data point will be allocated to its closest K-centroid, and all the points that are near K-center forms a cluster.

The complete K-Means algorithm is shown as Algorithm 1. For step 1 of the algorithm, the Elbow method is used in this work. In Elbow Method [12], a range of k values is considered and for each value of k in the selected range, this algorithm is executed and an average distortion score for all the clusters is computed. The distortion score is considered a performance metric in this work which is the square distance between each data point and its center. The performance metric score for different values of k is plotted in a graph with k values on the x-axis and performance metric score on the y-axis. In the graph, an elbow curve is reported and the value of k where a sharp curve is seen will be chosen as the optimal value of K in K-means clustering.

Algorithm 1 K-means Algorithm

```
1) Select K i.e. the total count of communities.
2) Choose K points or centroids arbitarily.
3) Allocate each data point to its closest centroid, this will
   generate initial K clusters.
4) Find the new centroid based on the different data points
   assigned to that cluster in the previous step.
5) Repeat step 3 and 4 until the convergence is reached i.e.
   no data point changes the cluster in step 4.
```

C. Gaussian Mixture Model

It considers that all the data points are made from the combination of some finite number of Gaussian distributions whose parameters are not known [13]. It is a probabilistic model. It uses the Expectation-Maximization(EM) algorithm to fit a mixture of Gaussian models as shown in Figure 4.

Here, the objective is to achieve the maximum likelihood by adjusting the parameters including means, covariances and mixture coefficients of the Gaussian distributions.

The various steps in Gaussian Mixture Model algorithm are as follows:

1) Initialise the value of means μ_j, covariances Σ_j, and mixing coefficients π_j and find the log-likelihood value.
2) E step: Find the value of responsibilities of each Gaussian distribution by using the current parameters using the following formula:

$$\gamma_k(a) = \left(\frac{\pi_k \; N\left(a \;\mid\; \mu_k,\; \Sigma_k\right)}{\sum_{j=1}^{k} \; \pi_j \; N\left(a \;\mid\; \mu_j,\; \Sigma_j\right)}\right)$$

3) M step: Estimate the value of parameters again using the value of responsibility that we obtained in step 2.

$$\mu_j = \frac{\sum_{i=1}^{N} \gamma_j(a_i) a_i}{\sum_{i=1}^{N} \gamma_j(a_i)} \tag{6}$$

$$\Sigma_j = \frac{\sum_{i=1}^{N} \gamma_j(a_i)(a_i - \mu_j)(a_i - \mu_j)^T}{\sum_{i=1}^{N} \gamma_j(a_i)} \tag{7}$$

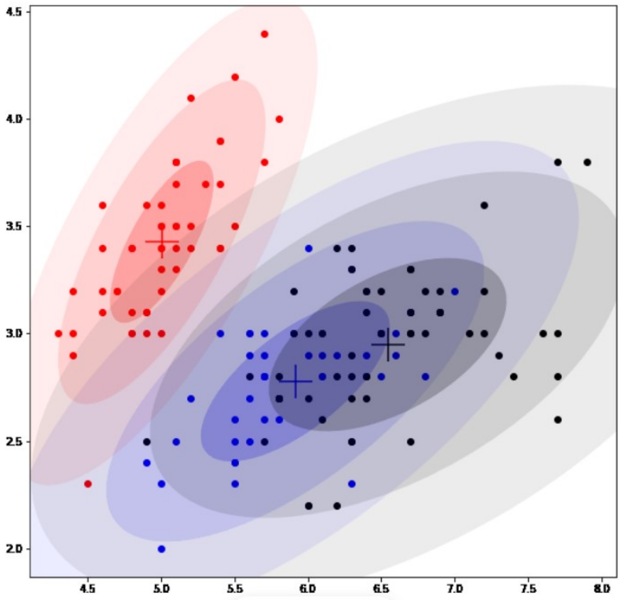


Fig. 4. Combination of Gaussian distribution

$$\pi_j = \frac{1}{N} \sum_{i=1}^{N} \gamma_j(a_i) \tag{8}$$

4) Calculate the value of log-likelihood.

$$lnp(A|\mu, \Sigma, \pi) = \sum_{n=1}^{N} ln \left\{ \sum_{k=1}^{K} \pi_k N(a_n|\mu_k, \Sigma_k) \right\} \tag{9}$$

5) If convergence is reached, STOP. Else, goto step 2.

IV. Experiments and Results

In this work, three unsupervised learning-based approaches namely Louvain Algorithm, K-Means Clustering, and Gaussian Mixture Model are implemented and compared for community detection.

A. Dataset

The dataset used in this work is DataCo Smart Supply Chain for Big Data Analysis [14]. This dataset consists of 180519 data instances with 54 features. This dataset contains structured data. Provisioning, Sales, Production and Commercial Distribution are some of the most important registered activities. For the data preprocessing, we have removed null and duplicate values and extracted two features namely Category Name and Order Region. The communities are identified as categories of those products which have the same region of orders.

B. Experimental Setup

The algorithms are implemented in Python and executed on a system code having an Intel Core i5 processor with 8 GB RAM running on Windows 10.

C. Performance Metrics Used

We have used three performance metrics to examine the effectiveness of the communities detected in the various algorithms which are as follows:

1) Calinski-Harabasz Index:
It was introduced by Calinski and Harabasz in 1974. It is also called the Variance Ratio Criterion. This performance metric is used when ground truth labels are unknown. It measures how similar a node is to its community (cohesion) in contrast to others (separation). Cohesion is calculated as the distance between the nodes in a community to its community centroid, while separation is computed as the distance between the community centroid and the global centroid. [15]
For a dataset D =[a_1 , a_2 , a_3 , ... a_N], CH index for K number of communities is described as follows:

$$CH = \left(\frac{\sum_{j=1}^{K} n_j \, ||c_j - c||^2}{K-1} \right) \; / \; \left(\frac{\sum_{j=1}^{K} \; \sum_{i=1}^{n_k} \; ||a_i - c_j||^2}{N-K} \right) \tag{10}$$

where, n_k is the total count of data points of k^{th} cluster, c_k is the total count of centroids of k^{th} cluster, N is the total count of data points and c is the global centroid of the whole dataset.

2) Silhouette Coefficient:
It is also called the Silhouette score [16]. It is used to examine how accurate the community detection technique is. The value ranges from -1 to +1. where, the value near 1 represents that the communities can be distinguished very clearly and the value near 0 represents that the distances between different communities are not significant. A silhouette Coefficient value near 0 represents that the communities are not correct. The formula of Silhouette Coefficient is as follows:

$$SilhoutterScore = (q-p) \, / \max(p,q) \tag{11}$$

where,
p is the mean intra-community distance
q is the mean inter-community distance

3) Davies-Bouldin score
It is the measure of mean similarity of each community with the community which is similar to that cluster [15]. Here, Similarity is the ratio of within-cluster distance to the between-cluster distance. For better quality of communities, a lower Davies-Bouldin Score is desired. The minimum value of Davies-Bouldin Score is zero. The formula of Davies-Bouldin Score is as follows:

$$DB = \frac{1}{n_c} \sum_{i=1}^{n_c} Q_i \tag{12}$$

where,

$$Q_i = \max_{j=1...n_c, i \neq j} (Q_{ij}), \quad i = 1......n_c \tag{13}$$

where,

$$Q_{ij} = \frac{s_i + s_j}{d_{ij}} \tag{14}$$

where, s_i is the mean distance between the centroid of the community and each data point of that community which is also called cluster diameter. d_{ij} is the distance between centroids of community i and j.

D. Result Analysis

Table I shows the total count of communities that are identified by the algorithms used. The results of all three unsupervised learning approaches used are shown in Table II. From the results, it can be interpreted that the Louvain Algorithm outperforms the K-means clustering Gaussian Mixture Model.

Algorithm	Number of Clusters Identified
Louvain Algorithm	3
K-Means Clustering	2
Guassian Mixture Model	3

TABLE I
NUMBER OF CLUSTERS IDENTIFIED IN EACH ALGORITHM

Algorithm	Calinski Harbasz Score	Silhoutte Score	Davies Bouldin Score
Louvain Algorithm	**303.567**	**0.835**	**0.316**
K-Mean Clustering	248.880	0.715	0.424
Gaussian Mixture Model	139.893	0.507	0.747

TABLE II
PERFORMANCE COMPARISON

Figure 5, Figure 6 and Figure 7 show the communities identified by the Louvain Algorithm, K-means clustering Gaussian Mixture Model respectively.

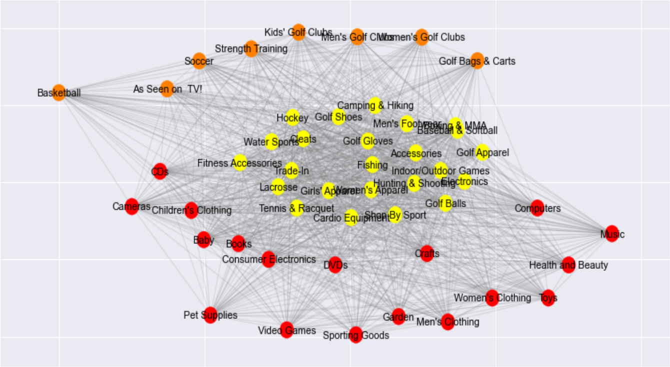


Fig. 5. Communities identified using Louvain Algorithm

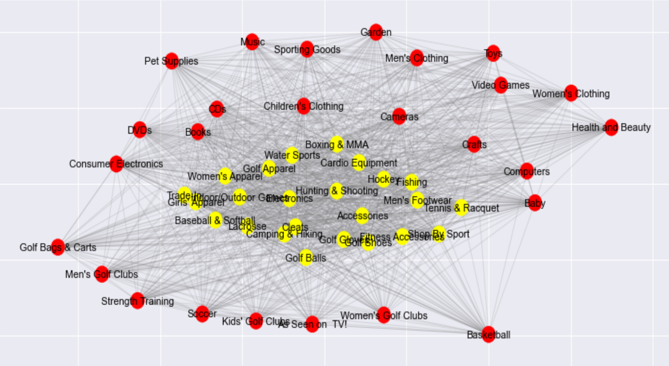


Fig. 6. Communities identified using K-means Clustering Algorithm

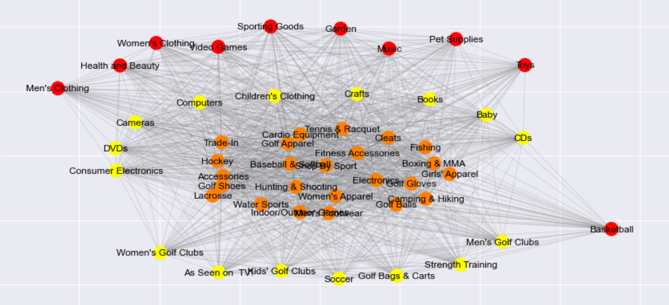


Fig. 7. Communities identified using Gaussian Mixture Model

Figure 8 shows the results of the Elbow method used to identify the optimal total count of communities in K-means clustering in our experiments. As can be seen in Figure 8, the optimal value of K comes out to be 2 for K-means clustering.

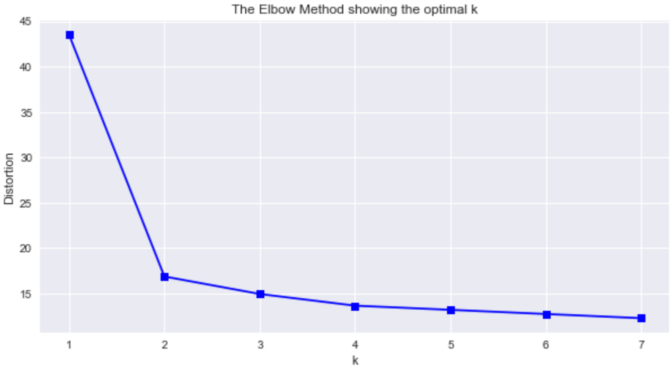


Fig. 8. Elbow method to show the optimal value of k

V. CONCLUSION

Community Detection has a plethora of applications across various domains like social networks, web networks, transportation networks, genetic networks etc. There are several supervised as well as unsupervised learning-based approaches proposed by researchers for community detection in past many years. In this work, three unsupervised learning-based techniques namely Louvain Algorithm, K-means clustering

Gaussian Mixture Model have been examined for detecting the various communities in social networks. The results demonstrated that the Louvain Algorithm outperforms the other two unsupervised learning techniques.

REFERENCES

[1] Wang, Yue, Xun Jian, Zhenhua Yang, and Jia Li. "Query optimal k-plex based community in graphs." Data Science and Engineering 2, no. 4 (2017): 257-273.

[2] Fried, Yael, David A. Kessler, and Nadav M. Shnerb. "Communities as cliques." Scientific reports 6, no. 1 (2016): 1-8.

[3] Ravasz, Erzsébet, Anna Lisa Somera, Dale A. Mongru, Zoltán N. Oltvai, and A-L. Barabási. "Hierarchical organization of modularity in metabolic networks." science 297, no. 5586 (2002): 1551-1555.

[4] Girvan, Michelle, and Mark EJ Newman. "Community structure in social and biological networks." Proceedings of the national academy of sciences 99, no. 12 (2002): 7821-7826.

[5] Newman, Mark EJ, and Michelle Girvan. "Finding and evaluating community structure in networks." Physical review E 69, no. 2 (2004): 026113.

[6] Wu, Xixi, Yun Xiong, Yao Zhang, Yizhu Jiao, Caihua Shan, Yiheng Sun, Yangyong Zhu, and Philip S. Yu. "CLARE: A Semi-supervised Community Detection Algorithm." In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 2059-2069. 2022.

[7] Li, Yunfan, Peng Hu, Zitao Liu, Dezhong Peng, Joey Tianyi Zhou, and Xi Peng. "Contrastive clustering." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 10, pp. 8547-8555. 2021.

[8] Que, Xinyu, Fabio Checconi, Fabrizio Petrini, and John A. Gunnels. "Scalable community detection with the louvain algorithm." In 2015 IEEE International Parallel and Distributed Processing Symposium, pp. 28-37. IEEE, 2015.

[9] http://networksciencebook.com/chapter/9

[10] Brandes, Ulrik, Daniel Delling, Marco Gaertler, Robert Gorke, Martin Hoefer, Zoran Nikoloski, and Dorothea Wagner. "On modularity clustering." IEEE transactions on knowledge and data engineering 20, no. 2 (2007): 172-188.

[11] Hartigan, John A., and Manchek 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.

[12] Cui, Mengyao. "Introduction to the k-means clustering algorithm based on the elbow method." Accounting, Auditing and Finance 1, no. 1 (2020): 5-8.

[13] Reynolds, Douglas A. "Gaussian mixture models." Encyclopedia of biometrics 741, no. 659-663 (2009).

[14] https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis (accessed Nov. 30, 2022)

[15] Maulik, Ujjwal, and Sanghamitra Bandyopadhyay. "Performance evaluation of some clustering algorithms and validity indices." IEEE Transactions on pattern analysis and machine intelligence 24, no. 12 (2002): 1650-1654.

[16] Aranganayagi, S., and Kuttiyannan Thangavel. "Clustering categorical data using silhouette coefficient as a relocating measure." In International conference on computational intelligence and multimedia applications (ICCIMA 2007), vol. 2, pp. 13-17. IEEE, 2007.

[17] Kumar, S., Hanot, R. (2021). Community Detection Algorithms in Complex Networks: A Survey. In: Thampi, S.M., Krishnan, S., Hegde, R.M., Ciuonzo, D., Hanne, T., Kannan R., J. (eds) Advances in Signal Processing and Intelligent Recognition Systems. SIRS 2020. Communications in Computer and Information Science, vol 1365. Springer, Singapore. https://doi.org/10.1007/978-981-16-0425-6-16.

[18] Kumar, S., Panda, B.S. Aggarwal, D. Community detection in complex networks using network embedding and gravitational search algorithm. J Intell Inf Syst 57, 51–72 (2021). https://doi.org/10.1007/s10844-020-00625-6.

[19] Kumar, S., Mallik, A. Sengar, S.S. Community detection in complex networks using stacked autoencoders and crow search algorithm. J Supercomput 79, 3329–3356 (2023). https://doi.org/10.1007/s11227-022-04767-y.

Contents lists available at ScienceDirect

Fuel

journal homepage: www.elsevier.com/locate/fuel

Full Length Article

Computational optimization of engine performance and emission responses for dual fuel CI engine powered with biogas and Co_3O_4 nanoparticles doped biodiesel

S. Lalhriatpuia [*], Amit Pal [1]

Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India

ARTICLE INFO

Keywords:
RSM
ANN
Co_3O_4 nanoparticles
Biogas
Linseed biodiesel
Optimization

ABSTRACT

Given the ongoing energy crisis, rapid depletion of fossil fuel globally, and emissions generated due to its utilization, this study investigates the combined effects of Engine load (20–100%), Cobalt oxide nanoparticles doped rate (NDR, 0–100 ppm), Linseed biodiesel blend rate (BBR, 0–20%) and Biogas flow rate (BFR, 0.5–1 kg/h) on engine performance and emission outputs. Computational methods such as response surface methodology (RSM) and artificial neural network (ANN) were used to establish a prediction model based on the Design of experiment results. The developed RSM model's F-value indicates engine load as the most significant input variable in deciding the value of output responses, followed by BFR, BBR, and NDR, respectively. The statistical analysis using different evaluation metrics suggests the prediction made by the RSM model is more accurate and reliable than the ANN model. The optimization for the RSM model observed an optimal response at 67.53% engine load, 99.998 ppm NDR, 12.084% BBR and 0.694 kg/h BFR, while an optimal response for the ANN model was observed at 79.11% engine load, 61.06 ppm NDR, 11.17% BBR and 0.776 kg/h BFR. The optimization study findings concluded that an optimal combination of nanoparticles, biodiesel, and biogas could significantly improve CI engine performance and emission responses.

1. Introduction

Over the last few decades, fossil fuels have been used as the primary energy source for many industries. However, due to the limited reserves and the environmental impact of extraction, the increasing energy demand raises sustainability concerns[1]. Also, automotive sectors primarily use fossil fuels as their power source, leading to the conclusion that the major emission emitted are due to the combustion of these fossil fuels. Hence there is a need to investigate alternate fuels to counter the issues on hand [2]. Concerning substituting diesel or supplementing diesel use in CI engines, biogas has emerged as an appealing fuel among the gaseous fuels explored over the years due to its easier production method and widespread accessibility of its feedstock. Utilizing dual fuel in diesel engines allows for the use of raw biogas. Dual fuelling ensures adaptability, allowing an engine to run alone or on diesel with biogas added. There is a clear financial gain because biogas partly replaces diesel[3]. When experimented with biogas, previous studies on CI engines observed a reduction in Brake Thermal Efficiency (BTE)[4,5], improvement in Nitrogen oxides (NO_X)[6,7] and Smoke opacity (SO) emissions[6,7], and an increase in Carbon monoxide (CO)[5,7] and Hydrocarbon (HC)[5]. For optimization purposes, conventional approaches are time and resource-consuming. To counter the barriers observed from traditional optimization approach, statistical analysis in the form of RSM and ANN has been used for modeling, simulation, and optimization purpose[8]. A study to analyze the impacts of CR(16–18), engine load(20–100%), and BFR(1.2–3.2 kg/h) on engine performance (BTE) and emission(SO, CO, HC, NO_X) parameters were carried out in CI engine. A decline in SO and NO_X was reported with an increment in BFR, while the effect observed for CO and HC was vice versa. The optimum engine responses using RSM optimization were observed at 18 CR, 80% engine load, and 2.8 kg/h BFR[9].

Due to its physiochemical characteristics, biodiesel has been extensively studied for its use in diesel engines, making it an appealing choice to use as a diesel fuel alternative. Oils from different sources, such as cooking oils, edible and non-edible oils, can be processed to create biodiesel [10]. A study in CI engine for analyzing the use of Shorea

[*] Corresponding author.
E-mail address: slalhriatpuia_2k17phdme07@dtu.ac.in (S. Lalhriatpuia).

[1] ORCID: 0000-0003-2979-8127.

https://doi.org/10.1016/j.fuel.2023.127892
Received 8 December 2022; Received in revised form 29 January 2023; Accepted 17 February 2023
Available online 11 March 2023

robusta biodiesel blended with diesel in BBR of B10-B40 (10%,20%,30%,40%) observed a decrease in BTE of 1.57%, 4.6%, 5.43%, and 7.39% respectively for each blended fuel because of biodiesel's lesser calorific value(CV) than diesel. An increase in Brake Specific Energy consumption (BSEC) and NO_X was observed with the rise in blend percentage, while the effect observed for CO and HC was vice versa. Higher oxygen availability with an increase in biodiesel blend can be attributed to more NO_X formation and the factor for lesser CO due to better combustion[11]. Research into enhancing engine performance and emissions by including nanoparticles has emerged to minimize the disadvantages of utilizing biodiesel in CI engines. Chicken fat biodiesel (B10, B20, B30) with inclusion of ferrous ferric oxide (Fe_3O_4) nanoparticles (50,100,150 ppm) was studied for its impact on CI engine outputs. Better engine performance and lower engine emission due to improved combustion characteristics were observed for all the nanoparticles blended fuels. The maximum optimum engine outputs were suggested for B20-NDR 100 ppm blend[12]. Cerium oxide nanoparticle(50–100 ppm) -infused biodiesel (B0-B20) generated from tyre oil was studied for its use in the CI engine. A drop in BTE of 0.6%, 1.5%. 2% and 2.5% for B5, B10, B15, and B20 were observed compared to pure diesel run due to the lower CV of biodiesel. An increment in BTE of 1.4% and 2% was observed for the B5 blend with 50 ppm and 100 ppm, respectively, due to the catalytic effect of nanoparticles, better atomization, and lower viscosity of nanoparticle fuel blends. A decline in SO emission of 6% was also reported for B5-100 ppm blended fuel in comparison to neat diesel run due to enhanced ignition properties[13]. Waste cooking biodiesel(B5-B10) with the addition of Alumina nanoparticles (30–90 ppm) was used for an ANN study in the CI engine. The study observed increment in BTE of 10.63% and a decrease in HC of 20.56% for the optimal fuel blend B10AL90. The ANN model could predict engine responses with high accuracy (R values > 0.95). It has also been observed that the incorporation of nanoparticles leads to a reduction in the amount of fuel used[14]. An RSM study in a CI engine with input parameters of Fusel oil derived Biodiesel blend(B0-B20), engine speed (1800–2600 rpm), and Biochar nanoparticle (25–125 ppm) observed improvement in NO_X and HC of 20.51% and 14.6% respectively, while setback in CO of 33% was regarded as opposed to the diesel run. The optimized engine performance and emission outputs were monitored at B10 blend, 2300 rpm engine speed, and 100 ppm NDR[15].

India's climate and soil characteristics are ideal for growing linseed crops. For the financial year 2021–22, India observed 120,000 tonnes of Linseed oilseed production and 28,000 tonnes of linseed oil production [16,17]. Hence linseed is chosen for biodiesel production in this study. Previous studies suggest biogas, when utilized as dual fuel without enrichment, can substitute up to 87.5% of diesel [6]. Also, From the above literature survey, nanoparticles in biodiesel have been shown to enhance CI engine parameters. No research has been done on optimizing the addition rate of nanoparticles, biodiesel, and biogas together in a CI engine. Hence in this study, the input variables of engine load, NDR, BBR, and BFR are optimized for their effect on engine performance and emission outputs using ANN and RSM computational methods.

2. Materials and methods

2.1. Nanoparticles properties

Cobalt oxide (Co_3O_4) was procured from Sigma Aldrich company. The morphological structure of nanoparticles is determined using a Zeiss scanning electron microscope (SEM) EVO 50. The SEM image shows non-uniform, larger agglomerated clusters with primarily round and oval shapes. The magnetic induction between particles causes agglomeration[18]. The images from Fig. 1(a) confirmed the nanoparticles have an average size lesser than 100 nm.

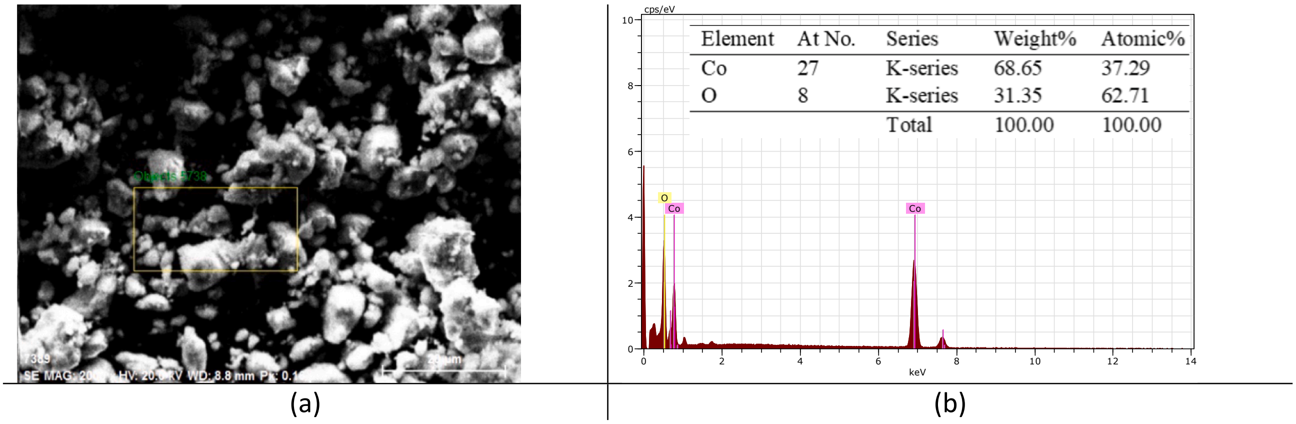
Utilizing the RONTEC EDX system model Quantax 200, the elemental composition of Co_3O_4 nanoparticles is identified. The presence of Co and O has been established in Fig. 1(b), and the absence of foreign element peaks indicates the nanoparticles utilized are of the highest purity. The Co_3O_4 theoretical ratio of 3:4 is also confirmed by the close proximity in observing atomic% of Co and O at 37.29% and 62.71, respectively. Co and O are evenly distributed across the lattice, according to the elemental mapping pictures in Fig. S1.

2.2. Biodiesel production and properties

In this study, raw linseed oil is converted to biodiesel by the transesterification method. Raw linseed oil was heated for moisture removal at 105–110 °C for 10 min. 20% v/v of Methanol and 1% KOH w/w were mixed homogenously, and this mixture was added to the linseed oil after the oil cooled down to 45 °C. The mixture is then stirred and heated to 50–55 °C for the transesterification process using a magnetic stirrer for 90–100 min. The mixture is then fed to a conical separator and let to rest for 12 h. The lower layer forming glycerol is removed thereafter. Water washing is then done with the linseed oil to remove the catalyst and impurities. The oil is heated to 110 °C for 2–5 min to eliminate the moisture content. The final product obtained biodiesel is then stored in a container for testing and blending purposes.

Agilent 8890 Gas chromatography-mass spectrometry(GC–MS) was used to validate the biodiesel synthesis and determine the chemical composition. The analysis observed 18 fatty acid methyl ester(FAME) peaks, as given in Fig. 2. The peaks were observed at unique acquisition time, as mentioned in Table 1. The fatty acid observed for the biodiesel prepared in this study adheres to the commonly observed fatty acid for most biodiesel tested[19].

Biodiesel is blended with diesel in a proportion of 10:90 and 20:80 for L10 and L20 blends, respectively. The measured physiochemical



Element	At No.	Series	Weight%	Atomic%
Co	27	K-series	68.65	37.29
O	8	K-series	31.35	62.71
		Total	100.00	100.00

(a)

(b)

Fig. 1. (a) SEM image of Co_3O_4 at 20 μm (b)EDX spectra of Co_3O_4.

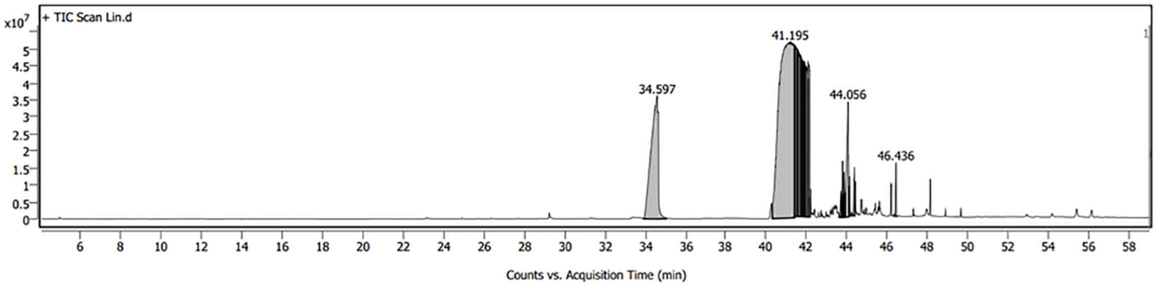
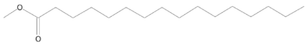
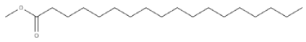
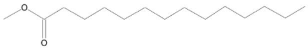
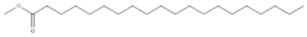
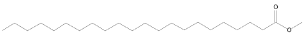


Fig. 2. GC–MS of Linseed oil.

Table 1
Composition of Linseed oil.

Sl. No.	Fatty acid	Peak area %	Acquisition time(min)	Fatty acids	Chemical Formula	Chemical structure
1.	Palmitic acid, methyl ester	14.87	34.597	Saturated	$C_{17}H_{34}O_2$	
2.	Linoleic acid, methyl ester	77.03	41.195, 41.412, 41.549, 41.584, 41.698, 41.767, 41.881, 41.904, 41.961, 44.056, 44.113	Unsaturated	$C_{19}H_{34}O_2$	
3.	Stearic acid, methyl ester	4.66	42.076	Saturated	$C_{18}H_{38}O_2$	
4.	Myristic acid, methyl ester	2.19	43.707, 43.787, 43.855	Saturated	$C_{15}H_{26}O_2$	
5.	Eicosanoic acid, methyl ester	0.73	44.37	Saturated	$C_{21}H_{42}O_2$	
6.	behenic acid methyl ester	0.50	46.43	Saturated	$C_{23}H_{46}O_2$	

properties of the diesel(D100), raw linseed oil(LO), and pure biodiesel (L100) are given in Table S1.

2.3. Nanoparticles blends preparation and properties

Nanoparticles and diesel/biodiesel were blended for one hour in an ultrasonication bath at 50 ppm(N10) and 100 ppm(N100) concentrations. The blend undergoes another round of mixing using an ultrasonicator probe at 50 Hz for 30 min. To minimize agglomeration caused by surface tension, 1% by weight of Triton X-100 surfactant was used. After 24 h of observation, there was no agglomeration or particle settling. The properties of the biodiesel blends and nanoparticles doped fuels employed in the study were then tested using multiple ASTM testing procedures, as shown in Table 2.

Because biodiesel is made from vegetable oils and animal fats, which are more viscous than diesel fuel's hydrocarbons, it has a higher viscosity and density. Due to increased surface tension, nanoparticles in biodiesel enhance their kinematic viscosity[20,21]. Higher oxygen concentration in biodiesel reduces CV compared to diesel. Nanoparticles in liquid fuels operate as catalysts to break down chemical bonds and release more energy, increasing CV[22,23]. The fuel ignites at its flash point when exposed to a spark or flame. Biodiesel's higher flash point makes it safer than diesel. ASTM D-97 defines the pour point as the temperature at which fuel stops flowing when cooled. The pour point of all the blended fuels ranges from −14 to −16 °C, indicating that the fuels are suitable for use in both mild cold weather and tropical climates. The Flash point of nanoparticle-blended fuels was seen to increase slightly while the pour point decreased. The Cetane number measures diesel fuel oil's igniting capability by comparing it to reference fuels in a regulated engine test. High Cetane numbers improve combustion by reducing fuel ignition delay. Because of nanoparticles' high surface-to-volume ratio, which enhances combustion and reduces ignition delay, cetane number increased with larger NDR. Biodiesel and nanoparticle blended fuels can be utilized in CI engines with no alteration to the engine due to their fuel properties similar to diesel[24,25].

2.4. Biogas properties

The composition of biogas varies depending on the feedstock and production factors; the two primary components are carbon dioxide and methane. In addition, there is hydrogen sulfide (H_2S) traces in their raw form. The biogas used for this study was produced using kitchen waste as its feedstock. Raw biogas was passed through an Iron sponge medium for the removal of H_2S. The final gas obtained composition(69% CH_4 and 24% CO_2) is presented in Table S2, and the calorific value was measured as 26 MJ/kg. Biogas composition is measured using a biogas analyzer (Make: Nunes VTPBGA-003), and the calorific value is measured using Junkers calorimeter (Make: Aditya RAP-147B).

2.5. Experimental setup and methodology

The experiment utilised a 3.5 kW rated power, single-cylinder, four-

Table 2
Physiochemical properties of fuel blends.

Properties	Unit	Testing methods	D100N50	D100N100	L10	L10N50	L10N100	L20	L20N50	L20N100
Kinematic Viscosity (40 °C)	cSt	D-445	3.07	3.16	3.15	3.25	3.35	3.36	3.46	3.56
Density (40 °C)	kg/m3	D-1298	843	845	828.3	830	832	834.2	836	838
Calorific Values	MJ/kg	D-240	44.1	44.25	43.5	43.63	43.83	43.03	43.16	43.27
Flash Point	°C	D-93	61	62	90	91	93	107	109	110
Pour Point	°C	D-97	−14	−15	−14	−14	−15	−15	−16	−16
Cetane Number	–	D-4737	51.9	52.6	50	50.8	51.5	49	49.7	50.4

stroke, constant-speed CI engine. An eddy current type dynamometer was employed to apply engine load. The engine load for this study ranges from 2.4 kg (20%) to 12 kg (100%). Rotameters were utilized to measure the flow of the cooling water. Airflow and fuel flow are measured using a manometer and a fuel flow meter, respectively. At various engine cylinder locations, temperature sensors(PT100) are installed to continuously gather data. These sensors are coupled with an NI unit, which records signals and transmits them to the computer using Engine Soft software. To enable dual fuel operation, a gas-air mixing device is incorporated to the setup. The gas-air mixing system was built using the data from engine specifications. Fig. 3 provides a schematic layout of the test setup and gas mixer based on our prior studies[3].

The BFR was measured using a biogas flow meter (Siya SI 2.5), and the change in gas flow rate was facilitated by modulating the ball valves. NO_X, HC, and CO emissions were assessed using a gas analyzer (Make: AVL DiGas 480), and smoke opacity was measured using a smoke meter (Make: AVL 437C).

2.6. Design of experiment (DOE)

Design Expert Software was utilized to create the DOE by applying the Central Composite Face-Centered Design (CCFCD) to the chosen input variables and output parameters. Since there are axial points on every face in the design domain, the model's alpha value is 1. CCFCD is the term for the Central Composite Design (CCD) with alpha equal to 1. For 16 factorial points, 8 axial points and 6 replicates, the total number of runs for CCFCD design translates to 30. The independent input variables selected and their coded levels are mentioned in Table 3.

2.7. RSM modelling

RSM was evaluated by fitting a second-order polynomial equation (Equation (1)) to the DOE experimental data to establish a correlation between independent inputs and dependent outputs.

$$Y = b_0 + \sum_{i=1}^{4} b_i x_i + \sum_{i=1}^{4} b_{ii} x_i^2 + \sum_{i \le 1 \le j}^{4} b_{ij} x_i x_j \qquad (1)$$

Y is the dependent output on the independent input factor of x_i. b_0

Table 3
Input Parameters and levels.

Input Parameters	Units	Symbol	Levels		
			−1	0	1
Engine load	%	Load	20	60	100
Nanoparticles doped rate	ppm	NDR	0	50	100
Biodiesel blend rate	%	BBR	0	10	20
Biogas Flow rate	kg/h	BFR	0.5	0.75	1

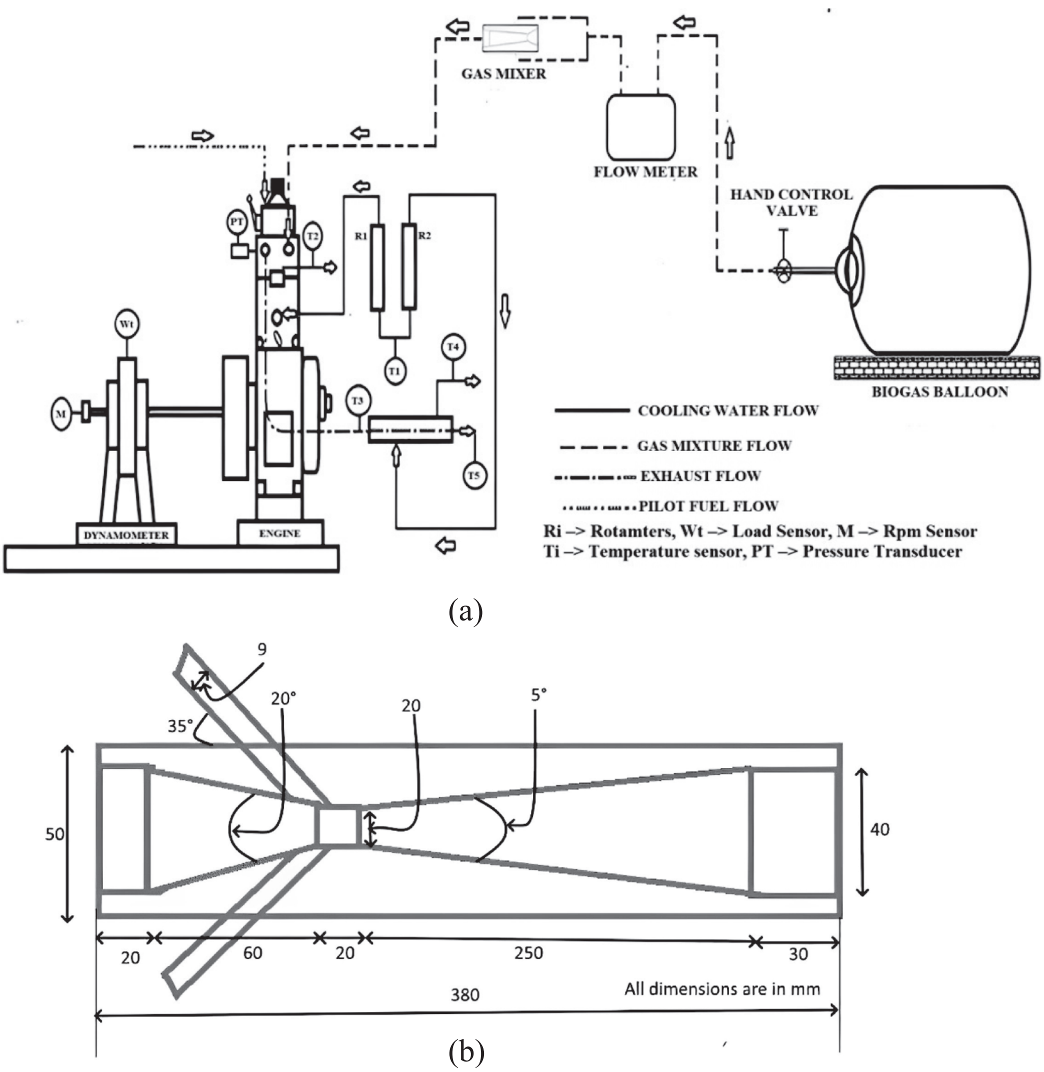


Fig. 3. Schematic layout of (a) the test setup; and (b) gas mixer.

represents the intercept coefficient of the polynomial equation. b_i represents the linear coefficients for Load, NDR, BBR, and BFR, while b_{ii} is the interactive coefficients, and b_{ij} is the quadratic coefficients. The model fitting was achieved through Design Expert software. Analysis of variance (ANOVA) was used to establish model reliability and the contribution significance of each input variable in determining output responses. The interaction impact of input factors on output was studied using a 3D response surface plot. For each 3D plot, two input interaction on output is studied at once, while the other input variable is kept constant, which allows visualization of output changes as different combinations of input factors are varied. The plot shows the output response on the vertical axis, while the two input factors that are being studied are plotted on the horizontal axes.

2.8. ANN modelling

ANN is an analytical tool for establishing the data's predictability regression and validating the correlation between input and output variables. In this ANN study, the backpropagation algorithm is utilized for the multilayer perceptron (MLP) neural network architecture, using Matlab R2020a for coding and execution purposes. This MLP architecture as shown in Fig. 4 holds the forms of A-X1-X2-Z, where the number of neurons in the input layer, first hidden layer (HL1), second hidden layer (HL2), and output layer is indicated by A, X1, X2, and Z respectively. The number of hidden layers is set to 2. Optimal neuron combinations for hidden layers can be obtained using the Trial and error method. A network is created for each output to train DOE results. The network is trained with different neuron combinations with each hidden layer neuron ranging from 1 to 10. The neuron combinations exhibiting the least RMSE are then selected as in Table 4. The optimal neuron combination obtained is then used to create a new network corresponding to each output. Levenberg Marquardt(trainlm) was employed as the training function, and Mean square error (MSE) was employed as the performance function. Log-sigmoid and Tan-sigmoid transfer functions were employed for the hidden layer and output layer, respectively. The DOE consists of 30 runs, of which 21 runs (70% of DOE) were used to train the network, four runs (15% of DOE) for testing, and four runs (remaining 15% of DOE) for validation. The interaction impact of different input variables on output is represented in a 2D graph, where the output parameter is represented on the y-axis. One input is represented on the x-axis, and a line or point on the graph represents the second input variable, while a constant third input is used for plotting the 2D graphs.

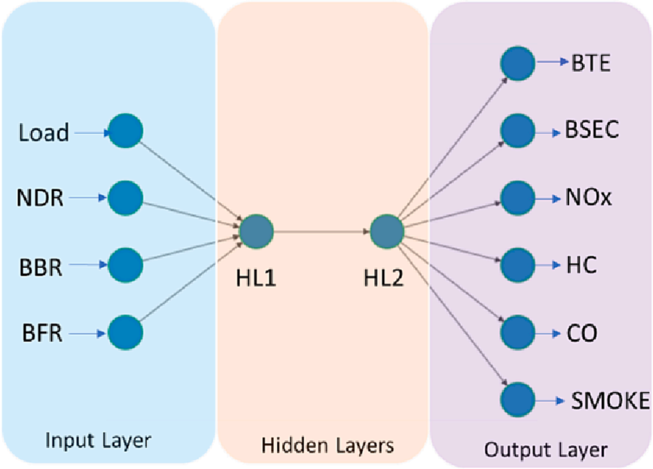


Fig. 4. ANN Architecture (4-X1-X2-6).

Table 4
ANN architecture for engine outputs.

BTE	BSEC	NO_X	HC	CO	SO
4-6-5-6	4-9-3-6	4-9-3-6	4-10-8-6	4-6-6-6	4-4-2-6

2.9. Comparison of the ANN model and RSM model

For each DOE run, a percentage of error is calculated for both the RSM and ANN predicted data and observed data using Equation (2). Coefficient of determination (R^2), Root Mean Square error (RMSE), and Mean Absolute deviation (MAD) are the evaluation metrics chosen to compare the predictive effectiveness of the ANN and RSM model, calculated using Equation. (3),4, and 5, respectively.

$$\text{Percentage of error} = \frac{|\text{Observed result} - \text{Predicted result}|}{\text{Observed result}} \tag{2}$$

$$\text{R2} = 1 - \left(\frac{\sum_{t=1}^{n} (A_t - F_t)^2}{\sum_{t=1}^{n} (F_t)^2} \right) \tag{3}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} (A_t - F_t)^2}{n}} \tag{4}$$

$$\text{MAD} = \frac{\sum_{t=1}^{n} |A_t - F_t|}{\text{n}} \tag{5}$$

where A_t is the actual observed data, F_t is the predicted data, and n denotes overall number of runs in DOE.

2.10. Optimization of RSM model with the desirability approach

A definite optimal solution without compromise is unattainable for more than one output parameter. RSM with desirability function is utilized to obtain the optimal solution for the multi-objective response. In the Desirability approach, the solution with the highest combined desirability factor is considered the optimal solution. The individual desirability(d_n) for the output parameter with the goal to maximize is determined with Equation (6), and for output parameter with the goal to minimize is determined with Equation (7). The combined desirability (CD) is determined as provided in Equation (8).

$$d_n = \left\{ \begin{array}{ll} 0 & n < L_n \\ \frac{n - L_n}{G_n - L_n} \times r_n & L_n < n < G_n \\ 1 & n > G_n \end{array} \right\} \tag{6}$$

$$d_n = \left\{ \begin{array}{cc} 0 & n > H_n \\ \frac{n - H_n}{G_n - H_n} \times r_n & G_n < n < H_n \\ 1 & n < G_n \end{array} \right\} \tag{7}$$

$$CD = [\Pi(\ d_n w_n)]^{\frac{1}{W}} \tag{8}$$

where n, G_n, L_n, H_n, r_n represent predicted, goal value, lower suitable values, higher suitable value, desirability function weight for n^{th} outputs respectively, $w_n = n^{th}$ output importance, W= Σw_n. The desirability factor varies from 0 to 1, with 1 being the most desired ideal option and 0 being the least desired. Each output parameter's goal has been specified. While the other five outputs goals are set to minimize, BTE goal is set to maximize. Each of the output parameter goals and weight was assigned equal importance.

2.11. Optimization of ANN model with genetic Algorithm(GA)

GA is an evolutionary algorithm that is an exploration and investigation method akin to the principle of natural selection. The trained network created corresponding to each output response in the ANN model is used to define a fitness function in GA. Fitness function is an objective function where the goals of either maximizing or minimizing the responses are fed. The input parameters' upper and lower bounds are then provided to the GA program, including the selection parameters listed in Table 5. Initial Population, Fitness function, Selection, Crossover, and Mutation are the five phases that are utilized for GA optimization. GA is accountable for grading and selecting the solution from the ANN-based objective function[26].

2.12. Validation of optimized results from ANN and RSM model

Validation is necessary to assess the correctness of the outcome attained from optimization. An experimental test is conducted for the optimized input parameters of the RSM and ANN model. The percentage of error for experimental results in comparison to the RSM and ANN models predicted optimized response parameters is calculated as per Equation (2).

3. Results and discussions

3.1. Design of experiment analysis

The experimental data observed from DOE runs are given in Table 6. Three repetitions of each experimental run were performed with consistent outcomes. After collecting data from each replicate, the mean value for each experimental condition was determined using a conventional averaging method.

3.2. RSM model analysis

Analysis of Variance (ANOVA) is evaluated by fitting Equation (1) into DOE data. The model formulated in the ANOVA Table S3, and S4 for engine parameter outputs is considered significant since the P-value for all output parameters was less than 0.05 for both the model and existing inputs. High F-values as given in Table 7, implies that it could conclusively interpret the variation in data obtained from the experiments. Engine load offered the highest F-value among the input parameters, followed by BFR, BBR, and NDR. It is construed that engine load has the most substantial influence in deciding the value of output responses, followed by BFR, BBR, and NDR, respectively. Since load is a major contributing factor for outputs, therefore the effects of other input parameters for outputs are taken in reference to engine load and studied using a 3D surface response plot as given in Figs. 5–10. The combined effects of different input parameters(other than engine load) on output are presented in the supplementary Fig. S2 to Fig. S7. High R^2 and Adjusted (adj.) R^2 values were observed for all the output responses, which implies that the model produces significantly similar data compared to the experimental data. The adj. R^2 and Predicted (Pred.) R^2 values had a lesser than 2% disparity for all the responses, indicating good prediction reliability. Tables S3 and S4 show that the lack of fit is insignificant since its *P* value is greater than 0.05. The second-order polynomial equation (Equation (1)) derives the coded relationship for the RSM model formed between the input variables and engine responses, as given in Table S5.

Table 5
Selection Parameters for GA optimization.

Population Type	Initial Population	Mutation Rate	Crossover Fraction	Selection Function
Double Vector	50	0.01	0.8	Tournament

3.2.1. Impact of input factors on BTE for RSM model

BTE indicates the effectiveness of converting fuel energy into mechanical work. The interaction impact of engine load vs. NDR for BTE is depicted in Fig. 5(a). BTE rises as a function of increased engine load., reaching its maximum at the peak load (100%) owing to an increase in fuel supply and a rise in cylinder temperature [27]. With an increment in NDR, fuel's evaporation time lowers, and the ignition delay minimizes, thereby improving the combustion efficiency and increasing BTE. Additionally, a BTE increase was observed for higher NDR fuel blends because of higher CV[28]. At each load, an average increase in BTE of 4% and 8% for NDR 50 and NDR 100 is observed compared to an engine run on non-doped blended fuel. Previous research on biodiesel-Co_3O_4 blend under comparable testing conditions reveals a similar increase in BTE with an increase in NDR. A fish oil biodiesel-Co_3O_4 blend (B20 + 120 ppm NDR) utilized in the CI engine reported a rise in BTE of 8.6% [29,30] The interaction impact of Load vs. BBR for BTE is depicted in Fig. 5(b). Due to higher viscosity and lower CV of linseed biodiesel, BTE declines with a rise in BBR[31]. In comparison to an engine run on non-biodiesel blended fuel, there was an average reduction in BTE of 2.25% and 4% for BBR 10 and BBR 20 at all loads. Previous research on linseed biodiesel under comparable testing conditions reveals a similar decline in BTE with an increase in BBR. A decrease in BTE of 4.63% and 13.82% was observed for an engine operated with BBR 10 and BBR 20 [32]. The interaction impact of Load vs. BFR for BTE is depicted in Fig. 5(c). With an increase in the BFR, BTE declines. The BTE has dropped due to incomplete combustion brought on by a lack of oxygen and the slower flame propagation speed of biogas[33]. Compared to an engine run on 0.5 kg/h BFR blended fuel, there was an average reduction in BTE of 7.2% and 14.4% for 0.75 kg/h BFR and 1 kg/h BFR, respectively. Research conducted under comparable testing conditions with varying BFR reveals a similar decline in BTE with an increment in BFR. A 16.6% decline in BTE was reported for engine run on 1.2 kg/h BFR compared to 0.3 kg/h BFR [34].

3.2.2. Impact of input factors on BSEC for RSM model

The quantity of fuel energy required to generate one kilowatt of output power is measured by BSEC[35]. The interaction impact of engine load vs. NDR for BSEC is depicted in Fig. 6(a). As the engine is subjected to a greater load, the combustion chamber's temperature also increases, resulting in a decrease in BSEC. As the engine's load increases, BSEC tends to decrease due to the catalytic chemical oxidation of nanoparticle fuel blends, enhancing fuel combustion and greater fuel efficiency[36]. Compared to an engine run on non-doped blended fuel, there was an average reduction in BSEC of 3% and 6.5% for NDR 50 and NDR 100, respectively, for all loads. Previous research on biodiesel-Co_3O_4 blend under comparable testing conditions reveals a similar decrease in BSEC with an increase in NDR. A jatropha oil biodiesel-Co_3O_4 blend(B20 + 100 ppm NDR) used in the CI engine reported decrease in BSEC of 7.95%[36,37]. The interaction impact of Load vs. BBR for BSEC is depicted in Fig. 6(b). BSEC declined with an increase in BBR due to lower volatility and lower CV[38]. Compared to an engine run on non-biodiesel blended fuel, there was an average increase in BSEC of 4% and 8% for BBR 10 and BBR 20, respectively, for all loads. Previous research conducted with linseed biodiesel under comparable test conditions revealed a similar increase in BSEC with an increment in BBR. An increase in BSEC of 10% was reported for engine run on BBR 20 [38]. The interaction impact of the Load vs. BFR for BSEC is illustrated in Fig. 6(c). As engine load increases, BSEC decreases for all BFR due to better combustion quality resulting from the combustion temperature and pressure rise. However, as BFR increases, BSEC also increases due to biogas' lower calorific value than diesel, resulting in a higher fuel consumption rate[34]. Compared to an engine run on 0.5 kg/h BFR blended fuel, there was an average increase in BSEC of 7.5% and 15% for 0.75 kg/h BFR and 1 kg/h BFR, respectively, for all loads. Previous research

Table 6
DOE with experimental responses.

Run	A:load	B:NDR	C: BBR	C:BFR	BTE	BSEC	NO_X	HC	CO	SO
	%	ppm	%	kg/h	%	MJ/kWh	ppm	ppm	%	%
1	60	50	10	0.75	15.53	23.27	216.37	63.92	0.0882	29.06
2	20	0	0	0.5	10.44	34.31	114.89	94.23	0.1495	20.47
3	20	0	20	0.5	10.02	37.05	129.83	76.32	0.1345	19.26
4	60	50	10	0.5	16.46	21.86	271.13	53.68	0.0838	34.99
5	20	100	20	0.5	10.62	34.83	122.04	71.74	0.1197	17.14
6	60	50	10	0.75	15.3	23.5	220	66	0.089	30
7	100	0	20	1	17.36	24.39	371.18	64.03	0.1109	37.51
8	100	100	20	0.5	21.86	19.58	473.81	44.26	0.0856	45.21
9	60	100	10	0.75	15.97	22.52	211.51	62.49	0.0863	27.23
10	100	0	20	0.5	20.43	21.06	509.47	47.59	0.0996	50.80
11	60	0	10	0.75	14.52	25.02	243.11	71.82	0.0999	30.59
12	20	100	0	1	9.28	38.71	77.014	123.85	0.1446	13.45
13	60	50	10	1	14.32	26.16	191.23	75.60	0.0933	25.83
14	100	100	0	0.5	23.13	18.13	419.30	49.17	0.0952	48.59
15	20	0	0	1	8.68	41.63	83.71	134.62	0.1625	15.12
16	100	50	10	0.75	20.77	20.57	373.47	52.32	0.0968	42.18
17	60	50	10	0.75	15.8	23	212	60	0.087	28
18	20	100	20	1	8.91	41.81	87.02	100.31	0.1301	12.66
19	60	50	20	0.75	14.93	24.16	224.54	60.37	0.0808	27.90
20	60	50	10	0.75	15.4	23.4	218	64	0.0885	29.5
21	20	0	20	1	8.33	44.96	94.59	109.04	0.1462	14.22
22	100	0	0	1	18.38	22.59	328.48	71.15	0.1233	40.32
23	100	100	0	1	19.85	20.78	298.92	64.74	0.1060	35.88
24	100	100	20	1	18.75	22.44	337.78	58.27	0.0954	33.39
25	60	50	10	0.75	15.6	23.1	215	63.92	0.0882	29
26	60	50	0	0.75	15.88	22.37	204.12	71.03	0.091	29.75
27	100	0	0	0.5	21.62	19.5	450.86	52.88	0.1107	54.6
28	60	50	10	0.75	15.45	23.2	219	65	0.0886	29.5
29	20	50	10	0.75	9.78	37.70	97.26	101.13	0.1363	15.85
30	20	100	0	0.5	11.06	32.25	108	88.57	0.1330	18.22

Table 7
ANOVA F-value for performance and emission responses.

Source	BTE F-value	BSEC F-value	NO_X F-value	HC F-value	CO F-value	SO F-value
Model	1427.07	2144.67	1220.70	246.86	1191.07	1405.3
A-Load	18517.73	22443.87	15030.67	2101.57	6331.88	16949.80
B- NDR	191.95	356.87	77.94	45.63	1136	280.15
C-BBR	103.28	378.47	150.39	188.09	725.42	96.67
D-BFR	970.84	1903.27	1139.68	669.39	579.28	1892.86
R^2	0.9992	0.9995	0.9991	0.9957	0.9991	0.9992

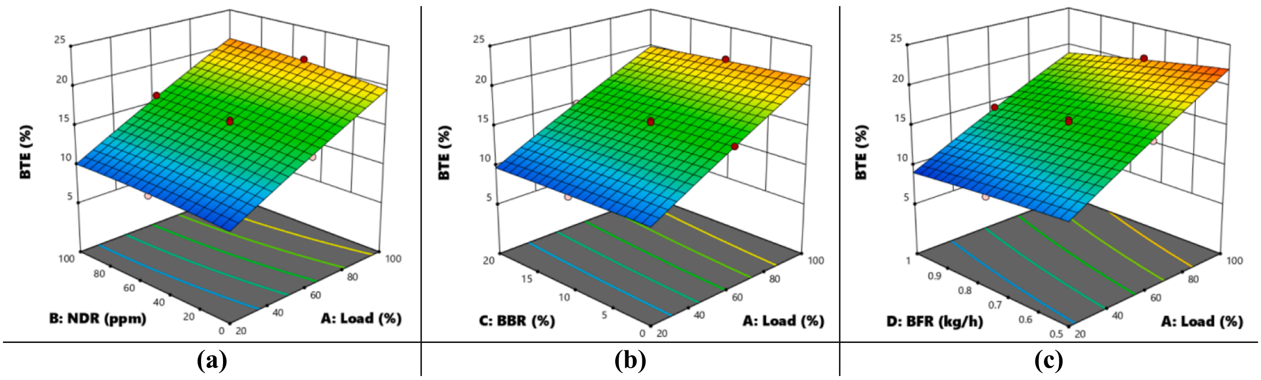


Fig. 5. Effect of (a) Load, NDR, (b) Load, BBR, and (c) Load, BFR on BTE.

conducted under comparable testing conditions with engine run with varying BFR reveals a similar increase in BSEC with an increase in BFR, reporting a 34.54% increase in BSEC for BFR 1.2 kg/h compared to BFR 0.3 kg/h[39].

3.2.3. Impact of input factors on NO_X for RSM model

The combustion chamber's high temperature and its prolonged duration are the factors that accelerate the oxidation of Nitrogen molecules resulting in NO_X formation[40]. The interaction impact of the Load vs. NDR for NO_X is depicted in Fig. 7(a). As the load on the engine rises, there is a significant rise in NO_X, highest at peak load(100%)

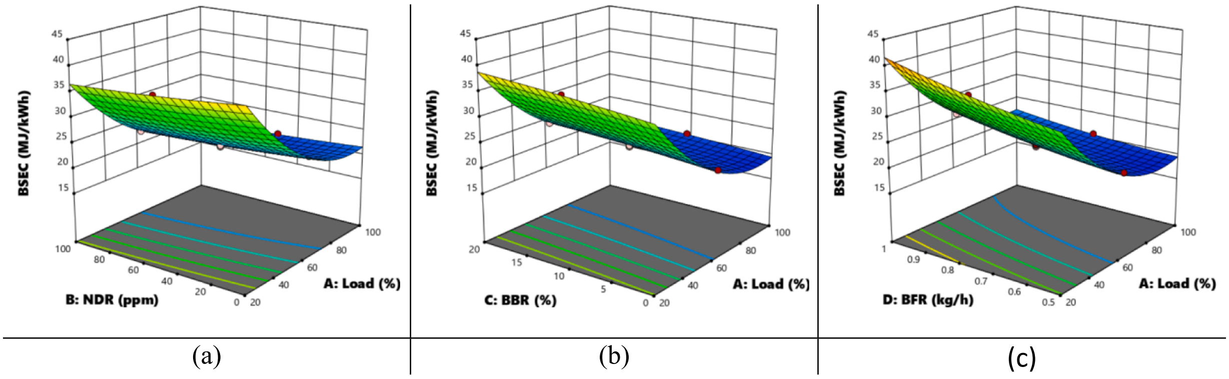


Fig. 6. Effect of (a) Load, NDR, (b) Load, BBR, and (c) Load, BFR on BSEC.

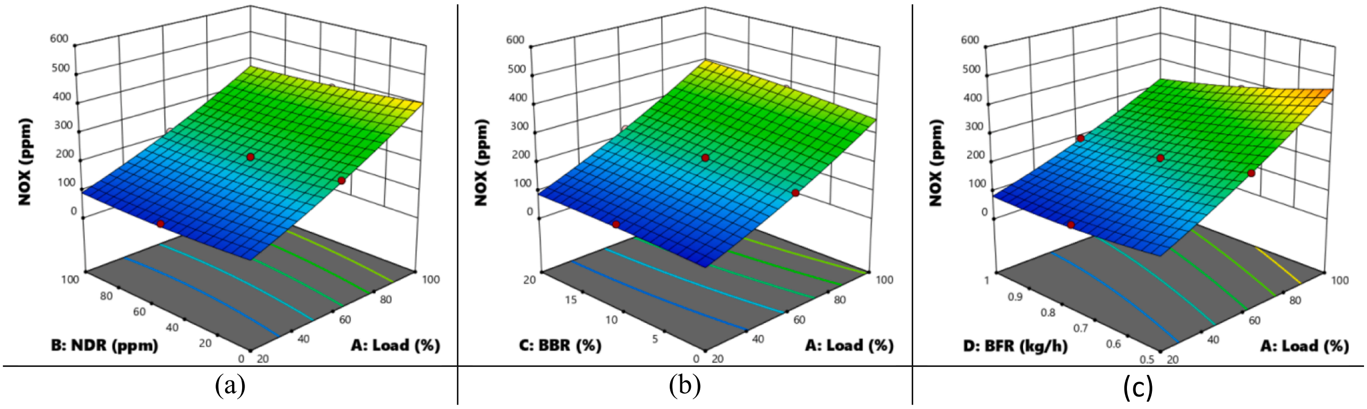


Fig. 7. Effect of (a) Load, NDR, (b) Load, BBR, and (c) Load, BFR on NO_X.

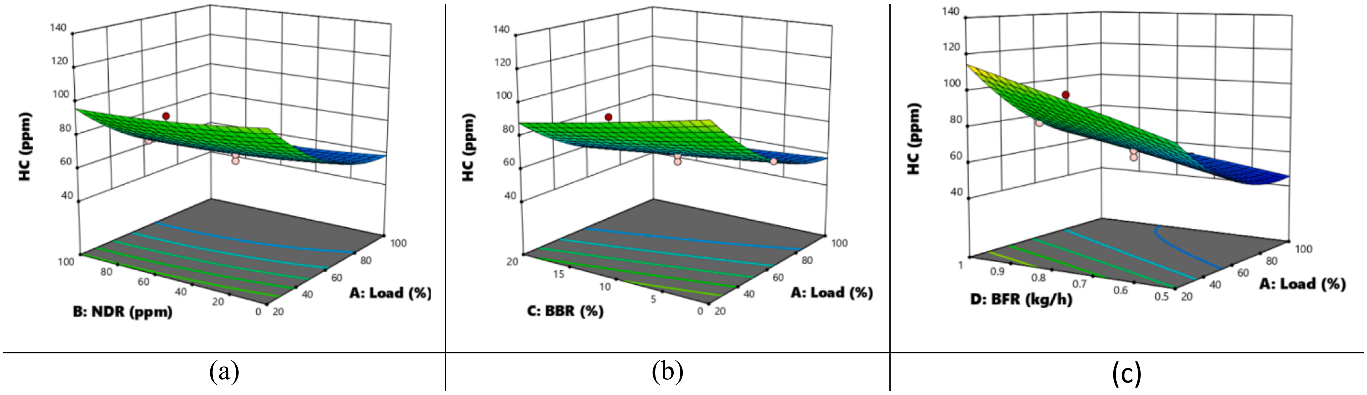


Fig. 8. Effect of (a) Load, NDR, (b) Load, BBR, and (c) Load, BFR on HC.

owing to the rise in fuel consumption and cylinder temperature [41]. The integration of nanoparticles reduces the ignition delay and optimizes the combustion progression by serving as a reducing element and converting NO_X to N_2 and O_2 [42]. Compared to an engine run on non-doped blended fuel, there was an average reduction in NO_X of 4% and 6% for NDR 50 and NDR 100, respectively, for each engine loads. Previous research conducted with biodiesel-Co_3O_4 blend under comparable testing conditions reveals a similar decrease in NO_X with an increase in NDR. The study with citronella oil biodiesel-Co_3O_4 blend(B25 + 100 ppm NDR) used in CI engine reported a decrease in NOx of 1.78% [42,43]. The interaction impact of Load vs. BBR for NO_X is depicted in Fig. 7(b). Using biodiesel in an engine increases the amount of oxygen accessible for combustion, resulting in a higher peak combustion temperature, which in turn causes an increase in NO_X as BBR increases[44]. Compared to an engine run on non-biodiesel blended fuel, there was an average increase in NO_X of 8% and 13% for BBR 10 and BBR 20, respectively, for all loads. Previous research with linseed biodiesel under comparable testing conditions reveals a similar increase in NO_X with an increase in BBR. The study reported an increase in NO_X of 1.92% and 10.70% for engine run on BBR 10 and BBR 20, respectively[45]. The interaction impact of Load vs. BFR for NO_X is depicted in Fig. 7(c). A decline in NO_X was observed with an increment in the BFR for all loads. The high specific heat of CO_2 in biogas, combined with the decreased oxygen in the fuel blend resulting from the substitution of biogas, leads to a decrease in peak temperature, which in turn minimizes the formation of NO_X[46]. Compared to an engine run on 0.5 kg/h BFR blended fuel, there was an average reduction in NO_X of 16% and 27.2% for 0.75 kg/h BFR and 1 kg/h BFR, respectively, for all loads. Previous research

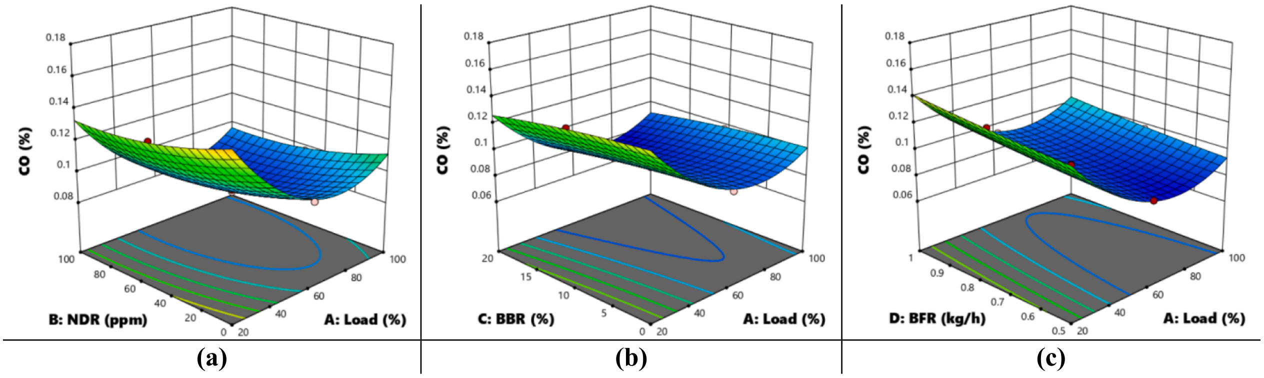


Fig. 9. Effect of (a) Load, NDR, (b) Load, BBR, and (c) Load,BFR on CO.

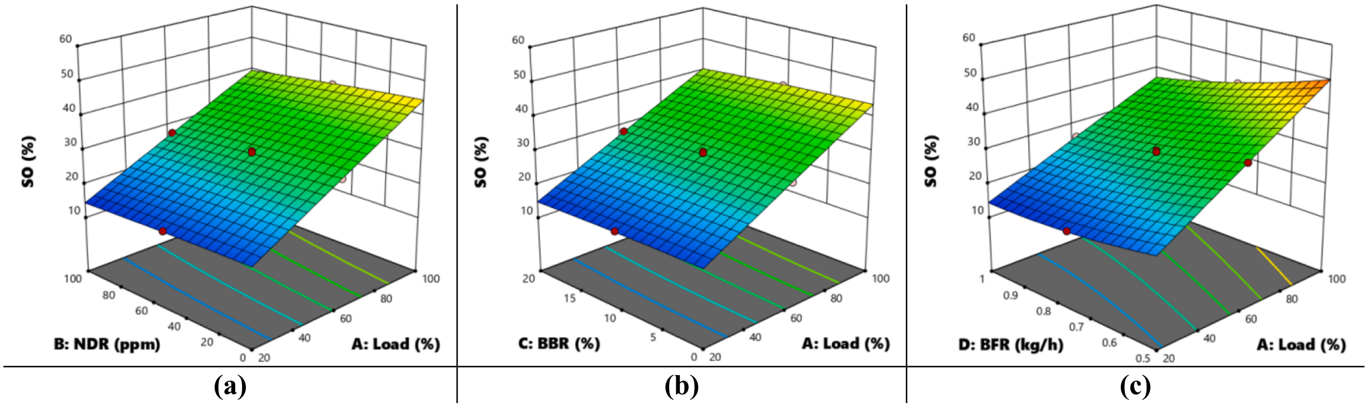


Fig. 10. Effect of (a) Load, NDR, (b) Load, BBR, and (c) Load, BFR on SO.

conducted under comparable testing conditions with varying BFR revealed a similar decline in NO_X with an increase in BFR, reporting a 38.6% decrease in BSEC for BFR 1.2 kg/h compared to BFR 0.3 kg/h [39].

3.2.4. *Impact of input factors on HC for RSM model*

Unburned fuels that are present close to the cylinder walls due to insufficient in-cylinder temperature are known as hydrocarbon (HC) emissions[47]. The interaction impact of the Load vs. NDR on HC is illustrated in Fig. 8(a). HC drops as the load increases. At low engine load due to low engine cylinder temperature, improper combustion is attained, resulting in higher HC. However, the homogenous mixing of fuel and air is more efficient at higher loads, resulting in lower HC emissions[48]. HC was found to decrease with the increase in NDR. The inclusion of Co_3O_4 reduced the quantity of HC emission owing to secondary atomization and oxidation of HC. Additionally, Co_3O_4 worked as an O_2 reservoir, increasing the oxidation rate of HC[42]. Compared to an engine run on non-doped blended fuel, there was an average reduction in HC of 4.5% and 6.4% for NDR 50 and NDR 100, respectively, for each engine load. Previous research on biodiesel-Co_3O_4 blend under comparable testing conditions revealed a similar decrease in HC with an increase in NDR. The study with citronella oil biodiesel- Co_3O_4 blend (B25 + 100 ppm NDR) used in the CI engine reported a decrease in HC of 4.91%[30,43]. The interaction impact of Load vs. BBR for NOx is depicted in Fig. 8(b). Increased oxygen availability in biodiesel leads to better combustion, resulting in lesser HC with increased BBR[16]. Compared to an engine run on non-biodiesel blended fuel, there was an average reduction in HC of 13% and 19% for BBR 10 and BBR 20, respectively, for each loads. Previous research on linseed biodiesel under comparable testing conditions revealed a similar decline in HC with an increase in BBR, reporting a decrease in HC of 7.27% for engine run on BBR 30[49]. The interaction impact of the Load vs. BFR on HC is depicted in Fig. 8(c). As BFR increases, there is a corresponding increase in HC emissions for all engine loads. Biogas has a low flame velocity, and the partial substitution of air for biogas reduces the amount of combustible oxygen, leading to increased HC emissions[6]. Compared to an engine run on 0.5 kg/h BFR blended fuel, there was an average increase in HC of 7.9% and 13.5% for 0.75 kg/h BFR and 1 kg/h BFR, respectively, for each load. Previous research conducted under comparable testing conditions with a BFR of 0.3 kg/h to 1.2 kg/h revealed a similar increase in HC with an increase in BFR, reporting a rise of 12.6% in HC for engine run on BFR 0.9 kg/h in contrast to BFR 0.3 kg/h[39].

3.2.5. *Impact of input factors on CO for RSM model*

Carbon monoxide is produced due to poor fuel-to-oxidant mixing and incomplete combustion of fuel[28]. The interaction impact of Load vs. NDR on CO is depicted in Fig. 9 (a). For engine operating at low to medium loads, the amount of carbon monoxide (CO) emitted is reduced because the air–fuel mixture is closer to the ideal stoichiometric ratio. However, as more fuel is added to the combustion process at higher loads, the amount of oxygen available for combustion is limited, resulting in a rich fuel mixture, thereby promoting an increase in CO emissions[50]. CO emissions were found to decrease as the amount of NDR in the fuel blend increased. Adding Co_3O_4 nanoparticles reduced CO emissions by shortening the ignition duration and promoting secondary atomization through micro explosions. Compared to an engine run on non-doped blended fuel, there was an average reduction in CO of 8% and 11% for each loads for NDR 50 and NDR 100, respectively. Previous research on biodiesel- Co_3O_4 blend under comparable testing conditions revealed a similar decrease in CO with an increase in NDR. The study with jatropha oil biodiesel- Co_3O_4 blend(B20 + 100 ppm NDR) used in CI engine reported a reduction in CO of 5.24%[37,43]. The interaction impact of Load vs. BBR for CO is depicted in Fig. 9(b). The stoichiometric air–fuel ratio of an engine running on biodiesel is lower than that of diesel, requiring less oxygen for combustion. Biodiesel contains more oxygen than diesel fuel, allowing carbon atoms to find enough oxygen to make CO_2, decreasing CO emissions[44]. Compared to an engine run on non-biodiesel blended fuel, there was an average reduction in CO of 5% and 10% for BBR 10 and BBR 20, respectively, for each loads. Previous research conducted with linseed biodiesel under comparable testing conditions revealed a similar decline in CO with an increase in BBR, reporting a decrease in BTE of 10% and 17.97% for engine run on BBR 10 and BBR 20, respectively[32]. The interaction impact of the Load vs. BFR on CO is illustrated in Fig. 9(c). Under each engine load, CO is realized to rise with an increment in BFR. The oxygen content of the air decreases as biogas is introduced into the cylinder, increasing CO emission as the BFR rises [46]. Compared to an engine run on 0.5 kg/h BFR blended fuel, there was an average increase in CO of 6% and 11.5% for 0.75 kg/h BFR and 1 kg/h BFR, respectively, for all loads. Previous research conducted under comparable testing conditions with a BFR of 0.3 kg/h to 0.9 kg/h revealed a similar increase in CO with an increase in BFR, reporting a rise of 14.1% in HC for engine run on BFR 0.9 kg/h contrast to BFR 0.3 kg/h[39].

3.2.6. *Impact of input factors on smoke Opacity(SO) for RSM model*

Diesel SO is an amalgamation of partially combusted fuel and soot particles[51]. The interaction impact of the Load vs. NDR on smoke opacity is depicted in Fig. 10(a). A rise in smoke opacity with increment in engine load is observed, with maximum smoke opacity to be found for the peak engine load(100%) owing to the rise in fuel consumption and cylinder temperature [41]. As NDR increased, smoke opacity emissions decreased due to the reduction in combustion temperature and the microexplosion action[50]. Compared to an engine run on non-doped blended fuel, there was an average reduction in smoke opacity of 5% and 11% for NDR 50 and NDR 100, respectively, for each loads. Previous research conducted with biodiesel- Co_3O_4 blend under comparable test conditions revealed a similar reduction in smoke opacity with rise in NDR. The study with a citronella oil biodiesel- Co_3O_4 blend(B25 + 100 ppm NDR) used in CI engine reported a decrease in smoke opacity of 10%[29,43]. The interaction impact of Load vs. BBR for CO is depicted in Fig. 10(b). Smoke opacity decreases with increase in BBR because of a better combustion process due to increased oxygen availability in biodiesel[38]. Compared to an engine run on non-biodiesel blended fuel, there was an average reduction in smoke opacity of 1.9% and 5.9% for BBR 10 and BBR 20, respectively, for each load. Previous research conducted with linseed biodiesel under comparable testing conditions revealed a similar decline in smoke opacity with an increase in BBR, reporting a decrease in smoke opacity of 6.54% for CI engine run on BBR 30[49]. The interaction impact of the Load and BFR on smoke opacity is illustrated in Fig. 10(c). Smoke opacity realized a decline for increment in BFR for each engine load. The drop in smoke opacity is attributed to the lower combustion temperatures caused by the presence of CO_2 in the biogas[52]. Compared to an engine run on 0.5 kg/h BFR blended fuel, there was an average reduction in smoke opacity of 16% and 28% for 0.75 kg/h BFR and 1 kg/h BFR, respectively, for each load. Previous research conducted under comparable test conditions with varying BFR revealed a similar decline in smoke opacity as BFR increases while reporting the highest drop of 25.88% in smoke opacity for BFR 1.2 kg/h compared to BFR 0.3 kg/h[39].

3.3. *ANN model analysis*

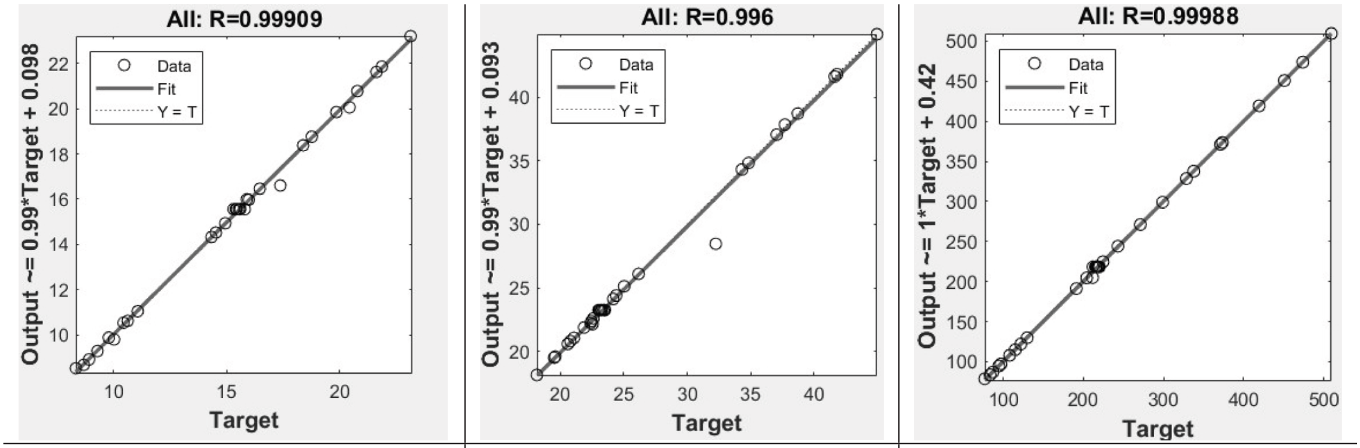
The coefficient of correlation (R) for each stage (i.e., training, testing, and validation) is obtained for the network created corresponding to output parameters. The overall coefficient of correlation (R) for every dependent response is given in Fig. 11 and Fig. 12. High R values for each network suggest good data training from DOE, and a further conclusion of the reliable regression model is reached. The predicted interaction impact graphs of NDR (Fig. S8 and Fig. S9), BBR (Fig. S10 and Fig. S11), and BFR (Fig. S12 and Fig. S13) on the output parameters reveal the congruency with interaction impact observed for the RSM model (section 3.2.1 to 3.2.6).

3.4. *Comparison of ANN model and RSM model*

For each run in the DOE Table (Table 6), the percentage error using Equation (2) is calculated for both the ANN and RSM predicted responses and presented in Fig. S14. In contrast, Table 8 shows the R^2, RMSE, and MAD evaluation metrics for prediction by the ANN and RSM models. Higher R^2 values, while lower RMSE and MAD values were analyzed mainly for the RSM prediction model compared to the ANN prediction model. Although the ANN model exhibits good prediction, lower error percentages were observed in the RSM model, indicating a better regression analysis for the input variables.

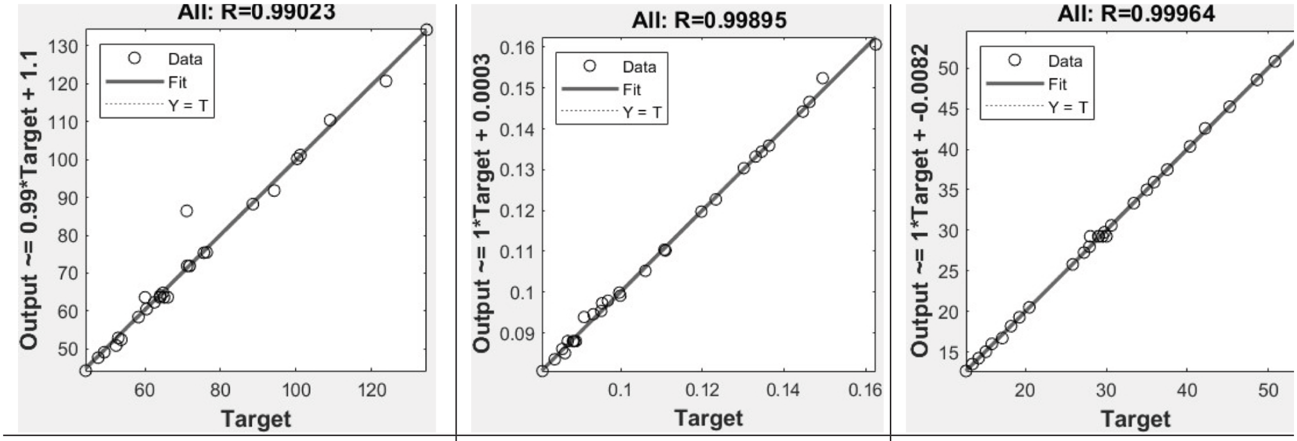
3.5. *Optimization of input parameters*

The RSM model with the optimized condition is obtained with the highest desirability of 0.742(Fig. S15), indicating the process being evaluated is operating at a level that is relatively close to the optimal conditions. RSM optimization predicted an optimal value of BTE, BSEC, NO_X, HC, CO, and SO at 17.02 %, 20.95 MJ/kWh, 256.58 ppm, 56.18



(a) BTE (b) BSEC (c) NO_X

Fig. 11. R for the trained network in response to outputs (a)BTE, (b)BSEC, and (c) $NO_{X.}$



(d) HC

(e) CO

(f) Smoke opacity

Fig. 12. R for the trained network in response to outputs (d)HC, (e)CO, and (f) SO.

Table 8
R^2, RMSE, and MAD evaluation metrics for RSM and ANN Model.

Responses	R^2		RMSE		MAD	
	ANN	RSM	ANN	RSM	ANN	RSM
BTE	0.9982	0.9992	0.184	0.116	0.092	0.087
BSEC	0.9920	0.9995	0.701	0.172	0.189	0.131
NO_X	0.9998	0.9991	1.875	3.601	0.815	2.657
HC	0.9805	0.9957	3.061	1.438	1.231	1.132
CO	0.9979	0.9991	0.001	0.001	0.001	0.001
SO	0.9992	0.9992	0.302	0.310	0.152	0.194

ppm, 0.081 %, and 30.45 % respectively at 67.53% engine load, 99.998 ppm NDR, 12.084 % BBR and 0.694 kg/h BFR.

ANN-GA optimization process terminated at 102 generations (Fig. S16), suggesting that the GA optimization process has reached a satisfactory level of convergence and the algorithm can be stopped. ANN-GA predicted optimal value of BTE, BSEC, NO_X, HC, CO, and SO at 17.16 %, 20.92 MJ/kWh, 264.75 ppm, 44.06 ppm, 0.093 %, and 33.78 % respectively at 79.11% engine load, 61.06 ppm NDR, 11.17% BBR and 0.776 kg/h BFR..

3.6. Validation of optimized results from ANN and RSM model

Three experiments were conducted at the optimum input conditions identified by RSM and ANN models. Data was collected from each repetition, and the average value for each experimental condition was calculated using a standard averaging technique. Table 9 display the optimized predicted results from RSM and ANN, experimental test validation results, and the percentage error. The percentage error is less than 5% for both model optimized results which are considered significant for acceptance. The RSM optimized result exhibits a lower error percentage than the ANN-GA results, thus confirming that RSM optimization is more accurate and reliable.

4. Conclusions

This study prepared the CCFCD matrix of 30 experimental runs for the DOE. From the DOE results, an RSM and ANN prediction model has been generated. The primary findings from the study are as follows:

1. The model generated gave an insight into the effects of input variables on output responses. F-value of engine outputs from the RSM model suggests that among the input variables, Engine load is the most substantial influence in deciding the value of output responses, followed by BFR, BBR, and NDR, respectively.
2. The evaluation metrics suggest low prediction error and high model performance for both regression analysis. Higher R^2, lower RMSE, and lower MAD were observed mainly for RSM model prediction, indicating RSM to be more accurate and reliable.
3. The optimization of the RSM model indicated an optimum response of 17.02 %, 20.95 MJ/kWh, 256.58 ppm, 56.18 ppm, 0.081 %, and 30.45 %, respectively, for BTE, BSEC, NO_X, HC, CO, and SO with operating input parameters of 67.53% engine load, 99.998 ppm NDR, 12.084 % BBR and 0.694 kg/h BFR. The optimization on the ANN-GA model indicated an optimum response of 17.16 %, 20.92 MJ/kWh, 264.75 ppm, 44.06 ppm, 0.093 %, and 33.78 %, respectively for BTE, BSEC, NO_X, HC, CO, and SO with operating input parameters of 79.11% engine load, 61.06 ppm NDR, 11.17% BBR and 0.776 kg/h BFR.
4. Although the validation test runs for the optimized result from RSM and ANN models suggest low error percentages for both, a lower error percentage was observed for the RSM optimized results.

Considering the optimized results from both the RSM and ANN model, a conclusion can be drawn that combining nanoparticles, biodiesel, and biogas is beneficial for CI engine performance and emissions. Also, RSM is found to be more accurate and reliable in studying the effects of Engine load, BBR, NDR, and BFR on CI engine.

Table 9
Validation test result and Percentage of error for the ANN & RSM optimized parameter.

Responses	Model Technique: RSM			Model Technique: ANN		
	Experimental	Predicted	Error	Experimental	Predicted	Error
BTE	17.32	17.018	1.774	18.02	17.6	2.36
BSEC	20.5	20.954	2.166	20.3	20.89	2.84
NOx	249	256.584	2.955	324.5	329.92	1.64
HC	55.15	56.18	1.833	44.25	45.18	2.07
CO	0.0792	0.081	2.22	0.0965	0.0988	2.4
SO	31.3	30.445	2.8	40.5	41.52	2.46

Funding

This study was supported by IRD, Delhi Technological University, for the project titled "Making DTU a zero organic waste campus" (Grant no. DTU/MED/HOD/2020/378).

Data availability

This research article (supplementary materials included) contains all the data evaluated during this investigation.

CRediT authorship contribution statement

S. Lalhriatpuia: Methodology, Software, Investigation, Validation, Writing – original draft. **Amit Pal:** Conceptualization, Validation, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fuel.2023.127892.

References

[1] Taqizadeh A, Jahanian O, Kani SIP. Effects of equivalence and fuel ratios on combustion characteristics of an RCCI engine fueled with methane/n-heptane blend. J Therm Anal Calorim 2020;139:2541–51. https://doi.org/10.1007/s10973-019-08669-9.

[2] Rahman M, Mohamme K, Bakar A. Air fuel ratio on engine performance and instantaneous behavior of crank angle for four cylinder direct injection hydrogen fueled engine. J Appl Sci. 2009;9:2877–86. https://doi.org/10.3923/jas.2009.2877.2886.

[3] Lalhriatpuia S, Pal A. Performance and emissions analysis of a dual fuel diesel engine with biogas as primary fuel. In: Kumar A, Pal A, Kachhwaha SS, Jain PK, editors. Recent Advances in Mechanical Engineering. Singapore: Springer Nature Singapore; 2021. p. 327–39.

[4] Mahla SK, Singla V, Sandhu SS, Dhir A. Studies on biogas-fuelled compression ignition engine under dual fuel mode. Environ Sci Pollut Res 2018;25:9722–9. https://doi.org/10.1007/s11356-018-1247-4.

[5] Makareviciene V, Sendzikiene E, Pukalskas S, Rimkus A, Vegneris R. Performance and emission characteristics of biogas used in diesel engine operation. Energy Convers Manag 2013;75:224–33. https://doi.org/10.1016/j.enconman.2013.06.012.

[6] Ambarita H. Performance and emission characteristics of a small diesel engine run in dual-fuel (diesel-biogas) mode. Case Stud Thermal Eng 2017;10:179–91. https://doi.org/10.1016/j.csite.2017.06.003.

[7] Aklouche FZ, Loubar K, Bentebbiche A, Awad S, Tazerout M. Experimental investigation of the equivalence ratio influence on combustion, performance and exhaust emissions of a dual fuel diesel engine operating on synthetic biogas fuel. Energy Convers Manag 2017;152:291–9. https://doi.org/10.1016/j.enconman.2017.09.050.

[8] Aklilu EG, Waday YA. Optimizing the process parameters to maximize biogas yield from anaerobic co-digestion of alkali-treated corn stover and poultry manure using artificial neural network and response surface methodology. Biomass Convers Biorefin 2021. https://doi.org/10.1007/s13399-021-01966-0.

[9] Mahla SK, Safieddin Ardebili SM, Mostafaei M, Dhir A, Goga G, Chauhan BS. Multi-objective optimization of performance and emissions characteristics of a variable compression ratio diesel engine running with biogas-diesel fuel using response surface techniques. Energy Sour, Part A: Recovery Utilization Environ Effects 2020;1–18. https://doi.org/10.1080/15567036.2020.1813847.

[10] Murugesan A, Umarani C, Subramanian R, Nedunchezhian N. Bio-diesel as an alternative fuel for diesel engines—a review. Renew Sustain Energy Rev 2009;13:653–62. https://doi.org/10.1016/j.rser.2007.10.007.

[11] Rai RK, Sahoo RR. Engine performance, emission, and sustainability analysis with diesel fuel-based Shorea robusta methyl ester biodiesel blends. Fuel 2021;292:120234. https://doi.org/10.1016/j.fuel.2021.120234.

[12] Suhel A, Abdul Rahim N, Abdul Rahman MR, bin Ahmad KA, Teoh YH, Zainal Abidin N. An experimental investigation on the effect of ferrous ferric oxide nano-additive and chicken fat methyl ester on performance and emission characteristics of compression ignition engine. Symmetry (Basel) 2021;13:265. https://doi.org/10.3390/sym13020265.

[13] Kumaravel ST, Murugesan A, Vijayakumar C, Thenmozhi M. Enhancing the fuel properties of tyre oil diesel blends by doping nano additives for green environments. J Clean Prod 2019;240:118128. https://doi.org/10.1016/j.jclepro.2019.118128.

[14] Hosseini SH, Taghizadeh-Alisaraei A, Ghobadian B, Abbaszadeh-Mayvan A. Artificial neural network modeling of performance, emission, and vibration of a CI engine using alumina nano-catalyst added to diesel-biodiesel blends. Renew Energy 2020;149:951–61. https://doi.org/10.1016/j.renene.2019.10.080.

[15] Safieddin Ardebili SM, Taghipoor A, Solmaz H, Mostafaei M. The effect of nano-biochar on the performance and emissions of a diesel engine fueled with fusel oil-diesel fuel. Fuel 2020;268:117356. https://doi.org/10.1016/j.fuel.2020.117356.

[16] Namdeo AK, Gupta R. Potential Of Linseed Oil Biodiesel As Fuel For CI-Engines In India. Int J Scientific Technol Res. 9, 2 (2020).

[17] AGRICULTURAL STATISTICS AT A GLANCE 2022: Government of India, Ministry of Agriculture & Farmers Welfare, Department of Agriculture, Cooperation & Farmers Welfare, Directorate of Economics and Statistics, Krishi-Bhawan, New Delhi.

[18] Latha KP, Prema C, Sundar SM. Synthesis and characterization of cobalt oxide nanoparticles. J Nanosci Technol 2018;4(475–477). https://doi.org/10.30799/jnst.144.18040504.

[19] van Gerpen J, Shanks B, Pruszko R, Clements D, Knothe G. Biodiesel Analytical Methods: August 2002–January 2004. , Golden, CO (United States) (2004).

[20] Altın R, Çetinkaya S, Yücesu HS. The potential of using vegetable oil fuels as fuel for diesel engines. Energy Convers Manag 2001;42:529–38. https://doi.org/10.1016/S0196-8904(00)00080-7.

[21] Ahmad KH, Hossain AK. Impact of nanoparticles and butanol on properties and spray characteristics of waste cooking oil biodiesel and pure rapeseed oil. E3S Web of Conferences. 23, 10001 (2017). 10.1051/e3sconf/20172310001.

[22] Pham PX, Nguyen KT, Pham Tv, Nguyen VH. Biodiesels Manufactured from Different Feedstock: From Fuel Properties to Fuel Atomization and Evaporation. ACS Omega. 5, 20842–20853 (2020). 10.1021/acsomega.0c02083.

[23] Yusof SNA, Sidik NAC, Asako Y, Japar W, Mohd AA, Mohamed SB, Muhammad NM. A comprehensive review of the influences of nanoparticles as a fuel additive in an internal combustion engine (ICE). Nanotechnol Rev 9, 1326–1349 (2020). 10.1515/ntrev-2020-0104.

[24] Bazooyar B, Ghorbani A, Shariati A. Physical properties of methyl esters made from alkali-based transesterification and conventional diesel fuel. Energy Sources Part A 2015;37:468–76. https://doi.org/10.1080/15567036.2011.586975.

[25] Budhraja N, Pal A, Mishra RS. Parameter optimization for enhanced biodiesel yield from Linum usitatissimum oil through solar energy assistance. Biomass Convers Biorefin 2022. https://doi.org/10.1007/s13399-022-03649-w.

[26] Tayyab M, Ahmad S, Akhtar MJ, Sathikh PM, Singari Ranganath M. Prediction of mechanical properties for acrylonitrile-butadiene-styrene parts manufactured by fused deposition modelling using artificial neural network and genetic algorithm. Int J Comput Integr Manuf. 1–18 (2022). 10.1080/0951192X.2022.2104462.

[27] Ramachander J, Gugulothu SK, Sastry GRK, Kumar Panda J, Surya MS. Performance and emission predictions of a CRDI engine powered with diesel fuel: a combined study of injection parameters variation and Box-Behnken response surface methodology based optimization. Fuel 2021;290:120069. https://doi.org/10.1016/j.fuel.2020.120069.

[28] Mehregan M, Moghiman M. Effects of nano-additives on pollutants emission and engine performance in a urea-SCR equipped diesel engine fueled with blended-biodiesel. Fuel 2018;222:402–6. https://doi.org/10.1016/j.fuel.2018.02.172.

[29] Krishnamoorthy R, Asaithambi K, Balasubramanian D, Murugesan P, Rajarajan A. Effect of cobalt chromite on the investigation of traditional CI engine powered with raw citronella fuel for the future sustainable renewable source. SAE Int J Adv Curr Pract Mobil 2020;3:843–50.

[30] Patil AK, Ganur SG. Comparative study on the effect of nano additives with biodiesel blend on the performance and emission characteristics of a laboratory ci engine. International journal of mechanical and production engineering research and development (ijmperd) ISSN (p). 2249–6890.

[31] Surendrababu K, Muthurajan KG, Prabhahar M, Prakash S, Saravana Kumar M, Jayakumar M. Performance, emission, and study of DI diesel engine running on pumpkin seed oil methyl ester with the effect of copper oxide nanoparticles as an additive. J Nanomater 2022;2022:3800528. https://doi.org/10.1155/2022/3800528.

[32] Nabi MN, Hoque SN. Biodiesel production from linseed oil and performance study of a diesel engine with diesel bio-diesel. J Mech Eng 2009;39:40–4. https://doi.org/10.3329/jme.v39i1.1832.

[33] Yoon SH, Lee CS. Experimental investigation on the combustion and exhaust emission characteristics of biogas–biodiesel dual-fuel combustion in a CI engine. Fuel Process Technol 2011;92:992–1000. https://doi.org/10.1016/j.fuproc.2010.12.021.

[34] Barik D, Murugan S. Investigation on combustion performance and emission characteristics of a DI (direct injection) diesel engine fueled with biogas–diesel in dual fuel mode. Energy 2014;72:760–71. https://doi.org/10.1016/j.energy.2014.05.106.

[35] Kumar ARM, Kannan M, Nataraj G. A study on performance, emission and combustion characteristics of diesel engine powered by nano-emulsion of waste orange peel oil biodiesel. Renew Energy 2020;146:1781–95. https://doi.org/10.1016/j.renene.2019.06.168.

[36] Ganesh D, Gowrishankar G. Effect of nano-fuel additive on emission reduction in a biodiesel fuelled CI engine. In: 2011 International Conference on Electrical and Control Engineering. IEEE; 2011. p. 3453–9.

[37] Sabarish R, Mohankumar D, Prem MJK, Manavalan S. Experimental study of nano additive with biodiesel and its blends for diesel engine. Int J Pure Appl Math 2018; 118:967–79.

[38] Ramalingam K, Kandasamy A, Balasubramanian D, Palani M, Subramanian T, Varuvel EG, et al. Forcasting of an ANN model for predicting behaviour of diesel engine energised by a combination of two low viscous biofuels. Environ Sci Pollut Res 2020;27:24702–22. https://doi.org/10.1007/s11356-019-06222-7.

[39] Barik D, Sivalingam M. Investigation on Performance and Exhaust Emissions Characteristics of a DI Diesel Engine Fueled with Karanja Methyl Ester and Biogas in Dual Fuel Mode. In: SAE Technical paper (2014).

[40] Agarwal AK, Gupta JG, Dhar A. Potential and challenges for large-scale application of biodiesel in automotive sector. Prog Energy Combust Sci 2017;61:113–49. https://doi.org/10.1016/j.pecs.2017.03.002.

[41] Abdel Razek S, Gad MS, Abd El Hakeem M. Experimental Investigation using CNTS as an Additive to Palm Biodiesel Blend on a DI Diesel Engine Performance, Emission and Combustion Characteristics. SJ Impact Factor:6. 887, (2017).

[42] Senthur NS, Anand C, Ramesh Kumar M, Elumalai PV, Shajahan MI, Benim AC, et al. Influence of cobalt chromium nanoparticles in homogeneous charge compression ignition engine operated with citronella oil. Energy Sci Eng 2022;10: 1251–63. https://doi.org/10.1002/ese3.1088.

[43] Ramalingam K, Perumal Venkatesan E, Aabid A, Baig M. Assessment of CI engine performance and exhaust air quality outfitted with real-time emulsion fuel injection system. Sustainability 2022;14:5313. https://doi.org/10.3390/su14095313.

[44] Şahin S, Öğüt H. Investigation of the effects of linseed oil biodiesel and diesel fuel blends on engine performance and exhaust emissions. Int J Automotive Eng Technol. 7, 149–157 (2018). 10.18245/ijaet.476775.

[45] Rashedul HK, Masjuki HH, Kalam MA, Ashraful AM, Rashed MM, Sanchita I, et al. Performance and emission characteristics of a compression ignition engine running with linseed biodiesel. RSC Adv 2014;4:64791–7. https://doi.org/10.1039/C4RA14378G.

[46] C J, Gumtapure V. Experimental investigation of methane-enriched biogas in a single cylinder diesel engine by the dual fuel mode. Energy Sources Part A 2022;44 (1):1898–911.

[47] Venkatesan SP, Kadiresh PN, Beemkumar N, Jeevahan J. Combustion, performances, and emissions characteristics of diesel engine fuelled with diesel-aqueous zinc oxide nanofluid blends. Energy Sources Part A 2019;1–15. https://doi.org/10.1080/15567036.2019.1666933.

[48] Praveena V, Martin MLJ, Geo VE. Experimental characterization of CI engine performance, combustion and emission parameters using various metal oxide nanoemulsion of grapeseed oil methyl ester. J Therm Anal Calorim 2020;139: 3441–56. https://doi.org/10.1007/s10973-019-08722-7.

[49] Govardhan Reddy K, Gupta T, Kumar Baghel P. The effect of linseed diesel blends and nano additives on diesel engine performance and emission characteristics. Int J Mech Eng 2022;7:974–5823.

[50] Hemadri V, Swamy M. Impact of cobalt oxide nanoparticles dispersed in water in diesel emulsion in reduction of diesel engine exhaust pollutants. Pollution. 2022, 579–593 (2022). 10.22059/POLL.2021.331156.1198.

[51] Rangabashiam D, Suresh Babu Rao H, Subbiah G, Vinayagam M. Study of Annona squamosa as alternative green power fuel in diesel engine. Biomass Convers Biorefin. (2021). 10.1007/s13399-021-01347-7.

[52] Verma S, Das LM, Kaushik SC. Effects of varying composition of biogas on performance and emission characteristics of compression ignition engine using exergy analysis. Energy Convers Manag 2017;138:346–59. https://doi.org/10.1016/j.enconman.2017.01.066.

Astrophysics and Space Science (2023) 368:16
https://doi.org/10.1007/s10509-023-04173-7

RESEARCH

Constraining the time-varying vacuum energy models in Brans-Dicke theory

Vinita Khatri[1] · C.P. Singh[1]

Received: 29 July 2022 / Accepted: 8 March 2023

Abstract
In this work, we constrain the time-varying vacuum energy models in Brans-Dicke theory within the framework of a flat Friedmann-Lamaître-Robertson-Walker space-time by using the latest observational data. In the first step, the analytical solution of field equations are found by considering the two functional forms of cosmological constant, viz. power-series form: $\Lambda = n_1 H + n_2 H^2$ and power-law form: $\Lambda \propto a^{-n}$, where n_1, n_2 and n are all constants, and H and a are the Hubble parameter and scale factor, respectively. Then, to test the viability of the models, the latest data sample such as Hubble $H(z)$ data, Type Ia supernovae and baryon acoustic oscillations are used to constrain the model parameters. We apply the Markov Chain Monte Carlo (MCMC) method to find the best-fit values of the space parameters of both the models. The cosmological implications of the models are discussed by using the best-fit values of parameters. It is found that both the models are in good agreement with the datasets and are consistent with the analytical solutions. We use jerk parameter and selection criteria (AIC and BIC) to find the consistency of the proposed models with the observation as compared to ΛCDM model. Both the models explain the late-time acceleration of the Universe.

Keywords Cosmology · Vacuum dark energy model · Brans-Dicke theory

1 Introduction

In the current view of modern cosmology, the nature and origin behind the current accelerating expansion of the Universe constitute a major problem. The analysis and interpretation of many observational data like Type Ia supernova (Perlmutter et al. 1999; Riess et al. 2004; Astier et al. 2006), galaxy clustering (Feldman et al. 2003), cosmic microwave background radiation (Spergel et al. 2007) and other cosmological observations (Komatsu et al. 2009, 2011; Sanchez et al. 2011; Ade et al. 2014, 2016) provide a cosmic expansion of the Universe that involves a recent accelerated expansion. This phenomena has been discussed either by adding an energy component in energy-momentum tensor usually called "dark energy" (DE) which has negative pressure, or modifying the general theory of relativity. The cosmological constant (CC), which was initially introduced by Einstein to make the static Universe, is a natural and simplest candidate of DE. This DE model, so-called standard Lambda-cold dark matter (ΛCDM) model, contains the cold dark matter for explaining cluster formation and a CC, Λ. Although the ΛCDM model fits accurately the current observational data and describes well the observed Universe, this model faces two serious problems, namely, the fine-tuning and the cosmological coincidence problems (Weinberg 1989; Copeland et al. 2006).

In the recent years, these longstanding problems have galvanized a variety of alternative theories for the cosmic acceleration beyond the ΛCDM model. One of such theories includes dynamical Λ instead of just assuming the Λ as a constant. A varying Λ has been proposed in literature to alleviate the CC problems. A number of works was proposed using varying Λ even before the discovering of the accelerating Universe (Ozer and Taha 1986; Peebles and Ratra 1988; Carvalho et al. 1992; Lima 1996; Overduin and Cooperstock 1998). This $\Lambda(t)$ model may act as an important alternative to the ΛCDM model. The $\Lambda(t)$ model is based on vacuum quantum fluctuations in the curved space-time. The resulting effective vacuum energy density depends on the

✉ C.P. Singh
cpsingh@dce.ac.in

V. Khatri
vinitakhatri_2k20phdam501@dtu.ac.in

[1] Department of Applied Mathematics, Delhi Technological University, Bawana Road, Delhi, 110 042, India

Published online: 20 March 2023

space-time curvature which decays from high initial values to smaller ones as the Universe expands (Carneiro 2003).

Due to lack of a concrete theory to model a time-varying Λ function, we generally parametrize the vacuum energy density by using phenomenological approach. In quantum field theory, the renormalization group (RG) (Shapiro and Solá 2000) describes a dynamical vacuum energy, in which the Λ-term varies as $\Lambda \sim H^2$, where H is the Hubble parameter. In a series of recent papers (Schützhold 2002; Borges and Carneiro 2005; Carneiro et al. 2006, 2008; Basilakos 2009; Basilakos et al. 2009; Perico et al. 2013; Bessada and Miranda 2013; Lima et al. 2013; Szydlowski and Stachowski 2015; Jayadevan et al. 2019), a number of flat Friedmann-Lemaître-Robertson-Walker (FLRW) type cosmologies have been studied by assuming the vacuum energy density as a truncated power-series in terms of Hubble parameter H. Carneiro et al. (2008) proposed a cosmological model by assuming the vacuum term as proportional to the Hubble parameter, $\Lambda \propto H$. Carvalho et al. (1992) and Grande et al. (2006) proposed a time-dependent vacuum, $\Lambda \sim c_1 H^2$ (here c_1 is a constant), which arises from the RG in quantum field theory. Bessada and Miranda (2013) have studied the evolution of the model with a phenomenological law $\Lambda = \Lambda_0 + 3\beta H^2$. Basilakos (2009) and Basilakos et al. (2009) investigated the properties of flat FLRW model by assuming the function form of $\Lambda(t)$ as a power series expansion in H up to the second order, $\Lambda \sim n_1 H + n_2 H^2$, where n_1 and n_2 are constants. Later on, Oliveira et al. (2014) have discussed the cosmological consequences of a model in which the vacuum varies as a truncated power series of the Hubble parameter.

On the other hand, the scalar tensor theories have been reconsidered extensively in literature, in particular, the Brans-Dicke (BD) scalar-tensor theory. Brans and Dicke (1961) introduced this scalar tensor theory to incorporate the Mach's principle in general relativity. In this theory, a scalar field ψ is included in the Einstein-Hilbert action that makes the Newtonian gravitational constant G as a function of coordinates. We replace the gravitational constant G by an inverse of time-varying scalar field ψ, which couples to gravity with a coupling parameter ω. However, in the limit $\omega \to \infty$, BD theory reduces to the corresponding general relativity. In recent years, this theory has received significant attention as it successfully describes the early inflationary era and late-time evolution of the Universe. Many authors (Pimental 1985; Johri and Kalyani 1994; Ram and Singh 1999; Sen et al. 2001; Banerjee and Pavon 2001a,b; Sen and Sen 2001; Mota and Barrow 2004; Das et al. 2006; Arik and Çalik 2006; Arik et al. 2008; Xu et al. 2010; Singh 2012; Karchi and Shojaie 2016; Kumar and Singh 2017; Singh and Kumar 2017; Srivastava and Singh 2018; Singh and Kaur 2019; Sharif and Syed Asit Ali Shah 2019; Karimkhani and Khoadam-Mohammadi 2019; Singh and Kaur 2020) have extensively studied FLRW model in the framework of BD theory. Since the vacuum energy models have the dynamical behavior, it is more suitable to consider the models in a dynamical framework such as BD theory.

In the present paper, we extend the successful approach recently presented on BD cosmology with decaying vacuum energy by Singh and Solà Peracaula (2021) with the some other suitable form of $\Lambda(t)$. However, the model could not gain the consistency with the analytical solutions. Here, we assume two different functional form of time-varying Λ: a power series up to the second order of H and a power-law form. We focus our attention on exact solutions and discuss the observational aspects of $\Lambda(t)$ models in the framework of BD theory. Additionally, we perform a Bayesian MCMC method to constrain the space parameters using the observational data of Type Ia supernova, Hubble data and baryon acoustic oscillations. Using the best-fit values of parameters we discuss the evolution of the Universe through Hubble parameter, deceleration and equation of state parameters, and check the consistency of the analytical solution sofar obtained for the both models. The model selection criteria and jerk parameters are also discussed and compared the models with the standard ΛCDM model.

The paper is organized as follows: In Sect. 2, we propose the model and basic field equations in BD theory with cosmological constant. The analytical solutions of the field equations are presented in Sects. 3 and 4 with two different functional forms of $\Lambda(t)$. In Sect. 5, we discuss the latest observational data and method to constrain the main parameters of our vacuum models. In Sect. 6, we analyze the models by using the best-fit values of model parameters. Section 7 discusses the statistical criteria of AIC and BIC in respect of the models along with ΛCDM. In Sect. 8, we draw our conclusions.

2 BD field equations with dynamical vacuum energy

We start with a spatially homogeneous and isotropic flat Friedmann-Lemaître-Robertson-Walker (FLRW) line element in standard spherical coordinates $x^i = (t, r, \theta, \phi)$

$$ds^2 = -dt^2 + a^2(t)\left[dr^2 + r^2(d\theta^2 + sin^2\theta d\phi^2)\right], \tag{1}$$

where $a(t)$ is the scale factor of the Universe. The field equations of BD theory in the presence of cosmological constant (in the unit $c = 1$) is given by (Uhera and Kim 1982; Kim 2005)

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi}{\psi} \tilde{T}_{\mu\nu} + \frac{8\pi}{\psi} T^{BD}_{\mu\nu}, \tag{2}$$

where ψ is the scalar field known as BD scalar field, ω is a dimensionless BD coupling constant and $\tilde{T}_{\mu\nu} \equiv$

$T_{\mu\nu} - g_{\mu\nu}\rho_\Lambda$ is the total energy-momentum tensor, that is, the sum of the matter and vacuum contributions. Here $\rho_\Lambda = \Lambda/8\pi G = \Lambda\psi/8\pi$ is the vacuum energy density which has the equation of state (EoS) $p_\Lambda = -\rho_\Lambda$, and $T_{\mu\nu} = (\rho_m + p_m)u_\mu u_\nu + p_m\, g_{\mu\nu}$ is the ordinary energy-momentum tensor of perfect fluid, where ρ_m is the energy density of matter, p_m is the corresponding pressure and u_μ is the four-velocity vector. Thus, we consider $\tilde{T}_{\mu\nu}$ as a perfect fluid form of energy-momentum tensor which is given by

$$\tilde{T}_{\mu\nu} = (\rho + p)\,u_\mu u_\nu + p\,g_{\mu\nu}, \tag{3}$$

where $\rho = \rho_m + \rho_\Lambda$ and $p = p_m + p_\Lambda$.

In Eq. (2), $T^{BD}_{\mu\nu}$ is considered as the energy-momentum tensor for the BD scalar which is defined by

$$T^{BD}_{\mu\nu} = \frac{1}{8\pi}\Big[\frac{\omega}{\psi}\left(\nabla_\mu\psi\nabla_\nu\psi - \frac{1}{2}g_{\mu\nu}\nabla_\alpha\psi\nabla^\alpha\psi\right) \\ +\nabla_\mu\nabla_\nu\psi - g_{\mu\nu}\nabla_\alpha\nabla^\alpha\psi\Big]. \tag{4}$$

The BD wave equation is given by

$$\Box\psi = \frac{8\pi}{(2\omega+3)}\left(T^\mu_\mu - 4\rho_\Lambda\right), \tag{5}$$

where T^μ_μ is the trace of $T_{\mu\nu}$.

For energy-momentum tensors (3) and (4), the BD field equations (2) and (5) for metric yield

$$3H^2 + 3H\frac{\dot\psi}{\psi} - \frac{\omega}{2}\frac{\dot\psi^2}{\psi^2} = \frac{8\pi}{\psi}\rho, \tag{6}$$

$$2\dot H + 3H^2 + \frac{\ddot\psi}{\psi} + 2H\frac{\dot\psi}{\psi} + \frac{\omega}{2}\frac{\dot\psi^2}{\psi^2} + = -\frac{8\pi}{\psi}p, \tag{7}$$

$$\ddot\psi + 3H\dot\psi = \frac{8\pi}{(2\omega+3)}(\rho - 3p), \tag{8}$$

where dots means time derivatives and $H = \dot a/a$ is the Hubble parameter. Let us assume the perfect fluid like form of BD energy-momentum tensor as $T^{BD}_{\mu\nu} = (\rho_{BD} + p_{BD})u_\mu u_\nu + p_{BD}g_{\mu\nu}$, where the energy density and pressure for BD are defined as

$$\rho_{BD} = \frac{1}{8\pi}\left[\frac{\omega}{2}\left(\frac{\dot\psi^2}{\psi}\right) - 3H\dot\psi\right], \tag{9}$$

$$p_{BD} = \frac{1}{8\pi}\left[\frac{\omega}{2}\left(\frac{\dot\psi^2}{\psi}\right) + 2H\dot\psi + \ddot\psi\right]. \tag{10}$$

The consistency of the BD field equations (2) yield

$$\nabla_\nu\left(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R\right) = 0 = \nabla_\nu\left(\frac{8\pi}{\psi}\tilde{T}^{\mu\nu} + \frac{8\pi}{\psi}T^{\mu\nu}_{BD}\right). \tag{11}$$

Assuming that the matter with vacuum and scalar field conserve separately, i.e., $\tilde{T}^{\mu\nu}$ obeys the usual conservation law, $\nabla_\nu\tilde{T}^{\mu\nu} = 0$, which leads to

$$\dot{\rho_m} + 3H(\rho_m + p_m) = -\dot{\rho_\Lambda}. \tag{12}$$

One should note here that the EoS of the vacuum energy density follows the same usual form, i.e., $p_\Lambda(t) = -\rho_\Lambda(t) = -\psi\Lambda(t)/8\pi$ despite it evolves with time. From (11), we obtain

$$8\pi\tilde{T}^{\mu\nu}\,\nabla_\nu\left(\frac{1}{\psi}\right) + \nabla_\nu\left(\frac{8\pi}{\psi}T^{\mu\nu}_{BD}\right) = 0, \tag{13}$$

which simplifies to

$$\dot\rho_{BD} + 3H(\rho_{BD} + p_{BD}) = (\rho_m + \rho_\Lambda + \rho_{BD})\frac{\dot\psi}{\psi}. \tag{14}$$

In this paper, we assume that the scalar field ψ is related to scale-factor a by a power-law relation (Pimental 1985; Banerjee and Pavon 2007; Sheykhi 2010):

$$\psi = \psi_0\,a(t)^\epsilon, \tag{15}$$

where ψ_0 and ϵ are constants. Using (15) into (6), we get

$$H^2 = \frac{2}{(6 + 6\epsilon - \omega\epsilon^2)}\frac{8\pi}{\psi}(\rho_m + \rho_\Lambda), \tag{16}$$

where $\Lambda = 8\pi\rho_\Lambda/\psi$. One can find that the standard cosmology is recovered in the limit of $\epsilon \to 0$.

Finally, using (12) and (16), we find

$$\dot H + \frac{(3+\epsilon)}{2}H^2 = \frac{3\Lambda}{(6 + 6\epsilon - \omega\epsilon^2)}. \tag{17}$$

In the following sections, we find the exact solutions with two forms of $\Lambda(t)$: power-series and power-law forms.

3 A power series $\Lambda(t)$ model

In this paper, we assume the time-varying Λ as a truncated power series of the Hubble parameter up to the second order [hereafter, Λ_{PS}-model], that is, (Basilakos 2009; Basilakos et al. 2009; Oliveira et al. 2014)

$$\Lambda(t) = n_1 H + n_2 H^2, \tag{18}$$

where n_1 is a constant with dimension of H, while n_2 is a dimensionless constant. The first term, i.e., $\Lambda \propto H$ was discussed in Refs. (Schützhold 2002; Borges and Carneiro 2005; Carneiro et al. 2006, 2008) where as the second term, i.e., $\Lambda \propto H^2$ was proposed in Refs. (Carvalho et al. 1992;

Grande et al. 2006). Thus, Eq. (18) is a combination of linear and quadratic form of $\Lambda(t)$. Using (18) into (17), we get

$$\dot{H}+\left(\frac{3+\epsilon}{2}-\frac{3n_2}{(6+6\epsilon-\omega\epsilon^2)}\right)H^2=\frac{3n_1 H}{(6+6\epsilon-\omega\epsilon^2)}. \tag{19}$$

Now, considering the current value of Λ as $\Lambda_0=3H_0^2\Omega_\Lambda$, where Ω_Λ is the density parameter for vacuum, Eq. (18) gives $n_1=H_0(\beta-3\Omega_m)$, where $\Omega_\Lambda=1-\Omega_m$. Here, Ω_m is the matter density parameter. Using this value of n_1, Eq. (19) can be rewritten as

$$\frac{dh}{dx}+\left(\frac{(3+\epsilon)(6+6\epsilon-\omega\epsilon^2)-6(3-\beta)}{2(6+6\epsilon-\omega\epsilon^2)}\right)h$$
$$=\frac{3(\beta-3\Omega_m)}{(6+6\epsilon-\omega\epsilon^2)} \tag{20}$$

where $x=ln\ a$ and $h=H/H_0$ is the dimensionless Hubble parameter and $\beta=3-n_2$ (or $|n_2|\ll 1$). It is obvious from (20) that $\beta>3\Omega_m$, where $0<\beta<3$.

On solving (20), the dimensionless Hubble parameter as a function of the scalar factor can be written as

$$h(a)=\frac{6(\beta-3\Omega_m)}{(3+\epsilon)(6+6\epsilon-\omega\epsilon^2)-6(3-\beta)}$$
$$+\left(1-\frac{6(\beta-3\Omega_m)}{(3+\epsilon)(6+6\epsilon-\omega\epsilon^2)-6(3-\beta)}\right)a^{-k}, \tag{21}$$

where $k=\frac{(3+\epsilon)(6+6\epsilon-\omega\epsilon^2)-6(3-\beta)}{2(6+6\epsilon-\omega\epsilon^2)}$. It is to be noted that for $\epsilon\to 0$, we recover the result obtained in Ref. (Oliveira et al. 2014) and further for $\beta\to 3$ i.e. $n_2\to 0$ and $\epsilon\to 0$, we recover the dynamical Λ solution derived in Ref. (Carneiro et al. 2006).

Considering $a=(1+z)^{-1}$, we can define the normalized Hubble expansion $E(z)$ as

$$E(z)=\frac{H}{H_0}=\tilde{\Omega}_{\Lambda 1}+\tilde{\Omega}_{m1}(1+z)^k, \tag{22}$$

where

$$\tilde{\Omega}_{\Lambda 1}=1-\tilde{\Omega}_{m1}=\frac{6(\beta-3\Omega_m)}{(3+\epsilon)(6+6\epsilon-\omega\epsilon^2)-6(3-\beta)}. \tag{23}$$

One can observe that for $\epsilon\to 0$ and $\beta\to 3$, Eq. (23) reduces to $\tilde{\Omega}_{\Lambda 1}=\Omega_\Lambda$. Assuming the scale factor to unity at present, i.e., $a_0=1$, the scale factor evolves with time as

$$a(t)=\left(\frac{e^{(k\tilde{\Omega}_{\Lambda 1}H_0\,t)}-1+\tilde{\Omega}_{\Lambda 1}}{\tilde{\Omega}_{\Lambda 1}}\right)^{1/k} \tag{24}$$

It is obvious from (24) that the scale factor varies as $a\sim t^{1/k}$, i.e., power-law expansion during the early times. Therefore, the model expands with decelerated rate. It is followed by a transition to an accelerating epoch where the scale factor varies $a\sim\exp(\tilde{\Omega}_{\Lambda 1}H_0\,t)$ in late time.

The observational data suggest that the accelerated expansion of the Universe is a recent phenomenon. It means that the Universe might be decelerated phase in the early epoch when there was no DE or when its effect was subdominant. Therefore, the Universe must have a transition from decelerating to accelerating phase. In this context, the deceleration parameter, which is defined as $q=-a\ddot{a}/\dot{a}^2$, plays an important role to describe the evolution history of the Universe. For this model, the deceleration parameter in terms of redshift is calculated as

$$q(z)=-1+\frac{k\,\tilde{\Omega}_{m1}\,(1+z)^k}{\tilde{\Omega}_{\Lambda 1}+\tilde{\Omega}_{m1}\,(1+z)^k}. \tag{25}$$

The present-day value q_0 can be found by putting $z=0$ in (25), which is given by

$$q_0=k\,\tilde{\Omega}_{m1}-1. \tag{26}$$

We now discuss the deceleration-acceleration transition redshift, z_{tr} which is defined as a redshift where $q(z)=0$ and it is given by

$$z_{tr}=-1+\left[\frac{2(6+6\epsilon-\omega\epsilon^2)\,\tilde{\Omega}_{\Lambda 1}}{\left((1+\epsilon)(6+6\epsilon-\omega\epsilon^2)-6(3-\beta)\right)\,\tilde{\Omega}_{m1}}\right]^{1/k}. \tag{27}$$

At this stage, we also discuss another important parameter, known as equation of state (EoS) parameter which describes the dynamics of the Universe. In this model, we discuss the effective EoS parameter, which is defined as $w_{eff}=-1-\frac{2a}{3}\frac{dh}{da}$. Using (21) in this expression, we get

$$w_{eff}(z)=-1+\frac{2k\tilde{\Omega}_{m1}}{3h}(1+z)^k. \tag{28}$$

The present value of EoS parameter is given by

$$w_{eff}(z=0)=-1+\frac{2k}{3}\tilde{\Omega}_{m1}. \tag{29}$$

The condition for acceleration of the present universe is given by

$$3w_{eff}(z=0)+1=2\left(-1+k\tilde{\Omega}_{m1}\right). \tag{30}$$

It should ne noted here that for an accelerated expansion of the Universe the effective EoS parameter must be $w_{eff}<(-1/3)$. This condition is satisfied if $\tilde{\Omega}_{m1}<1/k$, which is also compatible with the analysis of deceleration parameter as given in Eq. (26). From Eq. (28), it is found that $w_{eff}(z) \to -1$ as $z \to -1$, i.e., as $a \to \infty$. Thus, the model exhibits to de Sitter Universe and coincides with the ΛCDM model in later stages of its evolution.

In what follows, we check the consistency of the model. Using (9), (10) and (16) into (14), we get

$$2(\omega\epsilon-3)\dot{H}+(\omega\epsilon^2+6\omega\epsilon-12)H^2=0. \tag{31}$$

We substitute the solution of Hubble function H from (22) into (31), we get

$$\begin{aligned}&2k(\omega\epsilon-3)\tilde{\Omega}_{m1}(1+z)^k+(12-\omega\epsilon^2-6\omega\epsilon)\\&\quad\times(\tilde{\Omega}_{\Lambda 1}+\tilde{\Omega}_{m1}(1+z)^k)=0.\end{aligned} \tag{32}$$

It can be observed that, in general, the above equation is not satisfied. However, we can get a relation between the constants at present epoch, i.e., for $z=0$, which is given by

$$2k(\omega\epsilon-3)\tilde{\Omega}_{m1}+(12-\omega\epsilon^2-6\omega\epsilon)=0. \tag{33}$$

Once we get the best-fit values of model parameters by observations, we can check the consistency of the Eq. (33) for the present epoch.

4 A power-law $\Lambda(t)$ model

Bertolami (1986), Ozer and Taha (1987), and Chen and Wu (1990) have studied the model with vacuum energy density in general relativity which evolves as $\Lambda \propto a^{-2}$. In the following, we assume that the vacuum energy density evolves as the general power of the scale factor (hereafter Λ_{PL}-model):

$$\Lambda(a)=3\gamma a^{-n}, \tag{34}$$

where γ is a constant. Using (34) in (17) and simplifying, we get

$$\frac{dH^2}{da}+\frac{(3+\epsilon)}{a}H^2=\frac{18\gamma}{(6+6\epsilon-\omega\epsilon^2)}a^{-n-1}. \tag{35}$$

Using the present value of vacuum energy density, $\Lambda_0 = 3H_0^2\Omega_\Lambda$ into (34), we get $\gamma=\Omega_\Lambda H_0^2$. Using this value of γ, the solution of (35) with $H(a=1)=H_0$ in terms of redshift is given by

$$E(z)=\frac{H}{H_0}=\left[\tilde{\Omega}_{m2}(1+z)^{(3+\epsilon)}+\tilde{\Omega}_{\Lambda 2}(1+z)^n\right]^{1/2}, \tag{36}$$

where

$$\tilde{\Omega}_{\Lambda 2}=\frac{18\Omega_\Lambda}{(6+6\epsilon-\omega\epsilon^2)(3+\epsilon-n)}, \tag{37}$$

and

$$\tilde{\Omega}_{m2}=1-\frac{18\Omega_\Lambda}{(6+6\epsilon-\omega\epsilon^2)(3+\epsilon-n)}. \tag{38}$$

The deceleration parameter which is defined in the previous section is redefined as $q=-1-\frac{a}{2H^2(a)}\frac{dH^2}{da}$ and it gives

$$q=-1+\frac{1}{2}\frac{(3+\epsilon)\tilde{\Omega}_{m2}(1+z)^{(3+\epsilon)}+n\tilde{\Omega}_{\Lambda 2}(1+z)^n}{\tilde{\Omega}_{m2}(1+z)^{(3+\epsilon)}+\tilde{\Omega}_{\Lambda 2}(1+z)^n}. \tag{39}$$

It is clear from (39) that $q(z)$ tends to -1 in the future (negative redshifts). The present value q_0 is obtained as

$$q_0=-1+\frac{(3+\epsilon)\tilde{\Omega}_{m2}+n\tilde{\Omega}_{\Lambda 2}}{2}. \tag{40}$$

The transition redshift is given by

$$z_{tr}=\left(\frac{(2-n)\tilde{\Omega}_{\Lambda 2}}{(1+\epsilon)\tilde{\Omega}_{m2}}\right)^{1/(3+\epsilon-n)}-1, \tag{41}$$

where as the effective EoS parameter is calculated as

$$w_{eff}=-1+\frac{1}{3}\frac{(3+\epsilon)\tilde{\Omega}_{m2}(1+z)^{(3+\epsilon)}+n\tilde{\Omega}_{\Lambda 2}(1+z)^n}{\tilde{\Omega}_{m2}(1+z)^{(3+\epsilon)}+\tilde{\Omega}_{\Lambda 2}(1+z)^n}. \tag{42}$$

The w_{eff} at $z=0$ is obtained as

$$w_{eff}(z=0)=-1+\frac{1}{3}\left((3+\epsilon)\tilde{\Omega}_{m2}+n\tilde{\Omega}_{\Lambda 2}\right). \tag{43}$$

From above equation, we observe that the condition for acceleration of the present Universe $3w_{eff}(z=0)+1<0$ is satisfied if $(3+\epsilon)\tilde{\Omega}_{m2}+n\tilde{\Omega}_{\Lambda 2}<2$, which is also compatible with the analysis of deceleration parameter. From Eq. (43), it is found that $w \to -1$ as $z \to -1$, i.e., as $a \to \infty$. Thus, the model attains to de Sitter Universe and coincides with the ΛCDM model in late time evolution.

Let us check also the consistency of the model. Substituting the solution of Hubble function $H(z)$ obtained in Eq. (36) in Eq. (31), we get

$$\begin{aligned}&(\omega\epsilon-3)[(3+\epsilon)\tilde{\Omega}_{m2}a^{-(3+\epsilon)}+n\tilde{\Omega}_{\Lambda 2}a^{-n}]\\&\quad-(\omega\epsilon^2+6\omega\epsilon-12)[\tilde{\Omega}_{m2}a^{-(3+\epsilon)}+\tilde{\Omega}_{\Lambda 2}a^{-n}]=0.\end{aligned} \tag{44}$$

We can get the relation between the constants at present epoch which is as follows.

$$(\omega\epsilon-3)[(3+\epsilon)\tilde{\Omega}_{m2}+n\tilde{\Omega}_{\Lambda 2}]-(\omega\epsilon^2+6\omega\epsilon-12)=0. \tag{45}$$

We will check the above consistency equation for the model once the best-fit values of model parameters by observations are obtained.

5 Data sample and methodology

In this section we discuss the observational constraints on the free parameters of Λ_{PS} and Λ_{PL} models by using the latest observational data of $H(z)$, Type Ia supernovae and baryon acoustic oscillations.

5.1 Hubble data

The Hubble parameter measurements (abbreviated as $H(z)$) is an effective tools to constrain the free parameters of the model. In literature, there are two different techniques, differential-age method (Stern et al. 2010) and radial BAO method (Gaztañaga et al. 2009) to measure the Hubble parameter. We use 30 data points of Hubble parameter obtained by the so-called differential-age technique applied passively evolving galaxies in the redshift range $0.07 \leq z \leq 1.965$ as listed in Table 3 of Ref. (Solà et al. 2017). These Hubble data inputs are uncorrelated with the BAO data points.

The chi-square is defined as

$$\chi^2_{H(z)} = \sum_{i=1}^{30} \left[\frac{H_{obs}(z_i) - H_{th}(z_i, \mathbf{p})}{\sigma_{H,i}} \right]^2, \tag{46}$$

where $H_{th}(z_i)$ is the theoretical values and $H_{obs}(z_i)$ represents observed values as given in Table 3 of Ref. (Solà et al. 2017) and $\mathbf{p}$ is the set of space parameters.

5.2 Type Ia supernovae

We use the recent Type Ia supernovae (SNe) data points, the so-called Pantheon sample which includes 1048 data points of luminosity distance in the redshift range $0.01 < z < 2.3$ (Scolnic et al. 2018). This sample contains PanSTARRS1 Medium Deep Survey, SDSS, Low-z and HST samples (Scolnic et al. 2018). The chi-square function for Pantheon SNe data is

$$\chi^2_{SNe(Pan)} = \Delta\mu^T \cdot C^{-1} \cdot \Delta\mu, \tag{47}$$

where $\Delta\mu = \mu_i^{obs} - \mu^{th}(z_i, \mathbf{p})$. The observed distance modulus, μ_i^{obs} reads $\mu_{obs} = m_B - M_B$, where m_B is the observed peak magnitude in the rest frame of the B band and M_B is the absolute magnitude nuisance of SNe. The theoretical distance modulus, μ^{th}, which depends on redshift and the cosmological parameters, is defined by

$$\mu^{th}(z, \mathbf{p}) = 5\ \log_{10}[d_L(z, \mathbf{p})/10\ pc] + \mathcal{M}, \tag{48}$$

where $d_L(z)$ is the luminosity distance which is given by

$$d_L(z, \mathbf{p}) = (1+z)c \int_0^z \frac{dz'}{H(z', \mathbf{p})}, \tag{49}$$

where c is the speed of light. Also, $\mathcal{M}$ is the nuisance parameter in which H_0 and M_B can be absorbed. It is to be noted that M has been assumed to be 23.83. It is mentioned that C is the total covariance matrix which takes the form $C = D_{stat} + C_{sys}$, where the diagonal matrix D_{stat} and covariant matrix C_{sys} denote the statistical uncertainties and the systematic uncertainties, respectively.[1]

Simple analytical models of light curve predict that the SNe peak luminosity is proportional to the mass of nickel synthesized which in turn, to a good approximation, is a fixed fraction of the Chandrasekhar mass ($M_{Ni} \propto M_{Ch}$), which satisfies $M \propto G^{-3/2}$ (Khokhlov et al. 1993; Gomez-Gomar et al. 1998; Karimkhani and Khoadam-Mohammadi 2019). Based on the fact that luminosity $L \propto M_{Ch}$, a modification is required to the absolute magnitude of a SNe in the case of varying G. Thus, for the luminosity distance we have $L \propto G^{-3/2}$, i.e., for a slow decrease of G with time, the distant supernovae should be dimmer than predicted for a standard scenario. Using the definition of absolute magnitude

$$M = -2.5 \log \frac{L}{L_\odot}, \tag{50}$$

the modulus distance relation (48) must be corrected as (Li et al. 2015; Karimkhani and Khoadam-Mohammadi 2019)

$$\mu^{th}(z, \mathbf{p}) = 5\ \log_{10}[d_L(z, \mathbf{p})/10\ pc] + \frac{15}{4} \log \frac{G}{G_0} + \mathcal{M}. \tag{51}$$

Since, in BD theory, $G \propto \psi^{-1}$, where $\psi = \psi_0 a^\epsilon$, we rewrite (51) as

$$\mu^{th}(z, \mathbf{p}) = 5\ \log_{10}[d_L(z, \mathbf{p})/10\ pc] + \frac{15}{4}\epsilon \log(1+z) + \mathcal{M}. \tag{52}$$

5.3 Baryon acoustic oscillations (BAO_{dz})

In recent years, measurements of BAO have been proven as an important geometric probe that we can employ to constrain the dark energy models. In this paper, we have used BAO estimator $d_z(z)$ collected by Blake et al. (2011). It can computed as follows:

$$d_z(z_i, \mathbf{p}) = \frac{r_s(z_d)}{D_V(z_i)}, \tag{53}$$

where

$$r_s(z_d) = \int_{z_d}^{\infty} \frac{c dz}{H(z)\sqrt{3\left(1 + \frac{\delta\rho_b}{\delta\rho_\gamma}\right)}} \tag{54}$$

[1] https://archive.stsci.edu/prepds/ps1cosmo/index.html.

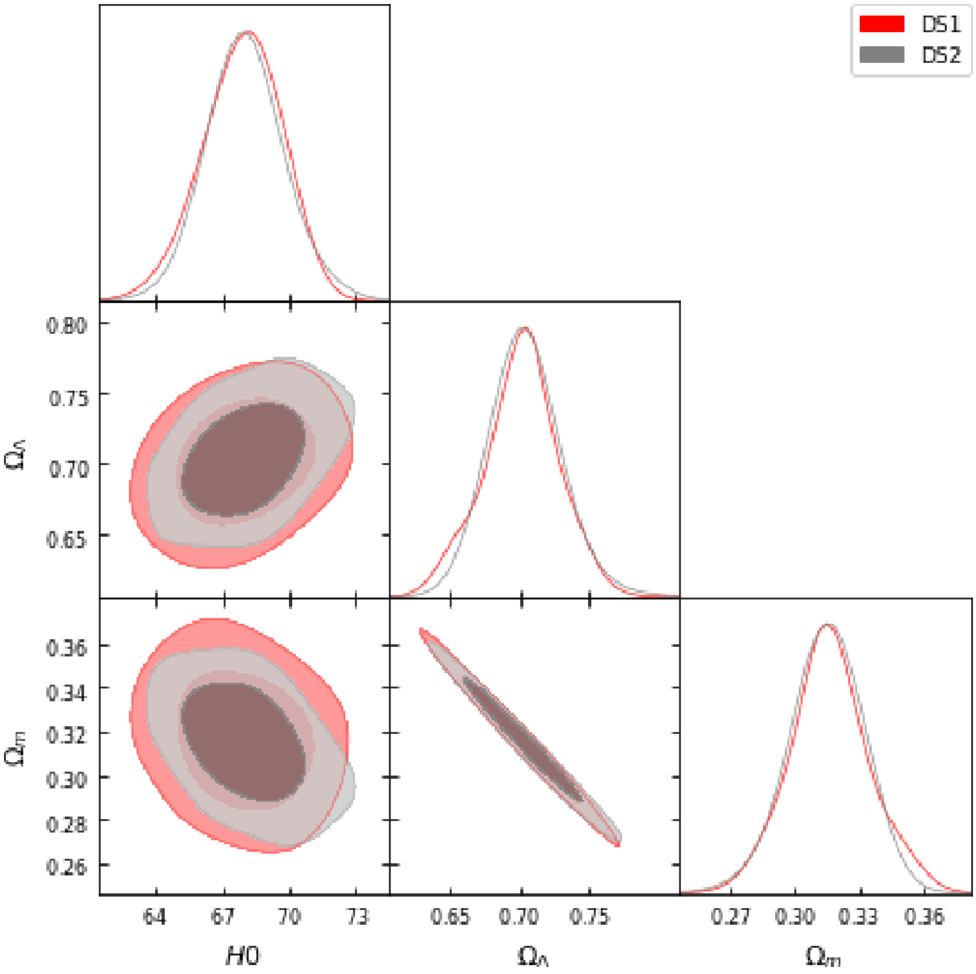


Fig. 1 Two-dimensional confidence contours and one-dimensional posterior distributions on free parameters in ΛCDM model obtained from the datasets $DS1 : SNe + H(z)$ (red contours) and $DS2 : SNe + H(z) + BAO_{dz}$ (grey contours)

is the comoving sound horizon prior to the drag redshift epoch, z_d, i.e., the epoch at which baryons are released from the Compton drag of photons, and ρ_b and ρ_γ are the baryon and photon densities, respectively.

The remaining term, $D_V(z)$ is the "dilation scale" introduced by Eisenstein et al. (2005) and can be calculated by

$$D_V(z) \equiv \left[(1+z)^2 D_A^2(z) \frac{c\,z}{H(z)}\right]^{\frac{1}{3}}. \tag{55}$$

Here, $D_A(z) = (1+z)^{-2} d_L(z, \mathbf{p})$ is the angular diameter distance.

The chi-square function for BAO_{dz} is defined as (Gomez-Valent et al. 2015a,b)

$$\chi^2_{BAO_{dz}}(\mathbf{p}) = \sum_{i=1}^{6} \left[\frac{d_{z,th}(z_i, \mathbf{p}) - d_{z,obs}(z_i)}{\sigma_{z,i}}\right]^2 \tag{56}$$

The values of z_i, $d_{z,obs}$, $\sigma_{z,i}$ can be found in Table 3 of Blake et al. (2011).

6 Results and discussion

In our analysis, we use the publicly available MCMC sampling algorithm in emcee python library (Foreman-Mackey et al. 2013) to generate the chain. In MCMC method, the best-fit of the parameters are maximized by using the probability function $\mathcal{L} \propto \exp(-\chi^2/2)$.

In order to find the best fit, we minimize the overall χ^2 function using two different combinations of datasets, namely, $DS1 : \chi^2_{min} = \chi^2_{Sne} + \chi^2_{H(z)}$ and $DS2 : \chi^2_{min} = \chi^2_{SNe} + \chi^2_{H(z)} + \chi^2_{BAO_{dz}}$. The main cosmological parameters are ϵ, ω and H_0 which are common for both models. In addition to this Λ_{PS}-model has two extra parameters Ω_m and β, and Λ_{PL}-model has two extra parameters n and Ω_Λ. We constrain the space parameters in three models: ΛCDM, Λ_{PS} and Λ_{PL}. The contours of our statistical analyses are shown in Figs. 1, 2, 3 and best-fit values of parameters are summarized in Table 1 that arise from the joint analysis described above. Using fitting values of parameters of ΛCDM, Λ_{PS} and Λ_{PL} models, a comparative study of Λ_{PS} and Λ_{PL} models with concordance ΛCDM are as follows:

Fig. 2 Two-dimensional confidence contours and one-dimensional posterior distributions on free parameters in the Λ_{PS} model obtained from the datasets $DS1 : SNe + H(z)$ (red contours) and $DS2 : SNe + H(z) + BAO_{dz}$ (grey contours)

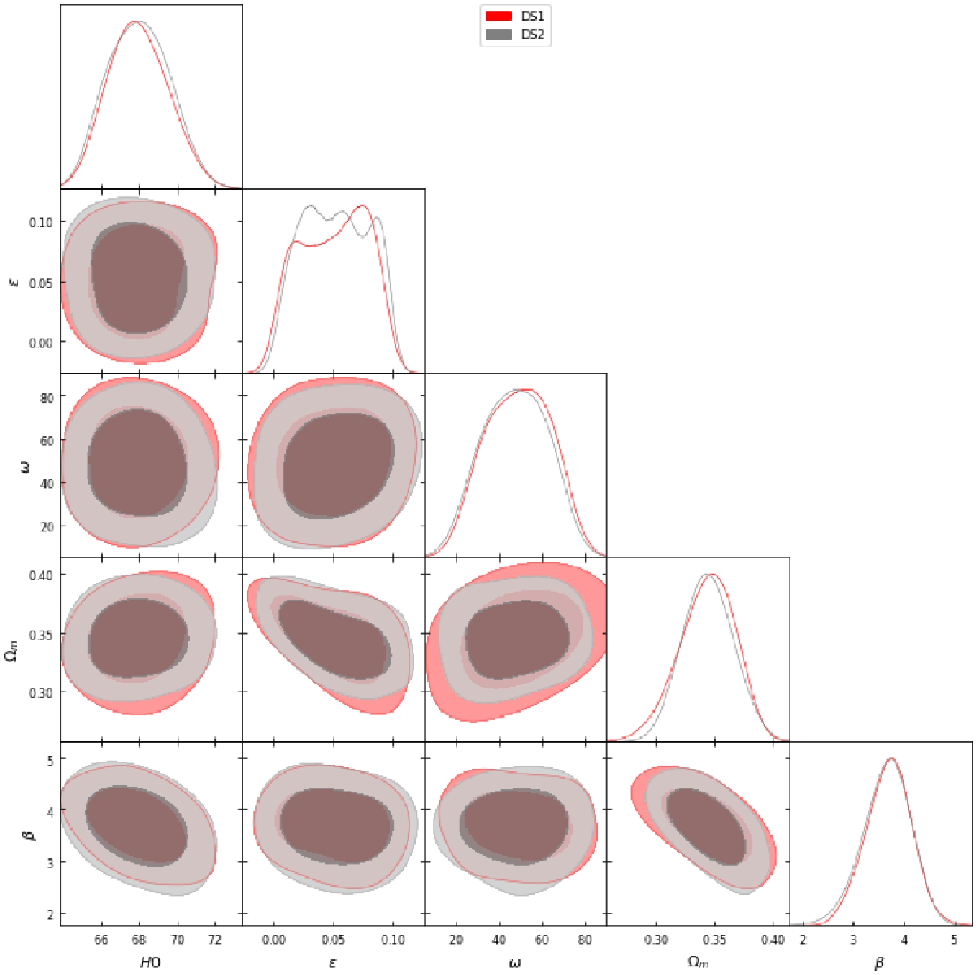


Table 1 The fit values of parameters of ΛCDM, Λ_{PS} and Λ_{PL}, respectively obtained from $DS1 : SNe + H(z)$ and $DS2 : SNe + H(z) + BAO_{dz}$ datasets. The H_0 parameter is expressed in Km s^{-1} Mpc^{-1}

$Models \rightarrow$	ΛCDM		Λ_{PS}		Λ_{PL}	
$Parameters \downarrow$	$DS1$	$DS2$	$DS1$	$DS2$	DS1	DS2
H_0	$68.179^{+2.008}_{-1.670}$	$68.126^{+1.311}_{-1.788}$	$67.801^{+1.688}_{-1.634}$	$67.706^{+1.617}_{-1.575}$	$67.266^{+1.440}_{-1.683}$	$67.182^{+1.615}_{-1.612}$
ϵ	–	–	$0.036^{+0.032}_{-0.037}$	$0.036^{+0.031}_{-0.037}$	$0.038^{+0.034}_{-0.030}$	$0.038^{+0.035}_{-0.029}$
ω	–	–	$48.234^{+18.385}_{-19.161}$	$48.384^{+18.593}_{-19.664}$	$46.489^{+20.241}_{-18.530}$	$46.151^{+19.415}_{-19.238}$
β	–	–	$3.710^{+0.414}_{-0.447}$	$3.709^{+0.419}_{-0.417}$	–	–
n	–	–	–	–	$0.219^{+0.101}_{-0.136}$	$0.222^{+0.116}_{-0.136}$
Ω_m	$0.313^{+0.018}_{-0.016}$	$0.314^{+0.015}_{-0.018}$	$0.344^{+0.023}_{-0.024}$	$0.341^{+0.026}_{-0.024}$	–	–
Ω_Λ	$0.708^{+0.023}_{-0.027}$	$0.706^{+0.027}_{-0.022}$	–	–	$0.672^{+0.021}_{-0.021}$	$0.669^{+0.021}_{-0.021}$
χ^2_{min}	569.617	10684.353	553.587	10659.909	544.063	10633.291
AIC	575.639	10690.355	563.642	10669.964	554.685	10643.346
ΔAIC	–	–	11.997	20.391	20.954	47.029
BIC	578.714	10693.458	568.750	10675.084	559.226	10684.474
ΔBIC	–	–	9.964	18.374	19.488	8.984

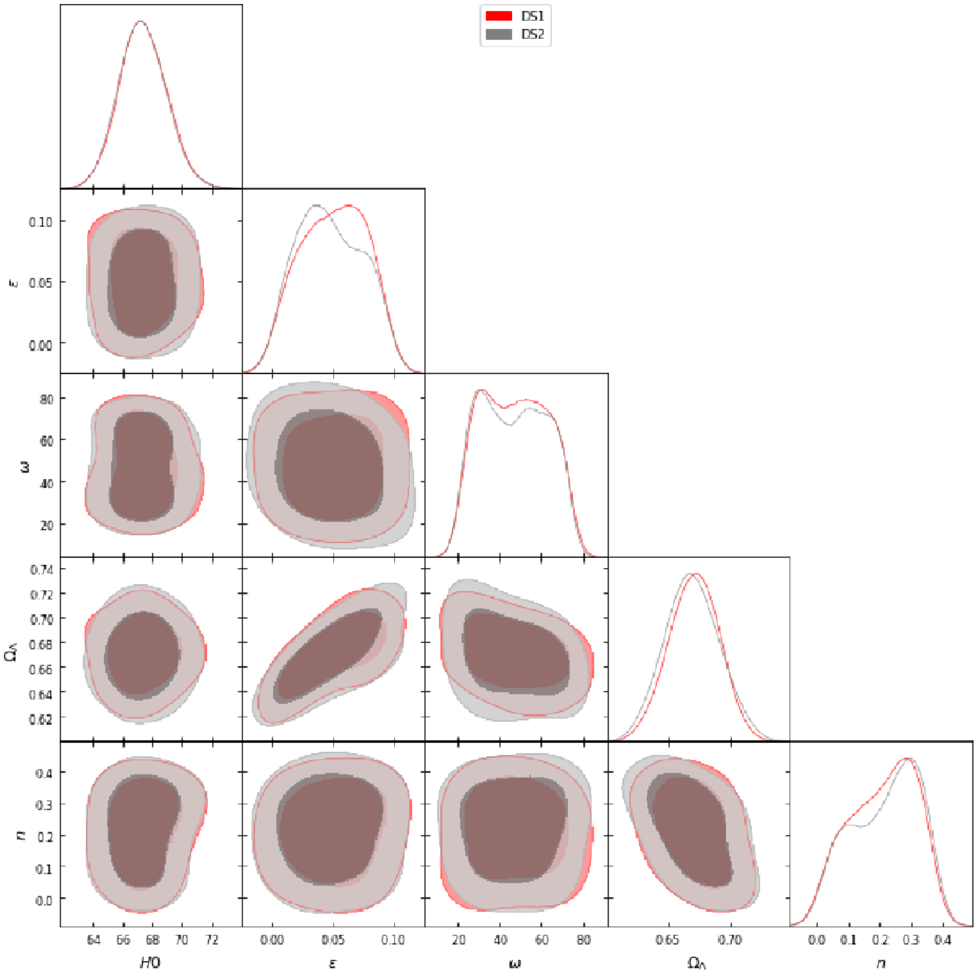


Fig. 3 Two-dimensional confidence contours and one-dimensional posterior distributions on free parameters in the Λ_{PL} model obtained from datasets $DS1: SNe + H(z)$ (red contours) and $DS2: SNe + H(z) + BAO_{dz}$ (grey contours)

Table 2 The transition value z_{tr} and the present values of q, w_{eff} of ΛCDM, Λ_{PS} and Λ_{PL}, respectively

$Model \rightarrow$	ΛCDM		Λ_{PS}		Λ_{PL}	
$Values \downarrow$	$DS1$	$DS2$	DS1	DS2	DS1	DS2
z_{tr}	$0.651^{+0.048}_{-0.048}$	$0.647^{+0.055}_{-0.055}$	$0.735^{+0.051}_{-0.180}$	$0.735^{+0.064}_{-0.180}$	$0.667^{+0.045}_{-0.052}$	$0.607^{+0.045}_{-0.052}$
q_0	$-0.541^{+0.032}_{-0.032}$	$-0.535^{+0.031}_{-0.031}$	$-0.770^{+0.23}_{-0.23}$	$-0.780^{+0.23}_{-0.23}$	$-0.459^{+0.018}_{-0.018}$	$-0.459^{+0.018}_{-0.018}$
$w_{eff}(z=0)$	$-0.694^{+0.021}_{-0.021}$	$-0.690^{+0.021}_{-0.021}$	$-0.850^{+0.15}_{-0.15}$	$-0.850^{+0.16}_{-0.16}$	$-0.639^{+0.012}_{-0.012}$	$-0.639^{+0.012}_{-0.012}$

In Λ_{PS} model, we find $\Omega_m = 0.344^{+0.023}_{-0.024}$ and $\Omega_m = 0.341^{+0.026}_{-0.024}$ from $DS1$ and $DS2$, respectively which are subsequently higher than the respective values $\Omega_m = 0.313^{0.018}_{-0.016}$ and $\Omega_m = 0.314^{+0.015}_{-0.018}$ of ΛCDM model. However, these results are close to $\Omega_m = 0.32^{+0.01}_{-0.02}$ obtained in Ref. (Basilakos et al. 2009) in general relativity.

The respective transition from deceleration to acceleration takes place at the redshift $z_{tr} = 0.735^{+0.051}_{-0.180}$ and $z_{tr} = 0.735^{+0.064}_{-0.180}$, which show that the transitions occur earlier than ΛCDM model as mentioned in Table 2, and also $z_{tr} = 0.660$ as obtained in Ref. Aghanim et al. (2020).

The present values of q and w_{eff} are listed in Table 2 which show that the q_0 and $w_{eff}(z=0)$ of Λ_{PS} are lower than the ΛCDM obtained from $DS1$ dataset. However, these values, which are obtained with dataset $DS2$, are little-bit higher than ΛCDM model. From Figs. 6 and 8, we observe that as $z \rightarrow -1$, both q and EoS parameter w_{eff} tend to -1. The present values of Hubble parameter are $H_0 = 67.801^{+1.688}_{-1.634}$ $\mathrm{Km\,s^{-1}\,Mpc^{-1}}$ and $H_0 = 67.706^{+1.617}_{-1.575}$ $\mathrm{Km\,s^{-1}\,Mpc^{-1}}$, which are good agreement with Planck result (Aghanim et al. 2020), where $H_0 = 67.7 \pm 0.46$ $\mathrm{Km\,s^{-1}\,Mpc^{-1}}$. However, these values are slightly lower than the values of ΛCDM obtained from the same datasets.

In this model we have two extra free parameters, namely ϵ and β with respect to the ΛCDM model. In dataset $DS1$ we find $\epsilon = 0.036^{+0.032}_{-0.037}$ and $\beta = 3.710^{+0.414}_{-0.447}$ whereas for $DS2$ dataset, we have $\epsilon = 0.036^{+0.031}_{-0.037}$ and $\beta = 3.709^{+0.419}_{-0.417}$.

The χ^2 is an important quantity which is used to data fitting process. In this analysis, we find $\chi^2 = 553.587$ and $\chi^2 = 10659.909$, respectively with respect to $DS1$ and $DS2$ datasets. The reduced chi-square is defined as $\chi^2_{red} = \chi^2_{min}/\nu$, where $\nu = (N - n)$ is the degree of freedom (dof). Here, N is total number of combined data, which are 1078 and 1084 and n is the number of estimated free parameters of model, which is 5 for each dataset $DS1$ and $DS2$, respectively. If $\chi^2_{red} \leq 1$, then the fit is good and the observed data is consistent with proposed model. For Λ_{PS} model, it is $\chi^2_{red} = 0.515$ and $\chi^2_{red} = 9.879$, respectively. Thus, the data $DS1$ is compatible with the considered model.

An another way to analyze the departure from the concordance ΛCDM model is through the *jerk parameter (j)*, which is a dimensionless third order derivative of the scale factor $a(t)$ with respect to cosmic time t. It is defined as (Blandford et al. 2004; Rapetti et al. 2007)

$$j = \frac{\dddot{a}(t)}{aH^3} = -q + (1+z)\frac{dq}{dz} + 2q(1+q), \tag{57}$$

where q is the deceleration parameter as given by (25) and (39). This parameter gives the information about the dynamics of DE corresponding to $j(z) = 1$ (constant) for ΛCDM model. Any deviation from $j = 1$ would favor a non-ΛCDM model. The plot of jerk parameter $j(z)$ is shown in Fig. 10 using the best-fit values of parameters obtained from DS1 and DS2 datasets in (57). It is found that $j(z) \to 1$ as $z \to -1$ which incorporates the flat ΛCDM model well in late times. The current value $j(z)$ at $z = 0$ is $j_0 = 0.6214$ and $j_0 = 0.6247$ with DS1 and DS2 datasets, respectively which differ from $j_0 = 1$ at present-day.

In Λ_{PL}, we find $\Omega_\Lambda = 0.672 \pm 0.021$ from $DS1$ and $\Omega_\Lambda = 0.669 \pm 0.021$ from $DS2$, which are comparatively lower than the value $\Omega_\Lambda = 0.708^{+0.023}_{-0.027}$ and $\Omega_\Lambda = 0.706^{+0.027}_{-0.022}$, respectively. The redshift transition values are $z_{tr} = 0.667^{+0.045}_{-0.052}$ and $z_{tr} = 0.607^{+0.045}_{-0.052}$, which are consistent with the values of ΛCDM model.

The present values of Hubble constant for this model are $H_0 = 67.266^{+1.440}_{-1.683}$ and $H_0 = 67.182^{+1.615}_{-1.612}$ obtained from DS1 and DS2 datasets which are slightly lower than the values of ΛCDM model. The evolution of $H(z)$ for this model with ΛCDM are shown in Fig. 4 and 5. The present values of q are higher than ΛCDM model whereas the $w_{eff}(z = 0)$ are very closed to standard model (refer to Table 2). From Fig. 6, 7, 8, 9, we observe that as $z \to \infty$, $q \to -0.779$ and -0.802, where as $w_{eff} \to -0.850$ and -0.864 for datasets $DS1$ and $DS2$, respectively. This model shows the quintessence-like behavior ($-1 < w \leq 0$) in late-time evolution. This model has two extra parameters, namely ϵ and n with respect to the ΛCDM model. The best-fit values of these parameters are $\epsilon = 0.038^{+0.034}_{-0.030}$ and $n = 0.219^{+0.101}_{-0.136}$ from $DS1$ dataset, and $\epsilon = 0.038^{+0.035}_{-0.029}$ and $n = 0.222^{+0.116}_{-0.136}$

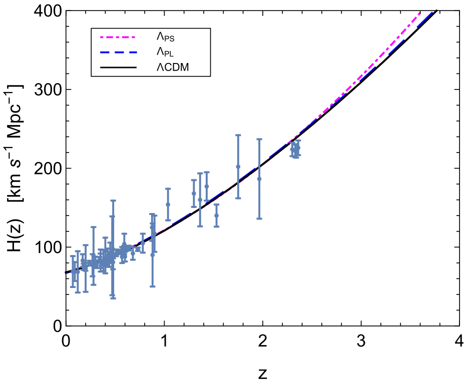


Fig. 4 Best fits over $H(z)$ obtained from $DS1$ dataset. The grey bars show the data points of $H(z)$

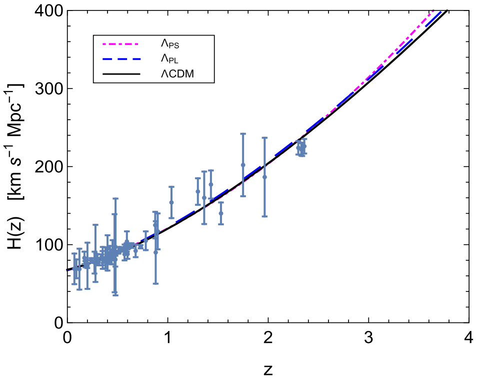


Fig. 5 Best fits over $H(z)$ obtained from $DS2$ dataset. The grey bars show the data points of $H(z)$

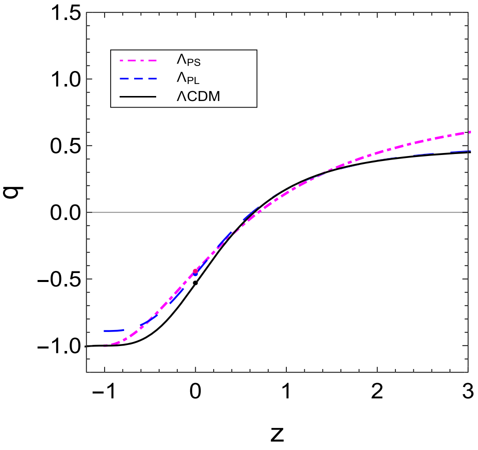


Fig. 6 Plot of evolution of deceleration parameter with redshift using fitting values of parameters obtained from $DS1$ dataset. The dot denotes the present value of deceleration parameter

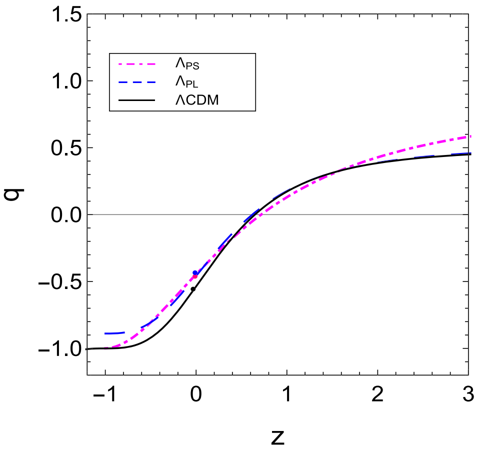


Fig. 7 Plot of evolution of deceleration parameter with redshift using fitting values of parameters obtained from $DS2$ dataset. The dot denotes the present value of deceleration parameter

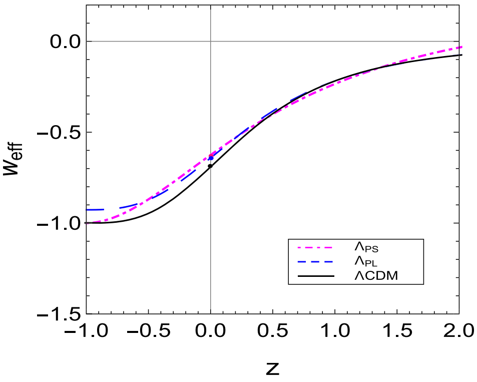


Fig. 8 Plot of evolution of EoS parameter with redshift using fitting values of parameters obtained by $DS1$ dataset. The dot denotes the present value of EoS parameter

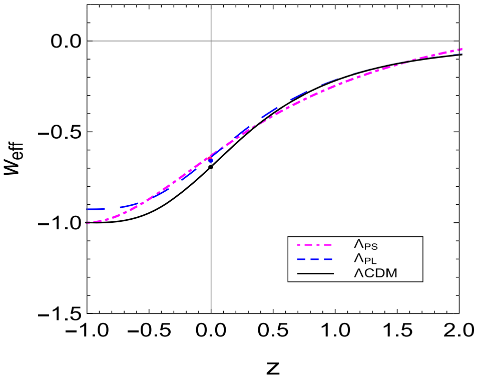


Fig. 9 Plot of evolution of EoS parameter with redshift using fitting values of parameters obtained by $DS2$ dataset. The dot denotes the present value of EoS parameter

from $DS2$. The values of n are much larger than the value $n = -0.06 \pm 0.04$ obtained in Ref. (Basilakos et al. 2009).

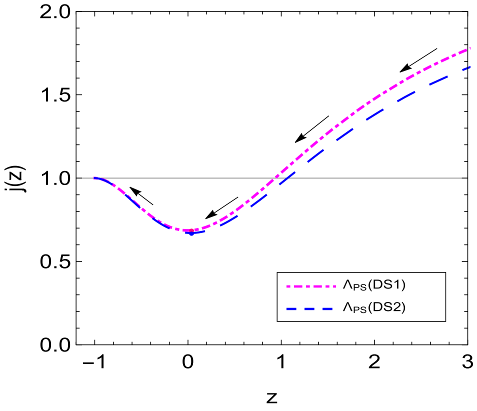


Fig. 10 Plot of evolution of jerk parameter $j(z)$ with redshift z using fitting values of parameters of Λ_{PS}. The horizontal line represents the ΛCDM model

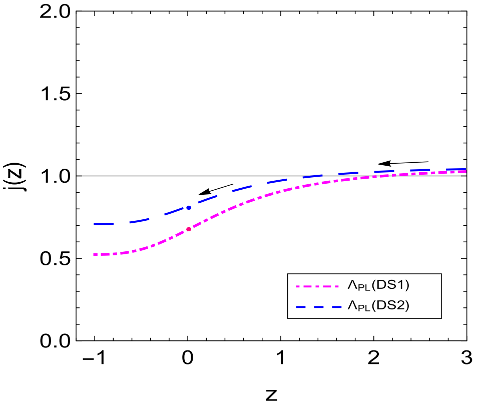


Fig. 11 Plot of evolution of jerk parameter $j(z)$ with redshift z using fitting values of parameters of Λ_{PL}. The horizontal line represents the ΛCDM model

The respective chi-square values from DS1 and DS2 datasets are $\chi^2 = 544.063$ and $\chi^2 = 10633.291$ for which $\chi^2_{red} = 0.507$ and $\chi^2_{red} = 9.854$. Thus, the χ^2_{red} is less than unity for $DS1$ dataset which show that the model provides a very good fit to this dataset.

The plot of jerk parameter $j(z)$ as a function of redshift z is shown in Fig. 11 using the best-fit values of parameters obtained from DS1 and DS2 datasets. It is observed that the Λ_{PL} deviates from the ΛCDM model at current epoch ($j_0 = 0.6004$ and $j_0 = 0.6574$, respectively) as well as $z \to -1$. These deviations from ΛCDM model need attention which would be found to know the real cause behind the cosmic acceleration.

In a paper (Singh and Solà Peracaula 2021), the authors explored two functional forms of Λ: $\Lambda = \text{const.}$ and $\Lambda = \sigma H$, so-called Λ_{H1} and Λ_{H2} models in BD theory. It was found that the BD version of the Λ-cosmology, i.e., Λ_{H1} is on an essentially equal footing position as compared to the concordance model in the light of observational fits. However, despite the quality fit of the Λ_{H2}, the model does not adapt to the consistency equation and Ω_Λ comes to very poor with this equation which is not acceptable in the present scenario. It has been shown that model Λ_{H1}, in the context of the BD theory, is more favored than Λ_{H2} and is comparable to the concordance ΛCDM model within general relativity.

In the present Λ_{PS} and Λ_{PL} models, the observational values obtained from DS1 and DS2 datasets are much more favored and satisfy the consistency equations (33) and (45), respectively, which show that the Λ_{PS} and Λ_{PL} models are also analytically consistent. The analytical values of Ω_Λ are much favored with the observed values obtained from DS1 and DS2 datasets in both the models. The version of the BD framework with these dynamical forms of Λ improve the efficiency with respect to the two datasets used. This can also be observed by information criteria as discussed below.

7 Selection criteria

In order to compare the proposed models with ΛCDM, we implement the selection information criteria in terms of the strength of the evidence according to Akaike information criteria (AIC) (Akaike 1974) and Bayesian information criteria (BIC) (Schwarz 1978). These information criteria penalize the presence of extra degree of freedom (d.o.f.). For detail discussion about these criteria, we refer to Ref. (Liddle 2007). The AIC and BIC are respectively defined as

$$AIC = \chi^2_{min} + \frac{2nN}{N-n-1}, \tag{58}$$

and

$$BIC = \chi^2_{min} + n\log(N), \tag{59}$$

where n is the number of free parameters and N is the size of the data sample. It is to be noted that the dataset $DS1$ has a total of 1078 data points (1048 data points of SNe and 30 points of $H(z)$), where as the dataset $DS2$ has 1084 data points (1048 data points of SNe, 30 points of $H(z)$ and 6 points of BAO_{dz}).

Assuming AIC (or BIC) value of ΛCDM as the reference, the AIC (or BIC) differences are defined as $\Delta AIC_i = AIC_i - AIC_{\Lambda CDM}$ (or $\Delta BIC_i = BIC_i - BIC_{\Lambda CDM}$), where i denotes either the Λ_{H1} or the Λ_{H2} model. A model having $0 \leq \Delta AIC$ (or ΔBIC) ≤ 2 gives "weak evidence in favor". In contrast, for $2 < \Delta AIC < 4$ and $2 \leq \Delta BIC < 6$, the model has "positive evidence in favor", where as for $6 \leq \Delta AIC$ (or ΔBIC) < 10, the model is considered to have "strong evidence in favor" and finally, for ΔAIC (or ΔBIC) > 10, the model has "very strong evidence in favor" (Liddle 2007). The AIC and BIC and their difference values ΔAIC and ΔBIC for models *PS-model* and *PL-model* with reference to the corresponding values of AIC and BIC of ΛCDM model are given in Table 1.

According to AIC and BIC in $DS1$ dataset, we find $\Delta AIC(\Delta BIC) = 11.997(9.964)$ for Λ_{PS} whereas it is $\Delta AIC(\Delta BIC) = 20.954(19.488)$ for Λ_{PL}. Similarly, in $DS2$ dataset, we find $\Delta AIC(\Delta BIC) = 20.391(18.374)$ for Λ_{PS} and $\Delta AIC(\Delta BIC) = 47.029(8.984)$ for Λ_{PL}. These values suggest that according to AIC, there is a *very strong evidence in favor* whereas as per BIC there is a *strong evidence in favor* of these two models.

8 Conclusion

In this work, we have discussed the dynamics of a flat FLRW model in BD theory with varying vacuum energy density. We have assumed two different functional forms of vacuum energy density, namely power series expansion in H up to the second order excluding constant term (Λ_{PS}-model) and power-law form in terms of scale factor (Λ_{PL}-model), in order to parametrize the vacuum energy density. In the first step, we have solved the BD field equations analytically using these two forms of vacuum energy density. These two models have different theoretical solutions. We have discussed the cosmological consequences of cosmic acceleration based on these two forms of interacting Λ scenarios. Secondly, we have performed two different combinations of joint likelihood analysis $DS1 = SNe + H(z)$ and $DS2 = SNe + H(z) + BAO_{dz}$ for each model including ΛCDM model in order to put the constraints on the main free parameters by χ^2 minimizing technique. It is noted that there are extra parameters, namely ϵ and β in Λ_{PS}, and ϵ and n in Λ_{PL} with respect to the ΛCDM model. The fit values of these free parameters are provided in Table 1.

Figures 1-3 show the two-dimensional confidence contours and one-dimensional posterior distributions on the free parameters in ΛCDM, Λ_{PS} and Λ_{PL} models obtained from two different datasets. The best-fit values of the model's parameters, transition redshift z_{tr}, q_0 and $w_{eff}(z=0)$ are displayed in the Tables 1 and 2, respectively. Using the fitting values we have discussed the dynamical behavior of various cosmological parameters, like $H(z)$, $q(z)$, $w_{eff}(z)$ and $j(z)$ by plotting the trajectories of evolution with redshift as shown in Figs. 4-11. In view of the observational datasets, we find datasets $DS1$ and $DS2$ are very much compatible for the considered models. The present values H_0, q_0 and $w_{eff}(z=0)$ are very close to the ΛCDM model. However, the current value of jerk parameter j_0 deviates from concordance model. We have found that the Λ_{PS} model behaves as a de Sitter model in late-time evolution of the Universe where as the Λ_{PL} model behaves as a quintessence DE with an EoS lying in $(-1 < w \leq 0)$. The χ^2_{red} implies the same goodness of the models considered here. Also, using the best-fit values of models parameters, we have found that the consistency equations (33) and (45) for the both models are satisfied. In what follows, we have summarized our main results in more detail.

Assuming ΛCDM as a reference model, we have discussed the performance of these two proposed models. We have found that both the Λ_{PS} and Λ_{PL} models show a smooth transition from deceleration ($q > 0$) epoch to acceleration ($q < 0$) epoch in recent past. The trajectories of $q(z)$ clearly show that the models generate decelerated expansion in past and late time cosmic acceleration in present. Figures 6 and 7 also show the transition from decelerated to accelerated expansion happen in the range $0.667 \le z_{tr} \le 0.735$ which are comparatively same as ΛCDM model. The parameters $q(z)$, w_{eff} and $j(z)$ tend to ΛCDM model in late-time evolution in Λ_{PS}. In Λ_{PL}, these parameters do not tend to respective values of ΛCDM in late-time evolution. It has been observed that both the model are well consistent with $H(z)$ data at low redshifts. Therefore, we conclude that both the models are well fitted with the present $H(z)$ data.

As for as the AIC and BIC statistical criteria is concerned, we have discussed these two criteria for the models against the ΛCDM to observe the performance of each model beyond the standard concordance ΛCDM model and have analyzed any deviation against or in favor of these models. According to ΔAIC and ΔBIC we have found large positive values which show that Λ_{PS} and Λ_{PL} models have *strong evidence in favor* over the ΛCDM model with reference to datasets $DS1$ and $DS2$. Finally, it should be mentioned that the results of our studied could be improved if more observational data is involved.

Acknowledgements The authors thank the anonymous reviewer for thoughtful remarks and suggestions. One of the authors, VK is thankful to Delhi Technological University, New Delhi for providing Research Fellowship to carry out this work.

Author contributions C.P. Singh has written the main manuscript text, conceptualization, investigation, methodology and observation part. Vinita Khatri has prepared the figures and observation part. Both the authors have reviewed the manuscript and agreed for submission.

Data Availability The manuscript has no associated data. The datasets used in this paper are openly available on https://doi.org/10.17909/T95Q4X.

Declarations

Competing interests The authors declare no competing interests.

References

Ade, P.A.R., et al.: Astron. Astrophys. **517**, A16 (2014)
Ade, P.A.R., et al.: Astron. Astrophys. **594**, 13 (2016)
Aghanim, N., et al. (Planck Collaboration): Astron. Astrophys. **641**, A6 (2020). arXiv:1807.06209
Akaike, H.: IEEE Trans. Autom. Control **19**, 716 (1974)
Arik, M., Çalik, M.: Mod. Phys. Lett. A **21**, 1241 (2006)
Arik, M., Çalik, M., Sheftel, M.B.: Int. J. Mod. Phys. D **17**, 225 (2008)
Astier, P., et al.: Astron. Astrophys. **447**, 31 (2006)
Banerjee, N., Pavon, D.: Phys. Rev. D **63**, 043504 (2001a)
Banerjee, N., Pavon, D.: Class. Quantum Gravity **18**, 593 (2001b)
Banerjee, N., Pavon, D.: Phys. Lett. B **647**, 447 (2007)
Basilakos, S.: Mon. Not. R. Astron. Soc. **395**, 2347 (2009)
Basilakos, S., Plionis, M., Solà, J.: Phys. Rev. D **80**, 083511 (2009)
Bertolami, O.: Nuovo Cimento B **93**, 36 (1986)
Bessada, D., Miranda, O.D.: Phys. Rev. D **88**, 083530 (2013)
Blake, C., et al.: Mon. Not. R. Astron. Soc. **418**, 1707 (2011)
Blandford, R.D., et al.: ASP Conf. Ser. **339**, 27 (2004). arXiv:astro-ph/0408279
Borges, H.A., Carneiro, S.: Gen. Relativ. Gravit. **37**, 1385 (2005)
Brans, C.H., Dicke, R.H.: Phys. Rev. **124**, 925 (1961)
Carneiro, S.: Int. J. Mod. Phys. D **12**, 1669 (2003)
Carneiro, S., Pigozzo, C., Borges, H.A.: Phys. Rev. D **74**, 023532 (2006)
Carneiro, S., Dantas, M.A., Pigozzo, C., Alcaniz, J.S.: Phys. Rev. D **77**, 083504 (2008)
Carvalho, J.C., Lima, J.A.S., Waga, I.: Phys. Rev. D **46**, 2404 (1992)
Chen, W., Wu, Y.S.: Phys. Rev. D **41**, 695 (1990)
Copeland, E.J., Sami, M., Tsujikawa, S.: Int. J. Mod. Phys. D **15**, 1753 (2006)
Das, S., Corasaniti, P.S., Khoury, J.: Phys. Rev. D **73**, 083509 (2006)
Eisenstein, D.J., et al. (SDSS Collab.): Astrophys. J. **633**, 560 (2005). arXiv:astro-ph/0501171
Feldman, H.A., et al.: Astrophys. J. **596**, L131 (2003)
Foreman-Mackey, D., Hogg, D., Lang, D., Goodman, J.: Publ. Astron. Soc. Pac. **125**, 306 (2013)
Gaztañaga, E., Cabré, A., Hui, L.: Mon. Not. R. Astron. Soc. **399**, 1663 (2009)
Gomez-Gomar, J., Isern, J., Jean, P.: Mon. Not. R. Astron. Soc. **295**, 1 (1998)
Gomez-Valent, A., et al.: J. Cosmol. Astropart. Phys. **01**, 004 (2015a). arXiv:1409.7048
Gomez-Valent, A., et al.: J. Cosmol. Astropart. Phys. **12**, 048 (2015b). arXiv:1509.03298
Grande, J., Sola, J., Stefancic, H.: J. Cosmol. Astropart. Phys. **8**, 11 (2006)
Jayadevan, A.P., et al.: Astrophys. Space Sci. **364**, 67 (2019)
Johri, V.P., Kalyani, D.: Gen. Relativ. Gravit. **26**, 1217 (1994)
Karchi, A.P.K., Shojaie, H.: Int. J. Mod. Phys. D **25**, 1650045 (2016)
Karimkhani, E., Khoadam-Mohammadi, A.: Astrophys. Space Sci. **364**, 177 (2019)
Khokhlov, A., Mueller, E., Hoeflich, P.: Astron. Astrophys. **270**, 223 (1993)
Kim, H.: Mon. Not. R. Astron. Soc. **364**, 813 (2005)
Komatsu, E., et al.: Astrophys. J. Suppl. Ser. **180**, 330 (2009)
Komatsu, E., et al.: Astrophys. J. Suppl. Ser. **192**, 11 (2011)
Kumar, P., Singh, C.P.: Astrophys. Space Sci. **362**, 52 (2017)
Li, J.-X., Wu, F.-Q., Li, Y.-C., Gong, Y., Chen, X.-L.: Res. Astron. Astrophys. **15**(12), 2151 (2015)
Liddle, A.R.: Mon. Not. R. Astron. Soc. **377**, L74 (2007)
Lima, J.A.S.: Phys. Rev. D **54**, 2571 (1996)
Lima, J.A.S., Basilakos, S., Solà, J.: Mon. Not. R. Astron. Soc. **431**, 923 (2013)
Mota, D.F., Barrow, J.D.: Mon. Not. R. Astron. Soc. **349**, 291 (2004)
Oliveira, F.A., Costa, F.E.M., Lima, J.A.S.: Class. Quantum Gravity **31**(04), 045004 (2014)
Overduin, J.M., Cooperstock, S.: Phys. Rev. D **58**, 043506 (1998)
Ozer, M., Taha, O.: Phys. Lett. B **171**, 363 (1986)
Ozer, M., Taha, O.: Nucl. Phys. B **287**, 776 (1987)
Peebles, P.L.E., Ratra, B.: Astrophys. J. **325**, L17 (1988)
Perico, E.L.D., et al.: Phys. Rev. D **88**, 063531 (2013)

Perlmutter, S., et al.: Astrophys. J. **517**, 565 (1999)
Pimental, L.O.: Astrophys. Space Sci. **112**, 175 (1985)
Ram, S., Singh, C.P.: Nuovo Cimento B **114**, 245 (1999)
Rapetti, D., Allen, S.W., Amin, M.A., Blandford, R.D.: Mon. Not. R. Astron. Soc. **375**, 1510 (2007)
Riess, A.G., et al.: Astrophys. J. **607**, 665 (2004)
Sanchez, A.G., et al.: Mon. Not. R. Astron. Soc. **425**, 415 (2011)
Schützhold, R.: Phys. Rev. Lett. **89**, 081302 (2002)
Schwarz, G.: Ann. Stat. **6**, 461 (1978)
Scolnic, D.M., et al.: Astrophys. J. **859**, 101 (2018)
Sen, S., Sen, A.A.: Phys. Rev. D **63**, 124006 (2001)
Sen, A.A., Sen, S., Sethi, S.: Phys. Rev. D **63**, 107501 (2001)
Shapiro, I.L., Solá, J.: Phys. Lett. B **475**, 236 (2000)
Sharif, M., Syed Asit Ali Shah: Mod. Phys. Lett. A **34**, 1950083 (2019)
Sheykhi, A.: Phys. Rev. D **81**, 023525 (2010)
Singh, C.P.: Astrophys. Space Sci. **338**, 411 (2012)
Singh, C.P., Kaur, S.: Phys. Rev. D **100**, 084057 (2019)
Singh, C.P., Kaur, S.: Astrophys. Space Sci. **365**, 2 (2020)
Singh, C.P., Kumar, P.: Int. J. Theor. Phys. **56**, 3297 (2017)
Singh, C.P., Solà Peracaula, J.: Eur. Phys. J. C **81**, 960 (2021)
Solà, J., Gomez-Valent, A., de Cruz Perez, J.: Astrophys. J. **836**, 43 (2017)
Spergel, D.N., et al.: Astrophys. J. Suppl. Ser. **170**, 377 (2007)
Srivastava, M., Singh, C.P.: Int. J. Geom. Methods Mod. Phys. **15**, 1850124 (2018)
Stern, D., Jimenez, R., Verde, L., Kaminokowski, M., Stanford, S.A.: J. Cosmol. Astropart. Phys. **02**, 08 (2010)
Szydlowski, M., Stachowski, A.: J. Cosmol. Astropart. Phys. **066**, 10 (2015)
Uhera, K., Kim, C.W.: Phys. Rev. D **26**, 2575 (1982)
Weinberg, S.: Rev. Mod. Phys. **61**, 1 (1989)
Xu, L., et al.: Mod. Phys. Lett. A **25**, 1441 (2010)

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

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-15118-1

Deep fusion framework for speech command recognition using acoustic and linguistic features

Sunakshi Mehra[1] · Seba Susan[1]

Received: 31 December 2021 / Revised: 6 October 2022 / Accepted: 13 March 2023

Abstract

The research problem addressed in this study is how to effectively combine multi-modal data from imperfect text transcripts and raw audio in a deep framework for automatic speech recognition. In this study, we suggest combining audio and text modalities late in the process. We propose a self-attention based deep bidirectional long short-term memory (SA-deep BiLSTM) for processing audio and text data independently. For training each type of feature, we use the SA-deep BiLSTM model which comprises of five BiLSTM layers and a self-attention module between the third and fourth layers. The linguistic data, like the word stem extracted from the text transcript, and acoustic features like Mel frequency cepstral coefficients (MFCC) and Mel-spectrogram are taken into consideration. The GloVe word embedding is used to vectorize the linguistic data. By fusing the posterior class probabilities of SA-deep BiLSTM models trained on individual modalities, we were able to achieve an accuracy of 98.80% on the 10-word categories of the Google speech command dataset. Numerous tests using the Google speech command dataset and ablation analysis prove that the suggested method performs better than the state of the art because of the high classification accuracies attained.

Keywords Acoustic features · Linguistic features · Fusion framework · Self-attention · Bidirectional long short-term memory

1 Introduction

Speech is a natural and efficient form of communication. However, the diversity of accents in a globalized society and the existence of background noise makes automatic speech identification from real-world audio samples a difficult task. To provide robust, efficient and natural interaction is one of the needs for speech-related tasks given the

✉ Sunakshi Mehra
mehra.sunakshi623@gmail.com

[1] Department of Information Technology, Delhi Technological University, Delhi, India

Published online: 23 March 2023

growing adaptive environment for voice interfaces in smart devices [42, 53]. In order to detect category labels of text and speech, natural language processing (NLP) is crucial [28]. However, due to the vast amount of multilingual speech data available online, voice search and open-ended dictation cannot cover the linguistic contents [7]. Along with natural language processing, acoustic properties taken from raw audio, including the Mel Frequency Cepstral Coefficients (MFCC), are frequently employed for speech recognition [12, 51].

Literature has extensively examined the fusion of speech elements for effective classification. One instance is the combination of two acoustic features: MFCC and Gammatone Frequency Cepstral Coefficients (GFCC) to classify speech signals in [23]. Effective features for the problem of Alzheimer's dementia (AD) recognition include MFCC and Log-mel-spectrograms [34]. For the purpose of classifying ambient sounds, many acoustic parameters including Log-mel Spectrogram, Chroma, MFCC, Tonnetz and Spectral Contrast have been combined by Su et al. in [47]. Combination of machine learning models is also investigated such as the fusion of neural network language models with recurrent neural network transducers in [22]. The encoders of the acoustic pre-trained model wav2vec 2.0 and the linguistic pre-trained model BERT were fused in [58] to classify low-resourced speech data.

In this paper, we explore multi-modal fusion in a deep framework that is trained from scratch on a speech command dataset. We thus combine the goodness of multi-modal representations and deep learning in a unified framework. The study investigates acoustic and linguistic learning for recognizing the speech commands. In particular, we analyze the auditory and linguistic information using tailored self-attention based deep recurrent neural network model and combine the probabilistic predictions for speech command identification. The following are the contributions made by our work:

1. We propose a self-attention based deep bidirectional long short-term memory (SA-deep BiLSTM) architecture for learning the acoustic and linguistic features independently from raw audio and text transcript, respectively.
2. The linguistic feature (morpheme) is the stem of each word in the text transcript, that is vectorized using GloVe embedding, and the acoustic features are MFCC and Mel-spectrogram.
3. Late fusion on the probabilistic predictions of the individual SA-deep BiLSTM models is performed for identifying the speech command.
4. To demonstrate the efficacy of the fusion framework in comparison to the individual models trained on smaller subsets of audio and text modalities, an ablation analysis is presented.
5. An extensive comparative analysis with the state of the art is performed to prove the efficacy of our method for speech command recognition.

The article's remaining sections are organized as follows. A brief synopsis of the research done so far on spoken word recognition is provided in Section 2. Section 3 presents the proposed deep fusion framework for multi-modal fusion. In Section 4, the experimental strategy is described, and the results are thoroughly analyzed. Section 5 wraps up the work and suggests possible future research trajectories.

2 Related work

The field of speech and language processing has engaged the interest of researchers since many decades, and a lot of noteworthy work has been achieved in this area, some of which are discussed in this section. Today, the usage of multimedia systems that teach users how to speak and enunciate sounds is gaining in popularity [4, 31]. Real-time applications for speech-based communication exist, and voice interface modules are built into smart appliances like the Alexa Echo, TV, and refrigerators. The usage of acoustic features such as MFCC and its variants have been well explored [48, 51] and, more recently, deep networks such as SincNet fused with X-Vectors, wav2vec 2.0 pre-trained model, and audio ALBERT have been proposed [10, 44, 52] that extract deep features from raw audio. As an alternative, text transcriptions that are produced from the raw audio with the use of API [35] may serve as the foundation for automatic speech recognition (ASR) models. Spectrogram [23] is another speech feature in addition to MFCC. The two-dimensional map known as a spectrogram shows the amplitude of sound on a frequency-time graph. A different field of study known as “spectrographic speech processing” analyzes the two-dimensional spectrogram to derive visual acoustic information [45]. Recently, deep neural networks trained on spectrograms were utilized to classify music [26]. Acoustic scene classification algorithms were found to be more effective when spectrograms were added, whether in the form of sub-spectrograms, ensembles, or combinations with auditory data [20, 40, 43]. According to the observations of Gallardo-Antolín et al. in [16], the speech intelligibility level can be adjusted by mixing acoustic and modulated spectrograms that are sent to an attention LSTM-based speech system. By introducing an attention mechanism, the final sentence embedding can use the attention summation to directly access earlier LSTM hidden states [27]. Despite the fact that neural networks have been successfully applied for ASR applications, they have some drawbacks such as overfitting, difficulties processing huge datasets, insufficient training with limited resources, and computational cost [5].

It is thus important to examine spoken word recognition from both a functional and a temporal standpoint [32]. Online spoken word recognition in templatic languages has only been the subject of a small number of investigations. The conventional gating paradigm was used to analyze the lexical (neighborhood density and frequency) and morphological (role of root morpheme) components of spoken word recognition in a templatic language [38]. The mapping is tested on a spoken word classification task that is akin to ASR, although it performs with a very low level of accuracy [2]. In unsupervised speech translation or query-by-example search, emphasizing semantic elements in embeddings can be helpful, but they are insufficient to reliably classify spoken words for robotic speech transcription [2]. On similar lines, certain efforts were made to detect spoken terms in Persian in broadcast news in [54].

Researchers claim that combining audio and text data for ASR improves recognition accuracy. Human perception of the world involves the integration of multimodal inputs [37]. Macary et al. in a recent work [30] demonstrated how to combine wav2vec 2.0 and CamemBERT pre-trained models to extract acoustic and contextual information for speech emotion recognition. On similar lines, acoustic and linguistic encoders of pre-trained models were fused in [58] for low-resourced speech recognition. All of these studies were motivated by the limitations of models that are trained on a single modality. To encode dynamic contextual information, the integration of shallow fusion, neural network language model, and trie-based deep biasing was investigated in [24].

A recent idea to fuse linguistic and acoustic information in a single representation was presented by Zheng et al. by combining acoustic and text-masked language models [61]. In the decoupled transformer model introduced by Zhang et al. in 2021 [60], audio-to-phoneme networks learn acoustic patterns while the phoneme-to-text network performs the classification task. A Recurrent Neural Network Transducer (RNN-T) is trained using pairs of YouTube audio and text transcripts in [33]. RNN-T has an implicit neural network language model (NNLM), which makes it challenging to use unpaired text inputs during training [22]. This is true of the majority end-to-end voice recognition model architectures. The RNN-T is used with source and target domain language models in the study proposed by [6] to assess ASR. The transfer learning method in [6] exhibited improved outcomes in speech emotion identification when acoustic and linguistic knowledge were combined. The authors of [36] proposed a transfer learning method based on cellular learning automata (CLA) to reduce negative transfers.

3 Proposed deep fusion framework for acoustic and linguistic features

In this work, multimodal fusion of audio and text modalities is investigated for speech command recognition. A self-attention based deep bidirectional long short-term memory (SA-deep BiLSTM) is proposed for classifying the speech commands by using acoustic features such as MFCC, Mel-spectrogram, and linguistic feature- morpheme of each word in the text transcript, i.e., stem. The process flow is shown in Fig. 1 which depicts the sequence in which the experiment was conducted.

The performance of the speech recognition system is anticipated to be improved by the merging of acoustic and linguistic decision streams in a deep framework. This is the primary assumption in our work. A late fusion strategy is employed to probabilistically fuse the output predictions. The complete algorithm of our proposed approach is shown below.

Algorithm Acoustic and linguistic feature extraction, classification with SA-deep BiLSTM, and Fusion

```
Input: Training Set $X_{train}$, Test Set $X_{test}$, Training Labels $Y_{train}$, Test Labels $Y_{test}$
Output: ACCURACY
    1.  For each audio file in [$X_{train}$, $X_{test}$] do
                ACOUSTIC_FEATURE_MFCC ← Extract MFCC features
                ACOUSTIC_FEATURE_MS ← Extract Mel-spectrogram features
                TEXT_TRANSCRIPT ← Extract Text-Transcript (using Google API)
                STEM ← TEXT_TRANSCRIPT to STEM (using Porter Stemmer)
                LINGUISTIC_FEATURE ←Vectorize STEM (using GloVe Embedding)
        End for
    2.  Instantiate three separate SA-deep BiLSTM models for training on samples of
        ACOUSTIC_FEATURE_MFCC, ACOUSTIC_FEATURE_MS and
        LINGUISTIC_FEATURE, belonging to $X_{train}$ having labels $Y_{train}$
    3.  Use SA-deep BiLSTM models to predict class label of $X_{test}$ for the three features
    4.  Create lists for storing posterior class probabilities computed by the three models
    5.  CLASS_PROBABILITY ← Fusion of posterior probabilities for each class
    6.  $Y_{pred}$ ← argmax (CLASS_PROBABILITY)
    7.  ACCURACY ← Calculate accuracy using ($Y_{pred}$, $Y_{test}$)
```

We first describe the proposed SA-deep BiLSTM model architecture that is used to independently learn the acoustic and linguistic modalities. One of the biggest problems of

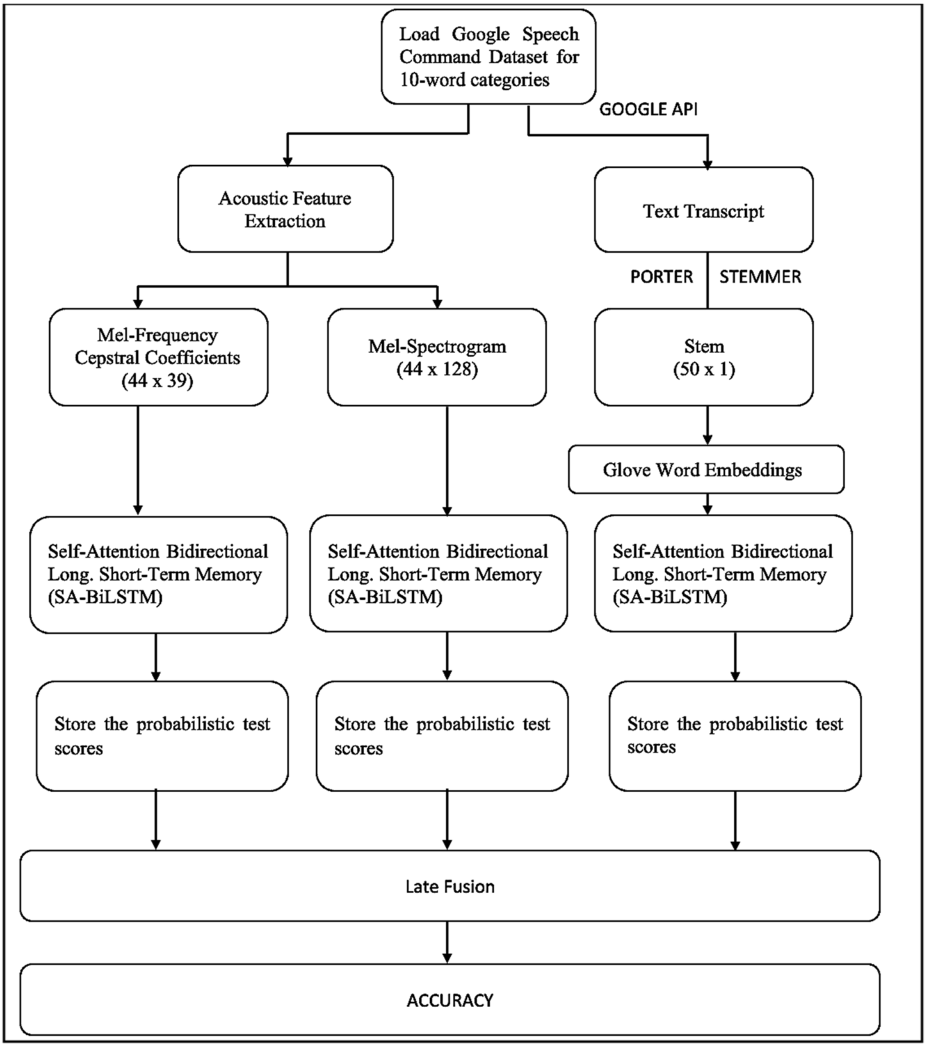

Load Google Speech Command Dataset for 10-word categories
GOOGLE API
Acoustic Feature Extraction
Text Transcript
PORTER STEMMER
Mel-Frequency Cepstral Coefficients (44 x 39)
Mel-Spectrogram (44 x 128)
Stem (50 x 1)
Glove Word Embeddings
Self-Attention Bidirectional Long. Short-Term Memory (SA-BiLSTM)
Self-Attention Bidirectional Long. Short-Term Memory (SA-BiLSTM)
Self-Attention Bidirectional Long. Short-Term Memory (SA-BiLSTM)
Store the probabilistic test scores
Store the probabilistic test scores
Store the probabilistic test scores
Late Fusion
ACCURACY

Fig. 1 Process flow for learning the acoustic and linguistic features using SA-deep BiLSTM model

the basic RNNs is that they lose vital information while working with long sequences. To overcome these difficulties and recall specifics from extended data sequences for understanding the correlations between dispersed data, long short-term memories (LSTMs) [19] were developed. These models include a cell state (i.e., the network's memory), a hidden state, and three gates that permit the gradient to continue to flow (used to make predictions). The trio is composed of an output gate, an input gate, and a forget gate. The forget gate determines which information will be removed from the cell state. When the data x_t and and the hidden state h_{t-1} are given as inputs, this gate, which is effectively a sigmoid function, outputs a number between 0 and 1 for each component of the cell state. A 0 indicates discarding of information, whereas 1 indicates retaining the information. Initially, the data x_t and hidden state h_{t-1} are passed through a sigmoid layer in the output gate. The new hidden state h_t is then produced by multiplying the results with c_t after it has been passed through a *tanh* layer. The term "bidirectional LSTM" describes the technique of first computing the hidden states from front to rear, and then vice versa, and then combining the two results. LSTM accepts inputs in the form of *samples* × *features* × *time-steps*. Each time-step input is recursively linked to the previous memory. It is popularly used to classify sequential data, including audio and text [9, 16].

Let x_t be the LSTM input which stands for the audio or text vectors originating from sequential data. h_t is the hidden state of the present timestamp and h_{t-1} is the hidden state of the previous timestamp. The input and previous hidden state combine to create the vector X as depicted in (1), which serves as the input for the forget, input and output gates (Eq. (2–4)). Here, W_i, W_f, $W_o \in R^{d \times h}$ are weight matrices, b_i, b_f, $b_o \in R^{h}$ are the biases of the input (i), forget (f) and output (o) gates, respectively, that are determined during the training phase. σ is the sigmoid function.

$$X = [h_{t-1}, x_t] \tag{1}$$

$$f_t = \sigma\left(W_f X + b_f\right) \tag{2}$$

$$i_t = \sigma\left(W_i X + b_i\right) \tag{3}$$

$$o_t = \sigma\left(W_o X + b_o\right) \tag{4}$$

In Eq. (2)–(4), f, i and o represent gate activations. The *tanh* is the hyperbolic tangent function, and * represents the element-wise multiplication. c_{t-1}, c_t in Eq. (5) represent the previous and current cell states, respectively.

$$ct = f_t * c_{t-1} + i_t * tanh\left(W_c * X + b_c\right) \tag{5}$$

The hidden state at time-step t is computed as.

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

The major difference between LSTM, BiLSTM and Deep BiLSTM is that LSTM is unidirectional and preserves information only from the past, while BiLSTM runs the inputs from past to future and also from future to past [11], whereas deep BiLSTM has a stack of multiple recurrent layers that enhances the efficiency of the BiLSTM model [21]. In BiLSTM, there are two hidden layers, one for the forward pass and the other for the backward pass. The final hidden state is the concatenation of the hidden states computed in the forward and backward passes is shown in Eq. (7).

$$h_t = \left[\overrightarrow{ht}; \overleftarrow{ht}\right] \tag{7}$$

Attention mechanisms allows the model to gather information about the context by looking at the neighbors of the target word [3]. Another type of attention mechanism is self-attention which is also termed intra-attention [9] which is different from the inter-attention proposed by Bahdanau et al. (2014). Attention mechanisms let a model directly look at and draw the state of earlier vectors. The clear utility of the attention mechanism was proved in neural machine translation (NMT) [3]. The computations for self-attention are given in Eq. (8)–(10).

$$l_t = \sum_{t'} \alpha_{t,t'} X_{t'} \tag{8}$$

$$\alpha_{t,t'} = softmax\left(\sigma\left(W_a\ h_{t,t'} + b_a\right)\right) \tag{9}$$

$$h_{t,t'} = tanh\ \left(x_t^T W_t + x_{t'}^T W_x + b_t\right) \tag{10}$$

We hypothesize that the deep BiLSTM model with self-attention would perform even better on the fusion of audio and linguistic-based information than when learning from the different modalities independently. We obtain the predictions of the individual deep models trained on acoustic and linguistic features, and fuse the predictions using a soft fusion technique. Soft fusion of prediction probabilities is a well-known technique in machine learning used for combining the predictions of individual models in an ensemble [49]. The fusion strategy opted in our work is late fusion in which we fuse the probabilistic predictions by taking the average or maximum probabilistic score associated with each class. The soft fusion procedure in our deep fusion framework is detailed below.

Let the posterior class probability associated with the spoken word category c be denoted by p_c. We fuse the three probabilistic decision scores of the SA-Deep BiLSTM models trained on MFCC, Mel-spectrogram (ms), and stem, using both maximum and average functions as shown below.

$$p_c = max\left({p_c}^{(mfcc)}, {p_c}^{(ms)}, {p_c}^{(stem)}\right) \tag{11}$$

$$p_c = mean\left({p_c}^{(mfcc)}, {p_c}^{(ms)}, {p_c}^{(stem)}\right) \tag{12}$$

The class of the test sample is calculated.

$$class = \forall c\ argmax\ (p_c) \tag{13}$$

To sustain the spectral information of the audio, the MFCC [12] and Mel-Spectrogram [46] acoustic features are extracted from the .WAV raw audio files and stored in the form of matrices of dimensions *samples* ×*features*. To sustain linguistic properties of the audio file, the stemming algorithm: Porter stemmer [41] is applied after acquiring the text transcript using Google API. The stemmed transcript is further converted to a feature-matrix using GloVe word embedding [39]. In Fig. 2, the architecture of our proposed self-attention based deep bidirectional long short-term memory (SA-deep BiLSTM) is shown. In Fig. 2, the “MATRIX” represents the feature vectors extracted from each of the modalities: audio and text. The frame length is fixed to 44 for all audio files. MFCC matrix dimension is (44 × 39), the shape of Mel-

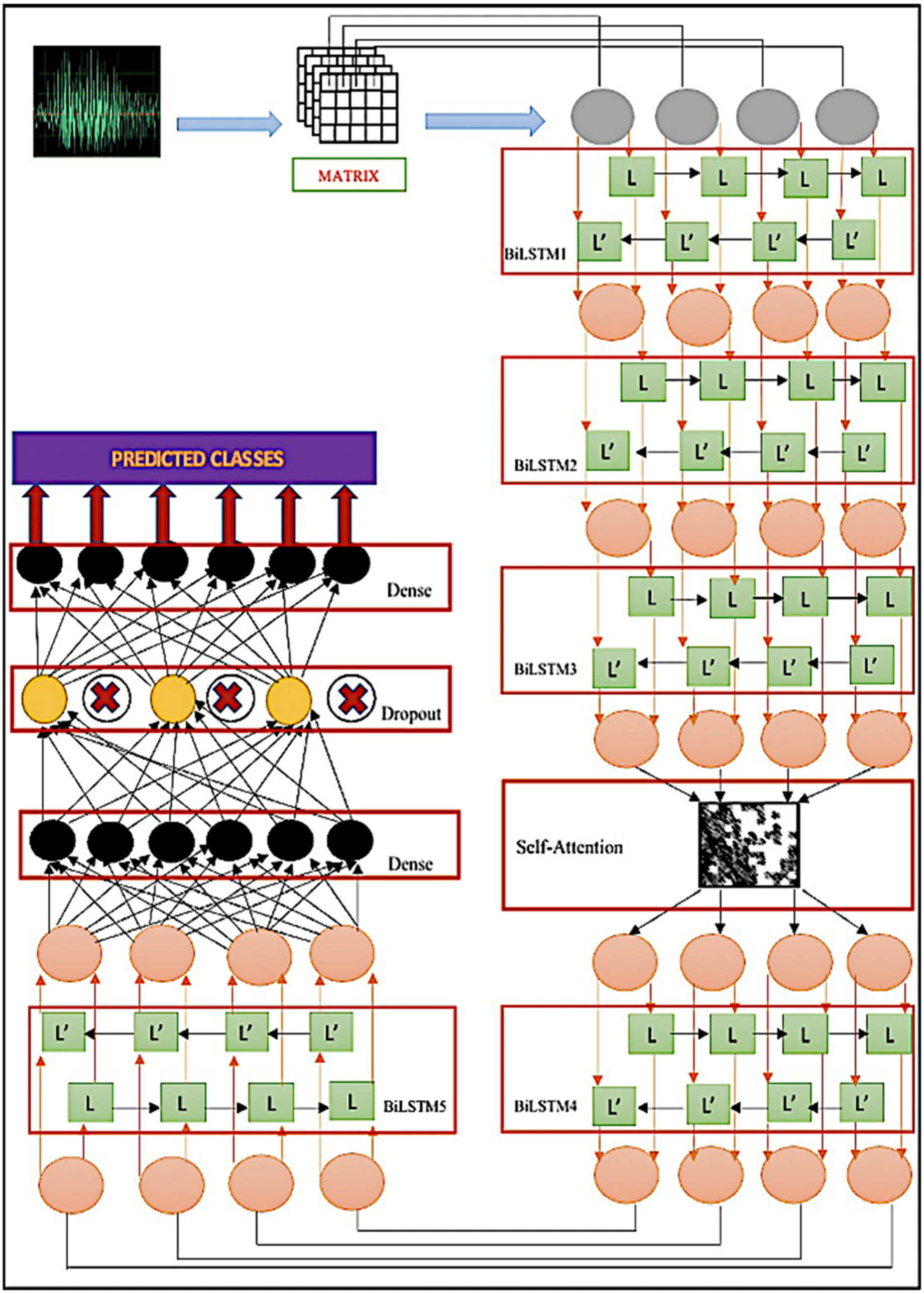


Fig. 2 Proposed architecture of deep BiLSTM with self-attention

spectrogram is (44 × 128), and the shape of the stem's GloVe word embedded vectors is (50 × 1). After extracting the acoustic and linguistic features (MFCC, Mel-spectrogram, stem), each feature is given as an input to a SA-deep BiLSTM model and the trained model is used to generate the posterior class probabilities for each test sample.

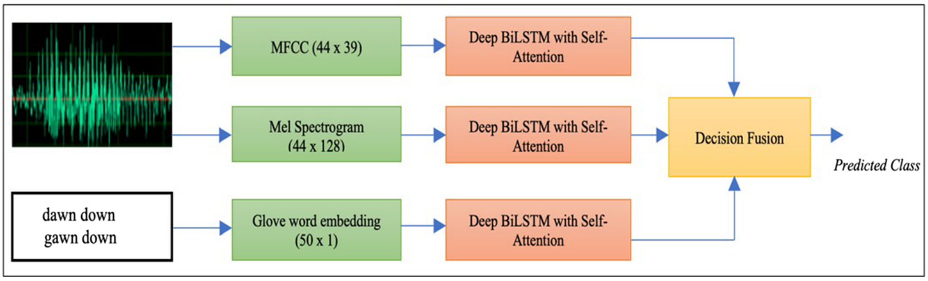

MFCC (44 x 39)
Mel Spectrogram (44 x 128)
Glove word embedding (50 x 1)
dawn down
gawn down
Deep BiLSTM with Self-Attention
Deep BiLSTM with Self-Attention
Deep BiLSTM with Self-Attention
Decision Fusion
Predicted Class

Fig. 3 Proposed deep fusion framework for speech command recognition using the acoustic and linguistic features

Our deep BiLSTM model consists of three initial BiLSTM layers with (512, 256, 128) units, self-attention layer (128) units, two high-level BiLSTM layers with (256, 128) units, and further, we have dense (32) layer, dropout (32) layer, and dense (10) layer. In total, we have 5 BiLSTM layers, hence making the model quite deep. The input features are fed to a forward LSTM, and in reverse order to a backward LSTM, thereby obtaining the forward and backward hidden states. The last layer in our SA-deep BiLSTM model is a dense layer which has the activation function "softmax", that generates a probability for each class prediction. In Fig. 2, L represents the cells of forward LSTM while L' represents the cells of backward LSTM, which together constitute one layer of BiLSTM.

Figure 3 depicts the flowgraph of the predictions made by the three deep models in our fusion framework. The decision fusion is performed using both maximum and average functions, as shown in Eqs. (11) and (12), to determine the more suitable choice of the two.

4 Experimental setup and results

4.1 Dataset and experimental settings

Our dataset is version 2 of the Google Speech Command Dataset [55]. It was released by Tensor-Flow and AIY teams, for the speech recognition challenge. The GSCD dataset has a variety of accents and speakers, which makes it harder to recognize spoken words. Because of this, the experiment is both difficult and applicable to real-world situations. Each audio file is a 16 kHz file in the .WAV format. The following speech commands are the 10-word target categories employed in our experiments: "YES", "NO", "UP", "DOWN", "LEFT", "RIGHT", "ON", "OFF", "STOP" and "GO" [55]. The training and testing folders are segregated at the source as shown in Fig. 4 which specifies the train: test split of each category. The unknown and silence categories are not considered for our experiments.

4.2 Results and discussions

The tests are carried out using the Python software 3.9.0 version, on a Mac running macOS High Sierra with an Intel Core i5 processor clocked at 1.8 GHz. We have made our code available online[1] to facilitate future research. The Librosa library is used to extract the MFCC

[1] https://github.com/sunakshimehra/Deep-Fusion-Framework-for-Speech-Command-Recognition-using-Acoustic-and-Linguistic-Features

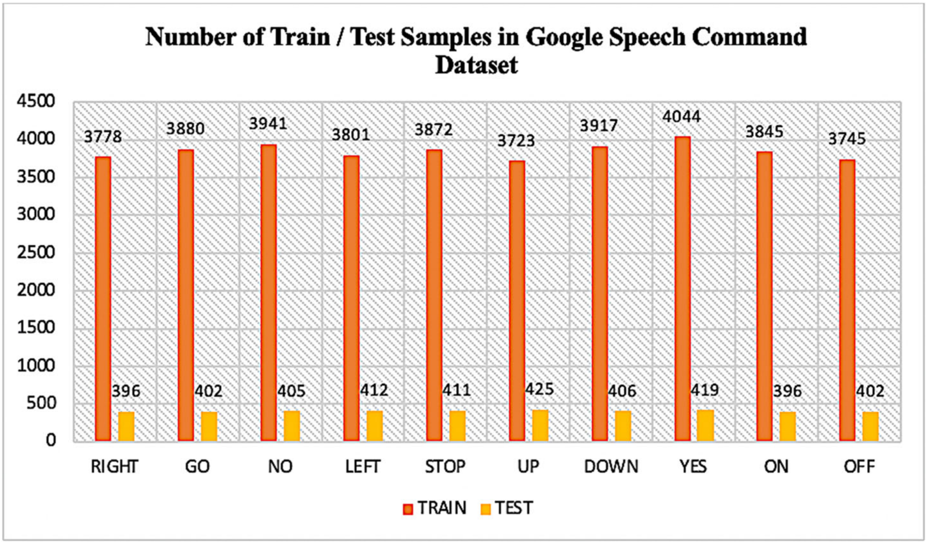


Fig. 4 Characteristics of Google speech command dataset

and Mel-spectrogram features. For extracting the MFCC features from a raw audio file, we use a hop length of 512 and 22,050 sample points per second. The 13 MFCC features are extracted and concatenated with delta (1st order) and delta-delta (2nd order) cepstral coefficients. After extracting the 39 MFCC features for timestamps fixed to a length of 44, the obtained feature matrix of dimension 44×39 is given as input to the proposed model SA-deep BiLSTM. Samples of the 2D representations of the frame-wise MFCC coefficients computed from raw audio files of the ten word categories of the Google speech command dataset are shown in Fig. 5.

The Mel-spectrogram is extracted from audio using the Librosa package by following the same process. Applying short-time Fourier transform to the speech signal, changing the amplitude to decibels, and then further converting the frequencies to Mel scale are the steps required to extract the Mel-spectrogram from the audio. The Mel-spectrogram 2D representations for samples of the ten word categories of the Google speech command dataset are shown in Fig. 6. The Mel-spectrogram feature matrix of dimension 44 × 128 derived for 44 timestamps from each raw audio file is passed as input to the SA-deep BiLSTM model. The trained model yields the probabilistic scores for each prediction. 100 epochs are chosen for each experiment. The activation function “ReLu” increases the model’s nonlinearity. The optimizer used to manage the learning rate for stochastic gradient descent is the Adam optimizer [14] known for its robust performance in various classification tasks. Adam optimizer calculates the current gradients by storing the average of past decaying gradients. The advantages of using Adam optimizer are fast convergence, and a healthy learning rate which does not vanish. The loss function used is sparse categorical cross- entropy.

The text transcript is used to record the linguistic characteristics. Using Google API, the written transcript of the speech is obtained. Stemming and conversion of the text’s words into 50-dimensional GloVe word embeddings are done before they are passed as input to the SA-deep BiLSTM model. As described in Section 3, the posterior class probabilities from the three deep models are combined.

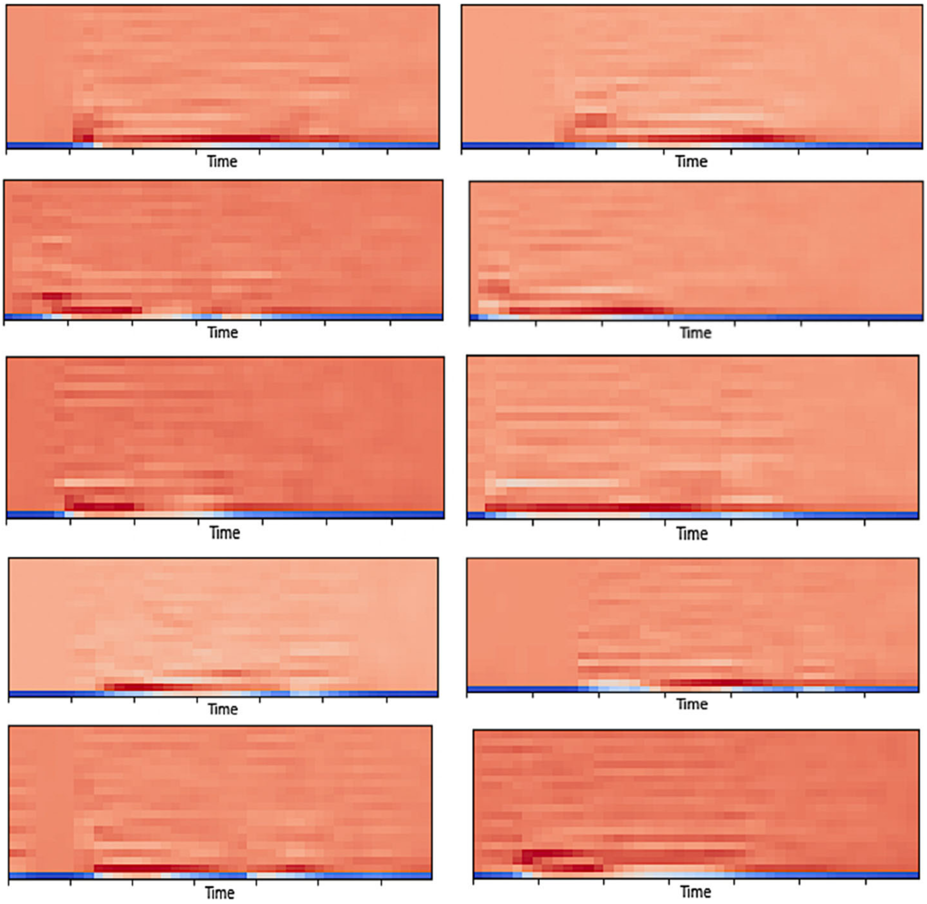


Fig. 5 Mel Frequency cepstral coefficients (2D-image representation) is shown for the ten word categories of the Google speech command dataset (from top to bottom and left to right)- "DOWN", "GO", "LEFT", "NO", "OFF", "ON", "RIGHT", "STOP", "UP" and "YES"

The proposed SA-deep BiLSTM model remains the same for each of the input modalities in our fusion framework. The probabilistic scores associated with each class are fused by soft fusion using the average or mean function, as shown in Eq. (11–13). The maximum function is less effective than the average function, as proved by results shown later on. The maximum (fused) probability indicates the class of the test sample. The model summaries for the acoustic and linguistic features used in our experiment are described in Table 1.

Table 2 compares the performance of our multimodal fusion approach to the state of the art. The outcomes demonstrate that in the current environment of adequately resourced data, our technique performs better than the alternatives, achieving the highest test accuracy of 98.80%. We compared our suggested approach with a convolutional neural network (CNN), whose input is provided as a 2D matrix of MFCCs features [17]. Our technique has been found to perform better at speech recognition than the MFCC with CNN by 5.52%. It has been noted that while using the Mel Spectrogram as an input feature, LSTMs perform well in sound classification [25]. Our suggested method, however, performs 3.73% better than Mel

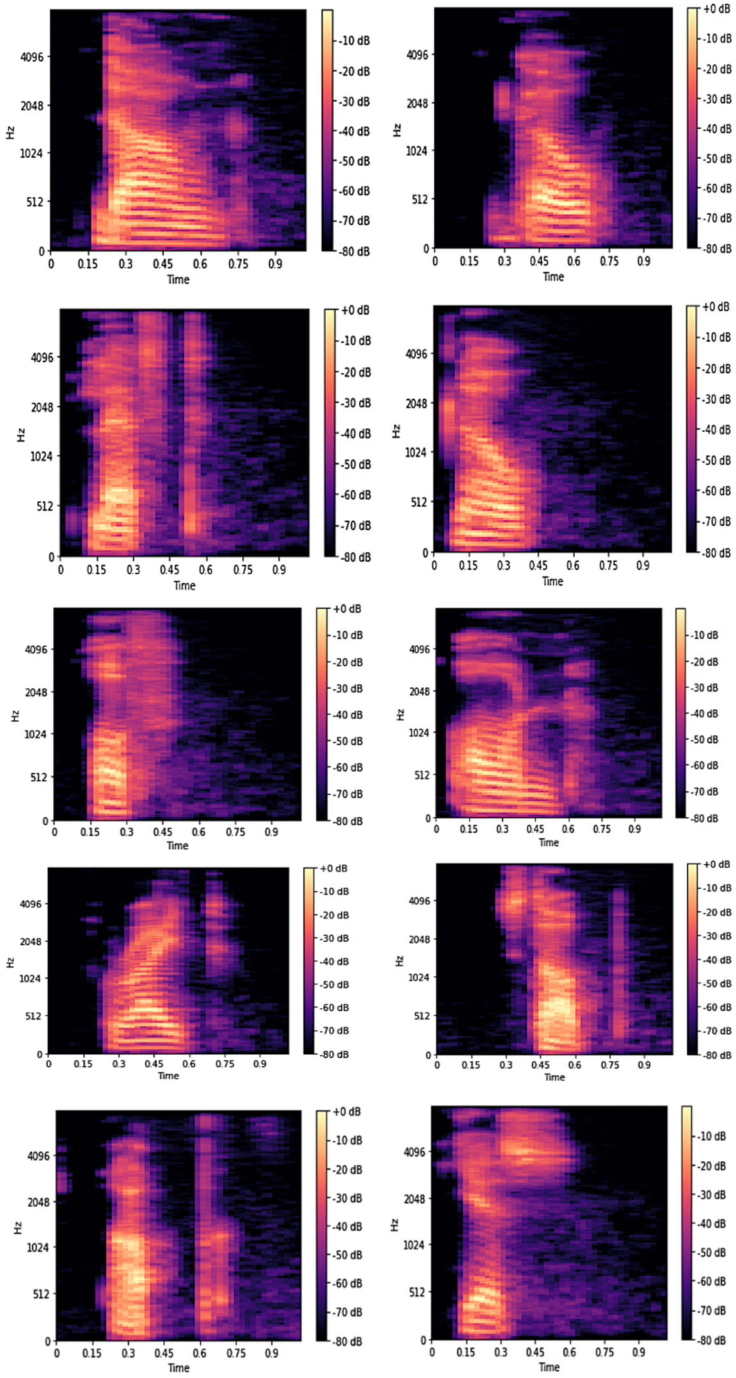


Fig. 6 Mel-spectrogram (2D-image representation) is shown for the ten word categories of the Google speech command dataset (from top to bottom and left to right)- “DOWN”, “GO”, “LEFT”, “NO”, “OFF”, “ON”, “RIGHT”, “STOP”, “UP” and “YES”

Table 1 SA-deep BiLSTM model summary for the MFCC, Mel-spectrogram and stem features

MFCC	Activation function	Output shape	Parameters
Layer (type)			
Bidirectional 1	_	(44, 512)	606,208
Bidirectional 2	_	(44, 256)	656,384
Bidirectional 3	_	(44, 128)	164,352
SeqSelfAttention	Sigmoid	(44, 128)	8257
Bidirectional 4	_	(44, 256)	263,168
Bidirectional 5	_	(128)	164,352
Dense 1	ReLu	(32)	4128
Dropout	_	(32)	0
Dense 2	Softmax	(10)	330
Total parameters: 1,867,179			
Trainable parameters: 1,867,179			
Non-trainable parameters: 0			
Mel-Spectrogram	Activation function	Output shape	Parameters
Layer (type)			
Bidirectional 1	_	(44, 512)	788,480
Bidirectional 2	_	(44, 256)	656,384
Bidirectional 3	_	(44, 128)	164,352
SeqSelfAttention	Sigmoid	(44, 128)	8257
Bidirectional 4	_	(44, 256)	263,168
Bidirectional 5	_	(128)	164,352
Dense 1	ReLu	(32)	4128
Dropout	_	(32)	0
Dense 2	Softmax	(10)	330
Total parameters: 2,050,276			
Trainable parameters: 2,050,276			
Non-trainable parameters: 0			
Stem2Vec	Activation function	Output shape	Parameters
Layer (type)			
Bidirectional 1	_	(50, 512)	528,384
Bidirectional 2	_	(50, 256)	656,384
Bidirectional 3	_	(50, 128)	164,352
SeqSelfAttention	Sigmoid	(128)	8257
Bidirectional 4	_	(256)	263,168
Bidirectional 5	_	(128)	164,352
Dense 1	ReLu	(32)	4128
Dropout	_	(32)	0
Dense 2	Softmax	(10)	330
Total parameters: 1,789,355			
Trainable parameters: 1, 789,355			
Non-trainable parameters: 0			

Spectrogram with LSTM [56]. The accuracy obtained with this method is 95.44%, which is 3.36% less accurate than ours.

EdgeCRNN [57] uses a feature-enhanced method that is based on residual structure and depth wise separable convolution. The accuracy obtained on 10 spoken word categories is 98.20%, which is 0.60% less accurate than our method. Our method outperformed [1] which used CNN and Gammatone Frequency Cepstral Coefficients as input by 5.71%. GFCCs are occasionally seen as superior signal representations for emotion perception [29]. The combination of DenseNet and BiLSTM was proposed recently by Zeng and Xiao [59] for keyword spotting. The experimental findings by Zeng and Xiao [59] demonstrate that DenseNet-Speech feature-maps effectively store time series information. DenseNet-BiLSTM can reach an accuracy of 94.88%; ours is better by 3.92%.

Table 2 Performance comparison with the state of the art for the 10-word category of Google Speech Command Dataset

Methods	Accuracy (%)
Attention based sequence to sequence model [18] (Higy and Bell, 2018)	97.50%
Semi supervised audio tagging [8] (Cances and Pellegrini, 2021)	95.58%
EdgeCRNN [57] (Wei et al., 2021)	98.20%
RNN Neural attention [13] (de Andrade et al., 2018)	94.11%
DenseNet + BiLSTM [59] (Zeng and Xiao, 2018)	94.88%
MFCC + LSTM-RNN [62] (Zia and Zahid, 2019)	95.14%
Mel Spectrogram with LSTM [25] (Lezhenin et al., 2019)	95.07%
MFCC + LSTM-RNN [56] (Wazir et al., 2019)	95.44%
GFCC + CNN [1] (Abdelmaksoud et al., 2021)	93.09%
MFCC + CNN [17] (Haque et al., 2020)	93.28%
Proposed Method	98.80%

We also compare our results with the LSTM architecture explored by Zia and Zahid in [62] for Urdu acoustic modelling. The utterances are preprocessed using the Python Speech Features Toolkit's MFCC approach. For comparison, a frame size of 10 ms, a frame shift of 5 ms, 40 filterbank channels, 20 cepstral coefficients, and 58 cepstral parameters are chosen. We outperform this method by 3.66%. For the Google Speech Command dataset, accuracy is presented as the mean of five runs; standard deviation, which is virtually always 0.1%, is not reported. Our method also outperformed the Deep CO-Training algorithm (DCT) [8] by 3.22%. Our approach transcends the attention convolutional recurrent neural network [13] by 4.69%. The architecture developed by [13] takes raw. WAV files as inputs, computes mel-scale spectrogram using a non-trainable Keras layer, extracts short- and long-term dependencies, and then employs an attention mechanism to determine which region contains the most useful information, which is then fed to a series of dense layers. On a short vocabulary keyword classification challenge, attention-based encoder-decoder models [18] have proved to outperform baselines, achieving 97.5% accuracy on TensorFlow's Speech Commands dataset. However, our method outperforms [18] by 1.30%.

It is summarized from Table 2 that with a high accuracy of 98.80%, the proposed method outperformed the state of the art for the categorization of the 10-word categories of the Google speech command dataset.

Table 3 Ablation-analysis of each method with accuracy score obtained for each class

Speech command	MFCC (%)	MS (%)	Stem (%)	OURS (%)
RIGHT	97.47	97.98	64.90	98.23
GO	97.26	95.77	52.99	97.26
NO	99.75	99.01	67.16	99.75
LEFT	99.51	99.76	47.57	100
STOP	99.51	99.76	53.28	99.51
UP	99.76	99.06	27.29	99.53
DOWN	96.06	96.80	49.26	96.80
YES	98.28	99.04	63.96	99.52
ON	98.23	97.73	44.95	98.48
OFF	98.26	98.26	97.26	98.76

Table 4 The results of soft fusion by averaging on combination of features for the 10-word Google speech command dataset

Combinatory results	Type of combination	Accuracy
Single Component	Stem	56.70%
Single Component	Lemma	56.31%
Single Component	MFCC	98.53%
Single Component	Mel-Spectrogram	98.33%
Two Components	MFCC + LEMMA	98.60%
Two Components	MS+LEMMA	98.43%
Two Components	MFCC + STEM	98.67%
Two Components	MS+STEM	98.53%
Two Components	MFCC + MS	98.70%
Two Components	LEMMA + STEM	62.13%
Three Components	MFCC + LEMMA + STEM	98.64%
Three Components	MS+LEMMA + STEM	98.43%
Three Components	MS+LEMMA + MFCC	98.72%
Three Components	MS+STEM + MFCC	98.80%

4.3 Ablation study

In this section, we investigate the impact of individual features (acoustic/linguistic), in an ablation study, on our SA-deep BiLSTM model. We investigate the significance of each acoustic/linguistic component (MFCC, Mel-spectrogram, stem), and pairwise combinations of these, for the proposed deep fusion framework involving our deep model: - SA-deep BiLSTM. The ablation analysis, whose results are summarized in Tables 3, 4, 5 yield the following observations.

- For the proposed deep framework, audio feature-based classification outperforms text-based classification.
- In our work, we have incorporated acoustic and text modalities as MFCC, Mel-spectrogram and stem, that have performed better than combinations of other popular acoustic/linguistic features such as GFCC, Log Mel Filter bank, Linear Predictive Cepstral Coefficients (LPCC) and lemma.

Table 5 The results of soft fusion by using maximum function on combination of features for the 10-word Google speech command dataset

Combinatory results	Type of combination	Accuracy
Single Component	Stem	56.70%
Single Component	Lemma	56.31%
Single Component	MFCC	98.53%
Single Component	Mel-Spectrogram	98.33%
Two Components	MFCC + LEMMA	98.55%
Two Components	MS+LEMMA	98.43%
Two Components	MFCC + STEM	98.60%
Two Components	MS+STEM	98.50%
Two Components	MFCC + MS	98.70%
Two Components	LEMMA + STEM	62.13%
Three Components	MFCC + LEMMA + STEM	98.60%
Three Components	MS+LEMMA + STEM	98.50%
Three Components	MS+LEMMA + MFCC	98.77%
Three Components	MS+STEM + MFCC	98.75%

- A key component in the recognition of spoken words is the self-attention module, which is inserted between the three initial BiLSTM layers and two higher-level BiLSTM layers of the proposed SA-deep BiLSTM model. Self-attention is used to highlight context in the input sequence that is specific to the classification task at hand.
- It was found that the MFCC and Mel-Spectrogram had individual accuracy values of 98.53% and 98.33%, respectively, which was increased to 98.80% following fusion.
- On applying stemming alone, the word recognition rate was observed to be 56.70%. However, including stemming in the proposed fusion framework boosted the accuracy to 98.80%. On substituting stemming with lemmatization, the accuracy dropped marginally to 98.72%. Lemmatization alone achieved an accuracy of 56.31% which is marginally lower than the performance of stemming.
- It is clear that the speech command “LEFT” was correctly identified with almost 100% accuracy using our suggested deep-BiLSTM attention strategy following soft fusion.
- The word categories “RIGHT,” “STOP,” “NO,” “LEFT,” “UP,” “YES,” “ON,” and “OFF” have shown high accuracies.
- The proposed technique was not able to significantly improve the accuracies of some word categories.

Table 3 yields the ablation study of accuracy obtained per word categories by applying stemming, MFCC, and Mel spectrogram individually and after decision-level fusion of all. It is evident that each word category’s recognition accuracy increased due to the soft fusion. The high accuracies could also be attributed to our SA-deep BiLSTM model that effectively learns from each modality separately.

Tables 4 and 5 provide the results of various homogeneous and heterogeneous combinations of the auditory and linguistic features used in our investigation. Morphological analysis of text is common in NLP and information retrieval [28, 35, 42, 53]. The stem and lemma are two common morphemes. The stemming algorithm is used to convert words from their affixes to their root form or morpheme, which is “stem,” in the text transcript [41]. The stemming technique supports vocabulary and text transcript size reduction in information retrieval. The WordNet lemmatizer is used to transform each word in the text transcript to its lemma, after which it is vectorized by 50-dimensional GloVe embedding, using the same process as in stemming [15]. Lemmatization is the process of removing the inflectional endings from a word to reveal its basic structure via morphological analysis. Lemmatization and stemming differ significantly in that lemmas carry contextual meaning, but stemming eliminates affixes without considering semantics.

The results of soft fusion by averaging (Eq. 12) for various combinations of acoustic/linguistic features is shown in Table 4. The results of soft fusion by using maximum function (Eq. 11) for various combinations of acoustic/linguistic features is shown in Table 5. We consequently draw the conclusion from our ablation study that improved spoken word categorization can be achieved by combining acoustic and linguistic modalities. Table 4 shows that, of the three feature types (MFCC, Mel-Spectrogram, and stem), the stem performed the least well since faulty transcriptions from raw audio files were used. The average score and maximum score have improved by 0.10% and 0.10%, respectively, as a result of the combination of acoustic components (MFCC and Mel- Spectrogram). However, the stem and lemma morphological features when combined have resulted in improvements of 5.43% and 5.82% over the respective baselines. The improvement over separate techniques are up to 42.10% for Stem, 42.49% for lemma, 0.27% for MFCC, and 0.47% for Mel-Spectrogram, for the average metric score, when MFCC, Mel-Spectrogram, and stem are combined as the recommended optimal fusion.

Table 6 displays the Precision, Recall and f1-scores for each word category. Precision or positive predictive value (PPV) is the ratio of the number of true positives to the total number of positives detected by the model as defined in Eq. (14).

$$precision/positive\ predictive\ value\ (PPV) = \frac{tp}{[tp + fp]} \tag{14}$$

The recall is true positive rate, also known as sensitivity. It is represented as

$$sensitivity/recall = \frac{tp}{[tp + fn]} \tag{15}$$

Here, *tp*, *tn*, *fn*, *fp* mean true positive, true negative, false negative and false positive values, respectively. The harmonic mean of precision and recall values is the f1-Score. The formula for the f1-Score is

$$f1{-}Score = \frac{2 \times precision \times recall}{[precision + recall]} \tag{16}$$

In our experiment, precision, recall and f1-Score each have high values, which indicates a good classification performance.

Table 7 displays the proposed deep BiLSTM's classification report for all word categories. The evaluation metrics listed in Table 7 for each word class include accuracy, negative predictive value (NPV), false positive rate (FPR), false negative rate (FNR), and false discovery rate (FDR). The percentage of cases with negative test findings that are already valid samples is known as the NPV. It determines the percentage of subjects that were genuinely scanned as negative in relation to all other test-negative participants (including samples that were incorrectly test as correct samples). The mathematical notation for NPV is shown in Eq. (17).

$$negative\ predictive\ value\ (NPV) = \frac{tn}{[fn + tn]} \tag{17}$$

Table 6 Statistical analysis per word-category

Categories	Precision	Recall	f1-Score
RIGHT	1.00	0.98	0.99
GO	1.00	0.97	0.99
NO	0.99	1.00	0.99
LEFT	0.96	1.00	0.98
STOP	0.99	1.00	0.99
UP	0.97	1.00	0.98
DOWN	0.99	0.97	0.98
YES	1.00	1.00	1.00
ON	1.00	0.98	0.99
OFF	0.98	0.98	0.99

Table 7 Classification report per word-category without roundoff

Categories	FNR	FDR	PPV	NPV	FPR
RIGHT	0.00	0.01	0.98	1.00	0.00
GO	0.00	0.02	0.97	1.00	0.00
NO	0.00	0.00	0.99	0.99	0.00
LEFT	0.04	0.00	1.00	0.99	0.00
STOP	0.00	0.00	0.99	0.99	0.00
UP	0.02	0.00	0.99	0.99	0.00
DOWN	0.00	0.03	0.96	0.99	0.00
YES	0.00	0.00	0.99	0.99	0.00
ON	0.00	0.01	0.98	1.00	0.00
OFF	0.01	0.01	0.98	0.99	0.00

The NPV value is 1 (100%) in a flawless test, one that yields no false negative results, and 0 (0%) in a test that yields no true negative results. The best possible PPV result in a perfect test is 1 (100%), and the worst result is zero. Also known as the conditional likelihood of a negative test result given the presence of the positives being tested for, the false negative rate is the proportion of positive test results that result in a negative test result. The false discovery rate is the anticipated proportion of type I errors (FDR). FPR, FNR, FDR are represented as

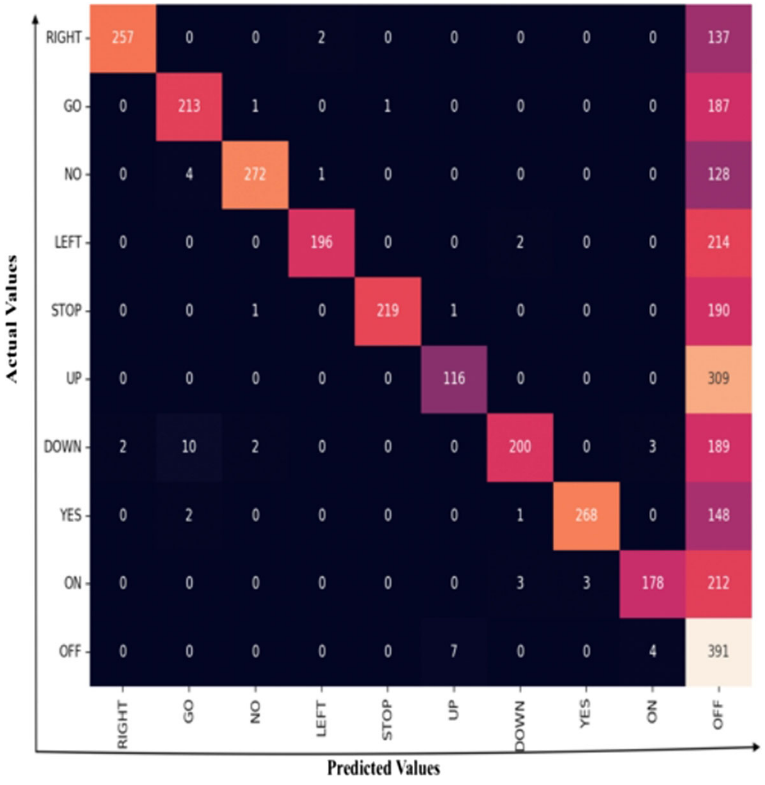
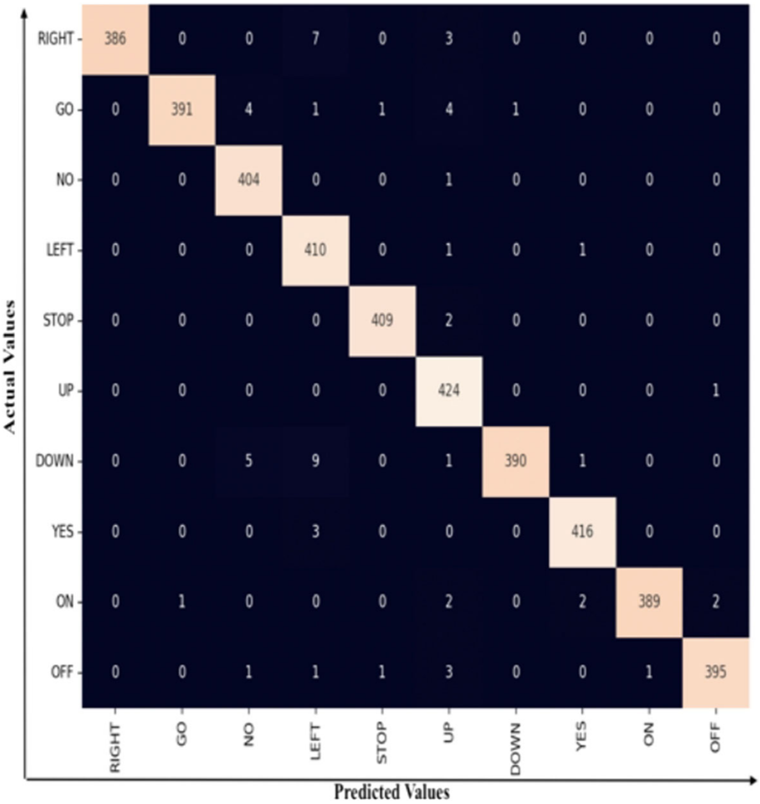


Fig. 7 Confusion matrix of stemming



	RIGHT	GO	NO	LEFT	STOP	UP	DOWN	YES	ON	OFF
RIGHT	386	0	0	7	0	3	0	0	0	0
GO	0	391	4	1	1	4	1	0	0	0
NO	0	0	404	0	0	1	0	0	0	0
LEFT	0	0	0	410	0	1	0	1	0	0
STOP	0	0	0	0	409	2	0	0	0	0
UP	0	0	0	0	0	424	0	0	0	1
DOWN	0	0	5	9	0	1	390	1	0	0
YES	0	0	0	3	0	0	0	416	0	0
ON	0	1	0	0	0	2	0	2	389	2
OFF	0	0	1	1	1	3	0	0	1	395

Fig. 8 Confusion matrix of MFCC

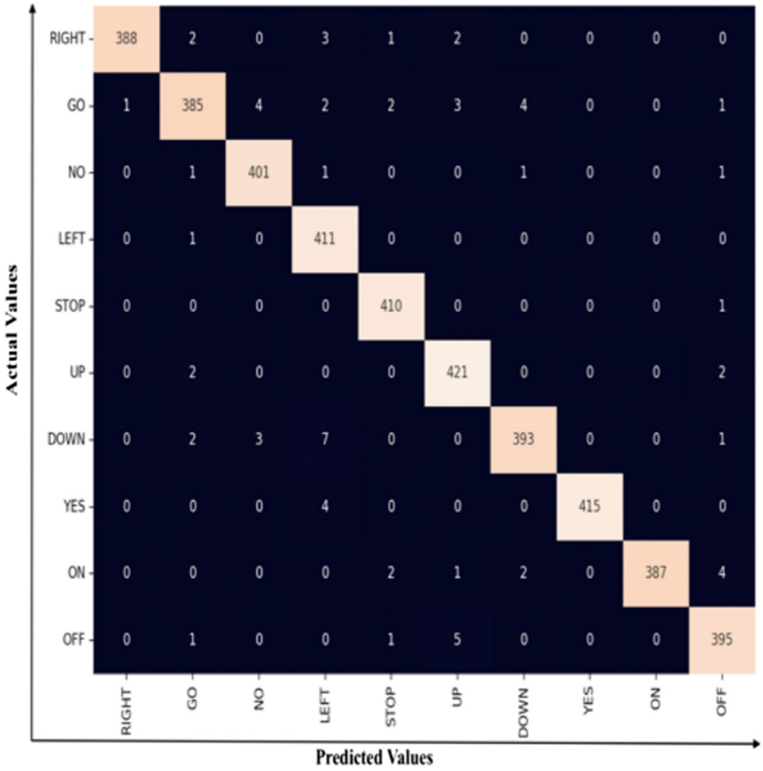
$$false\ positive\ rate\ (FPR) = \frac{fp}{[fp + tn]} \tag{18}$$

$$false\ negative\ rate\ (FNR) = \frac{fn}{[fn + tp]} \tag{19}$$

$$false\ discovery\ rate\ (FDR) = \frac{fp}{[fp + tp]} \tag{20}$$

Table 7 shows that almost all classes have high specificity and NPV, while some classes also have high sensitivity and PPV. The FPR values are consistently low across all classes, and the FNR and FDR values are on the lower side. The word category- LEFT performs the best overall.

Figures 7, 8, 9, and 10 show the confusion matrices of stemming, MFCC, Mel- Spectrogram, and soft fusion for the proposed SA-deep BiLSTM model. On the x-axis, we have predicted values and, on the y-axis, we have actual values. It can be easily observed from the confusion matrix that the stemming algorithm has poorly performed because of the inadequate information gathered from Google API while transcribing. It is verified from Fig. 10 that the



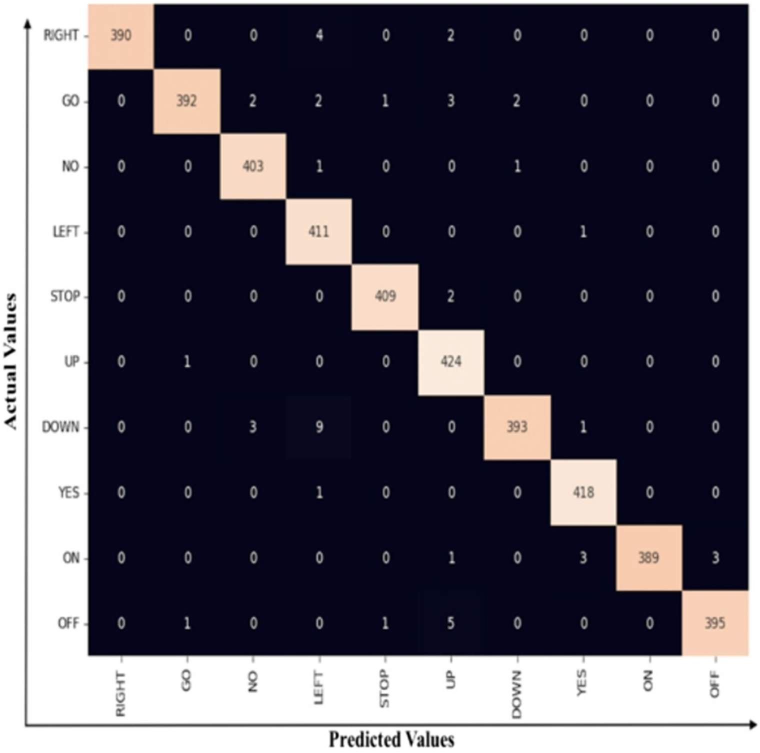
	RIGHT	GO	NO	LEFT	STOP	UP	DOWN	YES	ON	OFF
RIGHT	388	2	0	3	1	2	0	0	0	0
GO	1	385	4	2	2	3	4	0	0	1
NO	0	1	401	1	0	0	1	0	0	1
LEFT	0	1	0	411	0	0	0	0	0	0
STOP	0	0	0	0	410	0	0	0	0	1
UP	0	2	0	0	0	421	0	0	0	2
DOWN	0	2	3	7	0	0	393	0	0	1
YES	0	0	0	4	0	0	0	415	0	0
ON	0	0	0	0	2	1	2	0	387	4
OFF	0	1	0	0	1	5	0	0	0	395

Fig. 9 Confusion matrix of mel-spectrogram

proposed approach achieves the highest accuracies for all 10 categories. In Table 8, we have shown the misclassified words confused with other categories. And, the word category LEFT has 1 misclassifications, hence number of errors are one. GO category is misclassified by NO, LEFT, STOP, UP, DOWN, and reaches to 10 in number of errors, where the number of errors caused is 1, 2, 2, 2, 3 which sums up to 10. DOWN category is the most misclassified. The number of errors is 13, highest among others.

4.4 Implementation challenges

Working with multiple modalities, such as signal, speech, text, face and motion, is both intriguing and challenging because there are several implementation changes. The first challenge is the computational complexity involved for training the deep network from scratch for large datasets. Our framework presents a partial solution in the form of independent learning for different modalities. It should be noted that the computational complexity depends on the length of the input rather than the machine's real processing speed. The operation complexity of our model is as follows-: the operations for each BiLSTM layer are O (Lp^2), self-attention layer is O (L^2p) where p is the model dimension of hidden states and L is the length of the input features. We can reduce the complexity of the model by using restricted self-attention, at the cost of reduce in accuracy. For larger datasets, use of smaller mini-batch sizes is recommended.



	RIGHT	GO	NO	LEFT	STOP	UP	DOWN	YES	ON	OFF
RIGHT	390	0	0	4	0	2	0	0	0	0
GO	0	392	2	2	1	3	2	0	0	0
NO	0	0	403	1	0	0	1	0	0	0
LEFT	0	0	0	411	0	0	0	1	0	0
STOP	0	0	0	0	409	2	0	0	0	0
UP	0	1	0	0	0	424	0	0	0	0
DOWN	0	0	3	9	0	0	393	1	0	0
YES	0	0	0	1	0	0	0	418	0	0
ON	0	0	0	0	0	1	0	3	389	3
OFF	0	1	0	0	1	5	0	0	0	395

Fig. 10 Confusion matrix of proposed approach

Another obstacle in the experiments was the error-prone text transcripts obtained from online speech translators like Google API. Inclusion of the acoustic modality derived from raw audio in our deep framework helped to mitigate the errors induced due to this problem, to a good extent. Accents present one of speech's biggest challenges. Another factor that makes speech recognition a challenging task is the variability and diversity of the speakers. Further, the variety of phonemes, including vowels and diphthongs, in any language affects pronunciation, translation, word recognition and keyword tagging. ASR development can also be hampered by a lack of utterances, disorganized speech, or simple machine faults.

Table 8 List of misclassified words

Category	Confused with other categories	Number of errors
RIGHT	LEFT, UP	6
GO	NO, LEFT, STOP, UP, DOWN	10
NO	LEFT, DOWN	2
LEFT	YES	1
STOP	UP	2
UP	GO	1
DOWN	NO, LEFT, YES	13
YES	LEFT	1
ON	YES, UP, OFF	7
OFF	GO, STOP, UP	7

5 Conclusion and future work

In this paper, we propose late fusion of audio and text modalities using a novel SA-deep BiLSTM model for learning each modality separately. With 10-word categories from the Google speech command dataset, we were able to obtain an accuracy of 98.80% by training each modality on a deep self-attention BiLSTM model. We describe a soft fusion method based on posterior class probabilities for the stem (a linguistic feature) and MFCC and Mel-spectrogram (acoustic features) taken from each audio file. For training each type of feature, we suggest a deep model called SA-deep BiLSTM, which consists of five BiLSTM layers and a self-attention module between the third and fourth layers. In terms of classification accuracy, the suggested fusion method performs better than the current state of the art for spoken word recognition. It was observed that the word category “LEFT” was almost completely correctly predicted by our suggested deep fusion framework.

For the suggested deep fusion paradigm, it would be interesting to investigate early-cum-late fusion in the future [50]. The disadvantage of speech transcriptions is that during Google speech translation, a considerable amount of audio-to-text data is lost. Using a speech-to-text conversion method that produces fewer errors can significantly enhance results can be used in future scope. Working with articulatory features [31] and background noise are possible extensions of this work. In our work, we have merged linguistic and acoustic elements such that they work in harmony and supplement the information that one modality lacks. So, by combining the acoustic and linguistic information, and incorporating in a deep fusion framework, it is evident that spoken word classification is achieved more accurately, since it captures the information contributed by both audio and text modalities effectively.

Data availability The data will be made available by the authors on request.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

References

1. Abdelmaksoud ER, Hassen A, Hassan N, Hesham M (2021) Convolutional neural network for arabic speech recognition. The Egypt J Lang Eng 8(1):27–38
2. Aldarmaki H, Ullah A, Ram S, Zaki N (2022) Unsupervised automatic speech recognition: a review. Speech Comm 139:76–91
3. Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473
4. Bastanfard, Azam, Mohammad Aghaahmadi, Maryam Fazel, and Maedeh Moghadam (2009) Persian viseme classification for developing visual speech training application. In Pacific-Rim Conference on Multimedia, 1080–1085. Springer, Berlin, Heidelberg
5. Bastanfard A, Amirkhani D, Naderi S (2020) A singing voice separation method from Persian music based on pitch detection methods. In 2020 6th Iranian conference on signal processing and intelligent systems (ICSPIS), 1–7. IEEE
6. Boigne J, Liyanage B, Östrem T (2020) Recognizing more emotions with less data using self-supervised transfer learning. arXiv preprint arXiv:2011.05585
7. Cabrera R, Liu X, Ghodsi M, Matteson Z, Weinstein E, Kannan A (2021) Language model fusion for streaming end to end speech recognition. arXiv preprint arXiv:2104.04487

8. Cances L, Pellegrini T (2021) Comparison of deep co-training and mean-teacher approaches for semi-supervised audio tagging. In ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 361–365. IEEE
9. Cheng J, Dong L, Lapata M (2016) Long short-term memory-networks for machine reading. arXiv preprint arXiv:1601.06733
10. Chi P-H, Chung V, Wu T-H, Hsieh C-C, Chen Y-H, Li S-W, Lee H-y (2021) Audio albert: A lite bert for self-supervised learning of audio representation. In 2021 IEEE Spoken Language Technology Workshop (SLT), 344–350. IEEE
11. Cui Z, Ke R, Ziyuan P, Wang Y (2020) Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values. Transport Res Part C: Emerg Technol 118:102674
12. Davis S, Mermelstein P (1980) Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans Acoust Speech Signal Process 28(4):357–366
13. De Andrade DC, Leo S, Da Silva Viana ML, Bernkopf C (2018) A neural attention model for speech command recognition. arXiv preprint arXiv:1808.08929
14. Duchi J, Hazan E, Singer Y (2011) Adaptive subgradient methods for online learning and stochastic optimization. J Mach Learn Res 12(7)
15. Fellbaum C (1998) A semantic network of English verbs. WordNet: An electronic lexical database 3:153–178
16. Gallardo-Antolín A, Montero JM (2021) On combining acoustic and modulation spectrograms in an attention LSTM-based system for speech intelligibility level classification. Neurocomputing 456:49–60
17. Haque MA, Verma A, Alex JSR, Venkatesan N (2020) Experimental evaluation of CNN architecture for speech recognition. In: In First international conference on sustainable technologies for computational intelligence. Springer, Singapore, 507–514
18. Higy B, Bell P (2018) Few-shot learning with attention-based sequence-to-sequence models. arXiv preprint arXiv:1811.03519
19. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
20. Hyder R, Ghaffarzadegan S, Feng Z, Hansen JHL, Hasan T (2017) Acoustic scene classification using a CNN-supervector system trained with auditory and spectrogram image features. In Interspeech, 3073–3077
21. Kardakis S, Perikos I, Grivokostopoulou F, Hatzilygeroudis I (2021) Examining attention mechanisms in deep learning models for sentiment analysis. Appl Sci 11(9):3883
22. Kim S, Shangguan Y, Mahadeokar J, Bruguier A, Fuegen C, Seltzer ML, Le D (2021) Improved neural language model fusion for streaming recurrent neural network transducer. In ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 7333–7337. IEEE, 2021
23. Kumaran U, Radha Rammohan S, Nagarajan SM, Prathik A (2021) Fusion of mel and gammatone frequency cepstral coefficients for speech emotion recognition using deep C-RNN. International Journal of Speech Technology 24(2):303–314
24. Le D, Jain M, Keren G, Kim S, Shi Y, Mahadeokar J, Chan J, et al. (2021) Contextualized streaming end-to-end speech recognition with trie-based deep biasing and shallow fusion. arXiv preprint arXiv:2104.02194
25. Lezhenin I, Bogach N, Pyshkin E (2019) Urban sound classification using long short-term memory neural network. In 2019 federated conference on computer science and information systems (FedCSIS), 57–60. IEEE
26. Li J, Han L, Li X, Zhu J, Yuan B, Gou Z (2021) An evaluation of deep neural network models for music classification using spectrograms. Multimed Tools Appl 81:1–27
27. Lin Z, Feng M, dos Santos CN, Yu M, Xiang B, Zhou B, Bengio Y (2017) A structured self-attentive sentence embedding. arXiv preprint arXiv:1703.03130
28. Lin JC-W, Shao Y, Djenouri Y, Yun U (2021) ASRNN: a recurrent neural network with an attention model for sequence labeling. Knowledge-Based Systems 212:106548
29. Liu GK (2018) Evaluating gammatone frequency cepstral coefficients with neural networks for emotion recognition from speech. arXiv preprint arXiv:1806.09010
30. Macary M, Tahon M, Estève Y, Rousseau A (2021) On the use of self-supervised pre-trained acoustic and linguistic features for continuous speech emotion recognition. In 2021 IEEE Spoken Language Technology Workshop (SLT), 373–380. IEEE
31. Mahdavi R, Bastanfard A, Amirkhani D (2020) Persian accents identification using modeling of speech articulatory features. In 2020 25th international computer conference, Computer Society of Iran (CSICC), 1–9. IEEE
32. Marslen-Wilson WD (1987) Functional parallelism in spoken word-recognition. Cognition 25(1–2):71–102
33. McDermott E, Sak H, Variani E (2019) A density ratio approach to language model fusion in end-to-end automatic speech recognition. In 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 434–441. IEEE

34. Meghanani A, Anoop CS, Ramakrishnan AG (2021) An exploration of log-mel spectrogram and MFCC features for Alzheimer's dementia recognition from spontaneous speech. In 2021IEEE Spoken Language Technology Workshop (SLT), 670–677. IEEE
35. Mehra S, Susan S (2020) Improving word recognition in speech transcriptions by decision-level fusion of stemming and two-way phoneme pruning. In International Advanced Computing Conference, 256–266. Springer, Singapore
36. Minoofam SAH, Bastanfard A, Keyvanpour MR (2021) TRCLA: A transfer learning approach to reduce negative transfer for cellular learning automata. IEEE Trans Neural Netw Learn Syst
37. Nagrani A, Yang S, Arnab A, Jansen A, Schmid C, Sun C (2021) Attention bottlenecks for multimodal fusion. arXiv preprint arXiv:2107.00135
38. Oganyan M, Wright RA (2022) The role of the root in spoken word recognition in Hebrew: an auditory gating paradigm. Brain Sci 12(6):750
39. Pennington J, Socher R, Manning CD (2014) Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 1532–1543
40. Phaye SSR, Benetos E, Wang Y (2019) Subspectralnet–using sub- spectrogram based convolutional neural networks for acoustic scene classification. In ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 825–829. IEEE
41. Porter M (1999) Porter stemming algorithm. 2012-12-06]. http://tartarus.org/-martin/PorterStemmer
42. Ravuri S, Stolcke A (2015) Recurrent neural network and LSTM models for lexical utterance classification. In Sixteenth Annual Conference of the International Speech Communication Association
43. Sakashita Y, Aono M (2018) Acoustic scene classification by ensemble of spectrograms based on adaptive temporal divisions. Detection and Classification of Acoustic Scenes and Events(DCASE) Challenge
44. Schneider S, Baevski A, Collobert R, Auli M (2019) wav2vec: Unsupervised pre-training for speech recognition. arXiv preprint arXiv:1904.05862
45. Shah VH, Chandra M (2021) Speech recognition using spectrogram-based visual features. In Advances in Machine Learning and Computational Intelligence, 695–704. Springer, Singapore
46. Shen J, Pang R, Weiss RJ, Schuster M, Jaitly N, Yang Z, Chen Z, et al. (2018) Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP), 4779–4783. IEEE
47. Su Y, Zhang K, Wang J, Zhou D, Madani K (2020) Performance analysis of multiple aggregated acoustic features for environment sound classification. Appl Acoust 158:107050
48. Susan S, Kaur A (2017) Measuring the randomness of speech cues for emotion recognition. In 2017 Tenth International Conference on Contemporary Computing (IC3), 1–6. IEEE
49. Susan S, Malhotra J (2019) CNN pre-initialization by minimalistic part-learning for handwritten numeral recognition. In International Conference on Mining Intelligence and Knowledge Exploration, 320–329. Springer, Cham
50. Susan S, Malhotra J (2021) Learning image by-parts using early and late fusion of auto-encoder features. Multimed Tools Appl 80(19):29601–29615
51. Susan S, Sharma S (2012) A fuzzy nearest neighbor classifier for speaker identification. In: 2012 Fourth International Conference on Computational Intelligence and Communication Networks, IEEE, pp 842–845
52. Tripathi M, Singh D, Susan S (2020) Speaker recognition using SincNet and X-Vector fusion. In International Conference on Artificial Intelligence and Soft Computing, 252–260. Springer, Cham
53. Tur G, De Mori R (2011) Spoken language understanding: Systems for extracting semantic information from speech. John Wiley & Sons
54. Veisi H, Ghoreishi SA, Bastanfard A (2021) Spoken term detection for Persian news of Islamic Republic of Iran broadcasting. Signal and Data Processing 17(4):67–88
55. Warden P (2018) Speech commands: A dataset for limited-vocabularyspeech recognition. arXiv preprint arXiv:1804.03209
56. Wazir ASMB, Chuah JH (2019) Spoken arabic digits recognition using deep learning. In 2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), 339–344. IEEE
57. Wei Y, Zheng G, Yang S, Ye K, Wen Y (2021) EdgeCRNN: an edge-computing oriented model of acoustic feature enhancement for keyword spotting. J Ambient Intell Humaniz Comput 13:1–11
58. Yi C, Zhou S, Bo X (2021) Efficiently fusing pretrained acoustic and linguistic encoders for low-resource speech recognition. IEEE Signal Processing Letters 28:788–792
59. Zeng M, Xiao N (2019) Effective combination of DenseNet and BiLSTM for keyword spotting. IEEE Access 7:10767–10775
60. Zhang S, Yi J, Tian Z, Bai Y, Tao J (2021) Decoupling pronunciation and language for end-to-end code-switching automatic speech recognition. In ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 6249–6253. IEEE

61. Zheng R, Chen J, Ma M, Huang L (2021) Fused acoustic and text encoding for multimodal bilingual pretraining and speech translation. arXiv preprint arXiv:2102.05766
62. Zia T, Zahid U (2019) Long short-term memory recurrent neural network architectures for Urdu acoustic modeling. International Journal of Speech Technology 22(1):21–30

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

Kochi Chapter

Indian Geotechnical Conference
IGC 2022
15th – 17thDecember, 2022, Kochi

Deformation response of Twin Tunnels under the effect of static loading conditions.

Parvesh Kumar[1*], Amit Kumar Shrivastava[1]
[1]Department of Civil Engineering, Delhi Technological University, Delhi, India
*Corresponding Author: parvesh.kaushal2@gmail.com

Abstract. Underground structures can be used in various application such as trans- portation, sewage, gas pipelines, military purposes etc. Design parameters of under- ground structures play an important role in the stability of structures. Underground structures are subjected to various type of loads such as static, dynamic etc. So, the stability of underground structures is a major topic in order to keep structure safe in various loading conditions.This paper presents the deformation behavior of Twin Tunnels under the effect of static loading conditions. The objective of this work is to simulate the in-situ conditions through physical modeling. The spacing between the Twin Tunnels models is varied as 1.5D, 2D and 2.5D where "D" is the diameter of the tunnel. Both lined and unlined samples of Twin tunnel are prepared in laboratory. Plaster of Paris is used for making tunnel models. From the results it may be con- cluded that deformation of twin tunnels largely depends upon the spacing between the two tunnels. The results which are obtained from tests are studied for computation of stress and deformation in tunnels.

Keywords: Underground Structures, Deformation, Static Loading, Geo-material.

1. Introduction

Tunnels are horizontal, man-made underground passages that can be constructed without affecting the surrounding surface. Materials are typically transported through tunnels. Tunnels can be built through rocky terrain, including hills, rivers, etc. Tunnels are utilised for many different things today. There are many different uses for tunnels, including for highways, railroads, sewage and water supply tunnels, underground power plants, storage facilities, etc. Given the wide range of underground applications, it is crucial to take into account the many facets of underground openings as well as their stress and deformation characteristics. Any opening initially stresses rock, which creates early stress.The construction of the underground tube is quite old. In general, a tunnel structure is needed when a railway or highway route encounters an obstruction. In ancient times, tunnels were created to deliver pure water to key cities. Such tunnels are still used for the same purpose in Jammu and Kashmir, Egypt, Greece, Rome, and other places. Historically, manual mining techniques were used to create a number of tunnels in hard rock. Timber was employed as a temporary support to ensure the security of the tunnel workers. Brunel created the tunnelling shield in the 19th century, which prevented numerous fatalities due to timber collapses. The tunnelling technique, which is still in use today, was somewhat modified in the 20th century and given the term "open-faced" approach. Later in the 20th century, circular tunnel linings were employed to transport the weight of the rock and soil. The first tunnel was built some 4000 years ago. That tunnel was built in Babylon to connect two structures. Both the Egyptians and the Babylonians built it. That tunnel measured 3.6 metres in width, 4.5 metres in height, and 910 metres in length. Claudius, the Roman emperor, constructed the first tunnel in Europe later on to transport spring water through the Appennine Mountains.Chehade and Shahrour (2008) conducted a study on the interaction behaviour of twin tunnels with the help of numerical software. From the results, it can be concluded that higher settlement will occur if the upper tunnel is constructed first. Elshamy et al. (2013) conducted a study to determine the effect of different shapes of the tunnel on its deformation behavior. From the result obtained from the study, it is noticed that the circular tunnel is the best shape of twin tunnels. Yingjie et al. (2014) investigated the failure process of weak rock surrounding the tunnel using physical and computational methods. According to the findings, the weak rocks surrounding the tunnel fail predominantly due to shear wedge failure in the minimal principle stress direction, causing the tunnel arch to collapse. Oliaei and Manafi (2015) noticed that vertically aligned tunnels experienced the maximum settlement. Bayoumi et al. (2016) determine the effect of the construction of a twin tunnel on the structure with the help of PLAXIS 2D software. From the results, it has been concluded that the construction procedure affects the settlement of vertical twin tunnels.Kumar and Shrivastava (2017) and Kumar and Shrivastava (2019) reviewed the various factors which affect the stability of underground structures. A study conducted by Shrivastava and Rao (2011), Shrivastava and Rao (2015) and Shrivastava and Rao (2018) concluded that the shear behaviour of infilled rock joints depends upon the thickness of the infill material. Singh et al. (2018) conducted a numerical investigation to analyze the spacing and diameter effect on the stability of twin tunnels and concluded that the minimum spacing for twin tunnels should be 0.8 times the diameter of the circular opening. Kumar and Shrivastava (2021) conducted a study on the deformation behaviour of a single tunnel under static loading conditions and concluded that in the case of tunnels at shallow depths the extent of the damage along the tunnel axis depends upon the strength characteristics of the rock. Mishra et al. (2018) discuss the effect on the shallow tunnel under static and dynamic loading. The result shows that the strength of the rock plays an important factor in stability behaviour.Mishra et al. (2021) conducted a study to investigate the stability behaviour of a single tunnel in soft rock and found that depth of tunnel, intensity of drop load and strength of rock decide the extent of deformation in tunnel. Kumar and Shrivastava (2022) conducted a comparative study on the deformation behaviour of single and twin tunnels and concluded UCS value of the model material plays an important role in the deformation of tunnel.

1.1 Selection of the Model Material.

Finding a model material that can be utilised to imitate actual rock conditions is the main obstacle encountered during the testing phase because it is very difficult to incorporate all the challenges that must be confronted in the field circumstances in the laboratory. As a result, a material that can be utilised to prepare rock tunnel samples and imitate real-world field conditions is discovered in order to address this issue. Plaster of Paris is chosen as the model material because it is commonly available and has the ability to mould into any shape when mixed with water. Kumar and Shrivastava (2021) used plaster of paris as a model material in creating rock tunnels models. The compressive strength plaster of paris is around 8MPa which is greater than 1MPa therefore it represents the rock behavior. According to Deere Miller classification (1968) the classification of plaster of paris is done as EM (Medium Elastic). The following Table 1 gives a summary of properties of model material.

Table 1.Properties of Model Material

Properties	Value	Testing Method
Dry Density (kN/m^3)	12.19	ISRM (1972)
UCS (MPa)	10.6	ISRM(1979)
Modulus E_{t50} (MPa)	2510	ISRM(1979)
Tensile Strength (MPa)	0.79	ISRM(1979)
Deere–Miller Classification (1968)	*EM*	Deere Miller classification (1968)

1.2 Fixing dimension of tunnel models

The twin tunnel sample measures 425x375x230mm in size (LxWxH). The boundary conditions, i.e., r=4a, where "a" is the tunnel's radius, indicate the twin tunnel's width. Three distinct spacings, 1.5D, 2D, and 2.5D (where "D" is the tunnel's diameter), are evaluated for twin tunnel samples. The tunnel's cover depth is kept between 3 cm and 5 cm below the surface of the model. The tunnel's diameter is held constant at 5 cm. When it comes to twin tunnel models, PVC pipe is once again used as a liner material for lined tunnels.

1.3 Casting of Twin Tunnel Models

The plaster of paris used to create the twin tunnel samples. As indicated in Fig. 1, 18 of the approximately 36 tunnel samples cast for twin tunnel samples are unlined tunnel samples whereas 18 lined twin tunnel models are casted. After the casting of tunnel sample, they are left undisturbed for 28 days under air curing conditions before being evaluated under static loading circumstances. Fig.1.shows the Twin Tunnel Samples having different c/c spacing and cover depth.

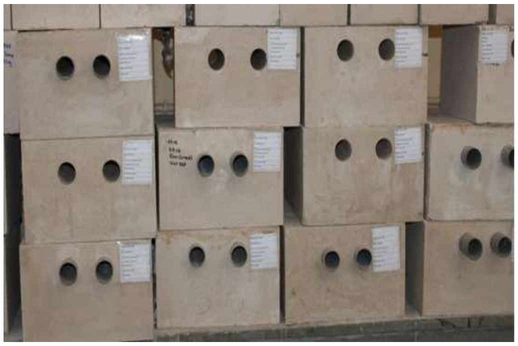


Fig.1. Twin Tunnel Samples having different c/c spacing and cover depth.

1.4 Physical modelling of Twin tunnel

Physical modelling technique is very useful to imitate field circumstances in the lab because it is impossible to conduct all the experiments in the field. Because there are many unfavourable circumstances in the field that prevent the ideal testing from being done successfully. Field experiments are challenging to carry out for practical reasons, so physical model testing must be utilised instead.

The sample used in the Twin Tunnel case is made of 100% plaster of paris with 60% water content. The twin tunnel sample remains 42.5x37.5x23 cm in size (LxWxH). Three distinct spacings, 1.5D, 2D, and 2.5D (where "D" is the tunnel's diameter), are investigated for twin tunnel samples. The tunnel's cover depth is kept between 3 cm and 5 cm below the surface of the model. Twin tunnels are prepared as lined and unlined samples in the laboratory. Six LVDTs are positioned in various positions to collect the tunnel sample's deformation. According to the degree of deformation that occurs in the tunnel sample, the placement of LVDTs is chosen.Each tunnel has three different locations for the three LVDTs. L is the length of the tunnel, and the distances between LVDTs are L/3, L/2, and 9L/15. The identical approach used in one tunnel is used to insert the LVDTs. Six 10mm-diameter holes are drilled from the surface's bottom for installation. The LVDT is then secured with the use of a three-pin clamp to keep it tight and prevent movement.

1.5 Result and Discussions

Plaster of Paris material is employed as the model material for Twin tunnel samples. The static loading condition is applied to the 1.5D c/c spaced twin tunnel sample. The maximum crown deformation value for 3 cm unlined tunnels is 0.25 mm, measured at L/2 distance. While at L/3 and 9L/15, the deformation measured was 0.03mm and 0.16mm, respectively. The crown deformation in 5 cm unlined tunnels is 0.20 mm at L/2 distance, 0.02 mm at L/3, and 0.12 mm at 9L/15, respectively. In the case of lined tunnels having 1.5D centre to centre spacing and 3cm cover depth, the crown deformation at L/2 distance is 0.12mm, whereas the deformation noticed at points L/3 and 9L/15 is 0.01mm and 0.07mm.The crown deformation at L/2 distance in lined tunnels with 1.5D center-to-centre spacing and 3cm cover depth is 0.12mm, but the distortion seen at points L/3 and 9L/15 is 0.01mm and 0.07mm. The deformation encountered at L/3, L/2, and 9L/15 in the case of 5cm lined samples is 0.01 mm, 0.10 mm, and 0.05 mm, respectively as shown in Fig 2.

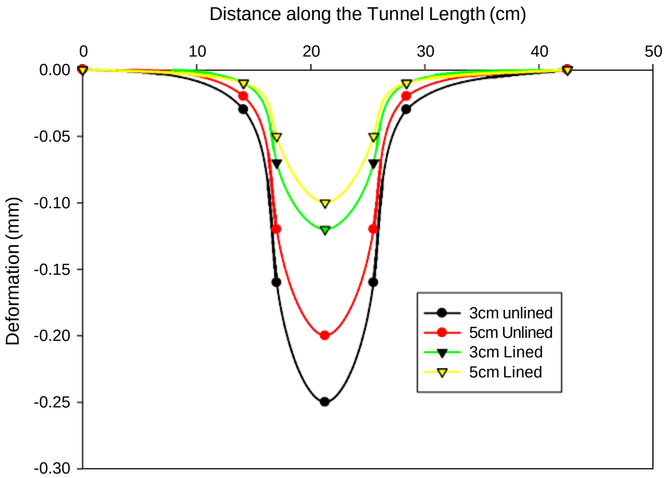


Fig.2. Deformation profiles of 1.5D c/c spacing twin tunnels models obtained from experimental results.

The highest crown deformation value for 2D c/c spacing twin tunnels of 3 cm unlined tunnels is 0.22 mm, obtained at L/2 distance. While at L/3 and 9L/15, the deformation measured was 0.02mm and 0.14mm, respectively. The crown deformation in 5 cm unlined tunnels is 0.18 mm at L/2 distance, 0.02 mm at L/3, and 0.11 mm at 9L/15, respectively. The crown deformation at L/2 distance in lined tunnels with 2Dcentre to centre spacing and 3cm cover depth is 0.10mm, while the distortion seen at L/3 and 9L/15 is 0.01mm and 0.06mm, respectively. The deformation encountered at L/3, L/2, and 9L/15 for 5 cm lined samples is 0.01 mm, 0.07 mm, and 0.04 mm, respectively as shown in Fig 3.

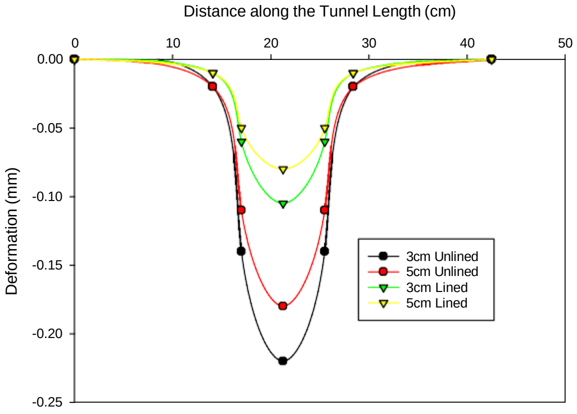


Fig.3. Deformation profiles of 2D c/c spacing twin tunnels models obtained from experimental results.The highest crown deformation value for 2.5D c/c spacing twin tunnels of 3cm unlined tunnels is 0.19mm, obtained at L/2 distance. While at L/3 and

9L/15, the deformations are 0.02mm and 0.10mm, respectively. In 5 cm unlined tunnels, the crown deformation at L/2 is 0.16 mm, whereas the deformation at L/3 and 9L/15 is 0.02 mm and 0.09 mm, respectively. The crown deformation in lined tunnels with 2.5D center-to-centre spacing and 3cm cover depth is 0.09mm at L/2 distance, whereas it is 0.01mm and 0.04mm at L/3 and 9L/15, respectively. The deformation encountered at L/3, L/2, and 9L/15 in the case of 5cm lined samples is 0.01mm, 0.08mm, and 0.03mm, respectively as shown in Fig 4.

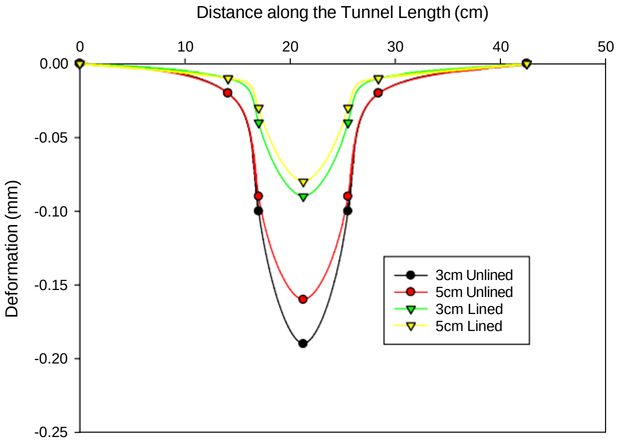


Fig.4. Deformation profiles of 2.5D c/c spacing twin tunnels models obtained from experimental results.

1.6 Conclusions

A comparative study is carried out in this study on the deformation behaviour of twin tunnels with the help of experimental investigation. Various unlined and lined twin tunnel models are prepared in the laboratory with varying strength properties, cover depth and spacing between the tunnel. The following conclusion can be made for the present study.

- The maximum deformation is recorded at center of the tunnel i.e at L/2 distance in all the cases and minimum at L/3.
- The extent of deformation in tunnels mainly depends upon the presence of liner material. Less deformation is observed in lined tunnels as compared to unlined tunnels.
- With increase in the spacing between the twin tunnel,the value of deformation decreases.

References

1. Bayoumi, A., Abdallah, M., and Chehade, F. H.:Non-Linear Numerical Modelling of the Interaction of Twin Tunnels-Structures. *International Journal of Computer and Systems Engineering*, 10(8), 1059–1063.(2016).
2. Chehade, H. F., and Shahrour, I.:Numerical analysis of the interaction between twin

tunnels: Influence of the relative position and construction procedure.*Tunnelling and Underground Space Technology*, 23, 210–214.(2008).

3. Elshamy, E. A., Attia, G., Fawzy, H., and Hafez, K. A.:Behaviour of Different Shapes of Twin Tunnels in Soft Clay Soil. *International Journal of Engineering and Innovative Technology (IJEIT)*, 2(7), 297–302.(2013).
4. Kumar,P. and Shrivastava,A.K.: Various Factors Effecting Stability of Underground Structure: State of Art.*7th Indian Rock Conference 25-27 October 2017,576-580.(2017).*
5. Kumar,P. and Shrivastava,A.K.: Deformation Behaviour of Tunnels under Different Loading Conditions: State of Art. *Proceedings of the 4th World Congress on Civil, Structural, and Environmental Engineering (CSEE'19) Rome, Italy-April, (2019).*
6. Kumar, P. and Shrivastava,A.K. :Physical Investigation of Deformation Behaviour of Single and Twin Tunnel under Static Loading Condition. *Applied Science,* Vol-11,11506,1-18.(2021).
7. Kumar,P. and Shrivastava,A.K.: Experimental and numerical analysis of deformation behaviour of tunnels under static loading conditions. *Sustainable Energy Technologies and Assessments*, Vol-52, 102057, 1-10.(2022).
8. Kumar,P. and Shrivastava,A.K.: Development of a testing facility to determine the Stress-Deformation behaviour of tunnel. *Journal of environmental protection and Ecology, 23(3),946-956.(2021).*
9. Mishra, S., Rao, K. S., Gupta, N. K., and Kumar, A. :Damage to shallow tunnels in different geomaterials under static and dynamic loading.*Thin-Walled Structures*, 126, 138–149.(2018).
10. Mishra, S., Kumar, A., Rao, K.S., and Gupta, N.K. (2021).Experimental and numerical investigation of the dynamic response of tunnel in soft rocks. *Structures*, 29, 2162-2173.(2021)
11. Oliaei, M., and Manafi, E.:Static analysis of interaction between twin-tunnels using Discrete Element Method (DEM).*Scientia Iranica*, 22(6), 1964–1971.(2015).
12. Shrivastava, A. K., and Rao, K. S.:Shear behaviour of non planar rock joints.*14th Asian Regional Conference on Soil Mechanics and Geotechnical Engineering*.(2011).
13. Shrivastava, A. K.: Physical and numerical modelling of shear behaviour of jointed rocks under CNL and CNS boundary conditions. *Ph.D Thesis*, IIT Delhi, (2012).
14. Shrivastava, A. K., and Rao, K. S.:Shear Behaviour of Rock Joints Under CNL and CNS Boundary Conditions.*Geotechnical and Geological Engineering*, 33, 1205–1220,(2015)
15. Shrivastava, A. K., and Rao, K. S.:Physical Modeling of Shear Behavior of Infilled Rock Joints Under CNL and CNS Boundary Conditions.*Rock Mechanics and Rock Engineering*, 51(3), 101–118,(2018).
16. Kumar,P. Modelling of Tunnel Behaviour under Static Load. *Ph.D Thesis*, Delhi Technological University, Delhi, (2022).
17. Singh, R., Singh, T. N., and Bajpai, R. K. (2018).The Investigation of Twin Tunnel Stability: Effect of Spacing and Diameter.*Journal of the Geological Society of India*, 91, 563–568.(2018).
18. Yingjie, L., Dingli, Z., Qian, F., Qingchun, Y., and Lu, X.:A physical numerical investigation of the failure mechanism of weak rocks surrounding tunnels.*Computers and Geotechnics*, 61, 292–307.(2014).

J Opt
https://doi.org/10.1007/s12596-023-01156-3

RESEARCH ARTICLE

Dehazing optically haze images with AlexNet-FNN

Anil Singh Parihar[1] · Sulaxna Gupta[2]

Received: 29 December 2022 / Accepted: 28 February 2023

Abstract Since the beginning of Computer Vision, the resolution of image de-hazing has been a problem. Due to the presence of numerous air particles, resulting in haze, fog, and so on, photographs obtained under unfavorable weather circumstances often seem to be of low quality. This, in turn, makes recognizing objects in a picture difficult. This poses issues for many computer vision problems that depend on picture visibility. The image captured under haze as well as other weather conditions has a process of image deterioration. Image dehazing is a difficult as well as ill-posed task. It overcomes the difficulties of manually constructing haze-related characteristics by using deep learning algorithms. We develop a neural network for image de-hazing in this study. The network model consists of two phases: first, the network is given a foggy image and is tasked with estimating the transmission map; next, the network is given the transmission map estimate and the ratio of the foggy image to the transmission map, and is used to perform haze removal. It avoids estimating ambient light as well as enhances dehazing performance. The haze and dehaze datasets are used as the training set for the proposed scheme. The experimental outcomes for the full-reference metrics SSIM, PSNR, RMSE, MSE, or BRISQUE validate the suggested method's reliability and effectiveness.

Keywords Haze Image dehazing · Deep learning · Neural network · AlexNet model

Introduction

Images having a high level of visibility are essential for tasks using computer vision. On the other hand [1], the quality of photographs that are taken on the hazy days tends to deteriorate because of the absorption of light by floating particles that are present in the surroundings. It is vital to create an efficient dehazing algorithm in order to accomplish the goal of restoring color and features of pictures that have been distorted [2].

Image de-hazing [3] is one of the primary obstacles to progress in computer vision research. Image dehazing remains difficult to achieve despite technological advancements. Both the field of computer vision and everyday life can benefit from solving the problem of image dehazing. One such use is in the removal of haze from photographs. It is possible to find its applications in many different facets of day-to-day living. The issue is a component of a larger group of issues in image processing that pertain to the procedure of de-noising images. Before it reaches the camera, the light that has been reflected from an object will be dispersed by the atmosphere. The abundance of aerosol particles in the atmosphere is responsible for the phenomena of light rays being scattered as they travel through the atmosphere. In turn, this phenomenon has an effect on the way in which a picture is caught by a camera. The quality of the picture is impacted when there are elements such as dust, fumes, fog particles, etc. are present. The lack of vividness and detail in these photographs is due to the circumstances in which they were shot. When used as a reliable source in areas like transportation or surveillance [4], photos like these that

✉ Sulaxna Gupta
sulaxna24296@gmail.com

Anil Singh Parihar
parihar.anil@gmail.com

[1] Department of Computer Engineering, Delhi Technological University, New Delhi, India

[2] Department of Software Engineering, Delhi Technological University, New Delhi, India

Published online: 14 March 2023

are lacking in detail present a risk that might have serious consequences. As a result, the need of picture dehazing has become more vital [5].

The formation of haze [6] may be attributed to either the scattering or the absorption of light that occurs as a result of droplets of water that are floating in the air or to a huge number of very small particles [7]. Images taken in hazy conditions have limited color fidelity and contrast, which is a problem for many optical imaging systems such as satellite remote sensing, aerial photography, outdoor monitoring, and target identification. It introduces a lot of difficulties that must be solved in order to complete the research.

Image processing-based improvements and physical model-based restoration are the two primary types of image processing technologies now accessible for hazy image processing. The more recent of the two is augmentation based on image processing. The technique of improving a picture that relies on image processing begins with the image itself, and it does not take into account the particular reason for the image's deterioration. By increasing the contrast and brightness of the image, the visual impact of the picture can be improved to meet the goal of clarity. These methods are typically mature and effective, and the outputs of clarity can occasionally meet the criterion for clarity. However, such systems are not capable of adapting to a variety of pictures and scenarios. In particular, the picture that has a greater variety of scene depth transitions is ineffective. Because the approach is predicated on picture enhancement and does not take into account the process of fog quality reduction, it is unable to significantly increase image definition. This is the most crucial aspect of the method. It is unable to clear away the fog to restore the original look, and as a consequence, the resulting distortion is even more severe. Not only does the treated picture have a disappointing visual appearance, but it also does not lend itself well to further processing [8].

The remainder of the study is summarized below: Section 2 discusses previous research that is pertinent to the topic of this research. Research methods will be covered in Section 3, while the experiment's results and in-depth analysis will be covered in Section 4. This part also includes the outcomes of the experiment, and Section 5, the study's last section, emphasizes the relevance of the experiment and identifies opportunities for further investigation.

Literature review

The following section is a review of the literature on image dehazing. This section offers information on earlier research work that is related to the present study. According to the findings of the provided literature review, it is akin to setting a precedent among the already accessible approaches. Numerous research has been conducted using image dehazing technology and techniques.

Yin et al. [9] provide an image dehazing approach based on a color-transfer image dehazing concept that outperforms modern techniques. This may be accomplished by employing a Deep CNN-based deep framework to develop an image-dehazing model that uses color-transfer image dehazing to clear away the haze and learn about the model's coefficient. The suggested technique outperforms currently available single-picture dehazing approaches, as shown by quantitative and qualitative assessments of synthetic and hazy images.

Golts et al. [10] explain an unsupervised training technique that involves decreasing the well-known energy function of the Dark Channel Before (DCP). We only utilize real-world outside photos to improve network performance by directly minimizing the difference between the best and worst-case variables, rather than providing the network with bogus data. The utilization of the network and the learning process has resulted in extra regularization, as indicated by this. Experiments show that the performance of our method is comparable to that of large-scale supervised algorithms.

Min et al. [11] provide a method for rating dehazing algorithms that takes into account picture structure recovery, colour rendition, and contrast enhancement in low-light areas. Both types of images can benefit from the proposed method; however, they have made it more suitable for aerial photographs by taking into consideration the particular qualities of these. The recommended approaches have been shown to be successful based on the results of experiments conducted on two different subsets of the SHRQ database.

Huang et al. [12] create a new model that results in the removal of the need for a haze/depth data set by using unsupervised learning and a cycle generative adversarial network. Although evaluated on both synthetic as well as actual haze photos, descriptive and analytical testing indicated that the proposed method outperformed existing state-of-the-art dehazing algorithms. This was the case regardless of whether the haze was actual or synthetic.

Du and Li [13] suggested that the dehazed picture be fed back into the input of the Deep Residue Learning (DRL) network in a recursive manner. An interpretation of this recursive extension as a nonlinear optimization of DRL, the convergence of which can be logically evaluated by applying fixed-point theory, is one possible interpretation. Extensive experimental research has been carried out by our team on both simulated and actual data derived from hazy environments. The efficacy of the suggested recursive DRL approach has been shown by the results of our experiments, and it has been demonstrated that the algorithm gives better than other competing approaches.

Li et al. [8] researchers have developed a dehazing method that is based on residual-based Deep CNNs as part of this body of work. After first providing the network model with a foggy picture, which it uses to derive an estimate of the transmission map based on this image, the network then receives a ratio of the foggy image to the transmission map, which causes the haze to be removed from the picture. Increases the efficiency of dehazing while also eliminating the need to estimate light levels throughout the environment. A training set based on the NYU2 depth datasets has been incorporated into the suggested method. The exploratory results indicate that the proposed method is effective and trustworthy in terms of full-reference metrics peak ratio of signal to noise and correlation, in addition to feature similarity and the non-reference metrics SSIM, PSNR, RMSE, and MSE. Additionally, the results show that the proposed method is good in terms of feature similarity.

Research methodology

Describe the research methodology that was employed for this study effort in the third subsection of this part. This section outlines the entire process, which includes the various steps, tools, and workflow.

Proposed methodology

The most challenging inverse problem is frequently ranked as image dehazing. Deep learning methods have appeared as an addition to traditional model-based techniques, helping to define a fresh state-of-the-art in regards to the level of dehazed pictures that can be obtained. used its deep learning model in this study to solve the aforementioned issue. To begin this study, use the dataset that was gathered. Gather the haze and dehaze datasets first. The collection consists of 55 comparisons between haze-free and hazy images. This dataset is split into both testing and training halves in a 90:10 ratio. 30 s for training. Apply the next preprocessing method, which normalizes images, converts BRG images to RGB images, and converts images to NumPy arrays. Following this, carry out an EDA that displays histogram plots and implements AlexNet using a functional neural network that makes use of the Adam optimizer and a variety of activation functions. Because this takes a while, we have set the number of epochs to five and the batch size to eight. The experimental results verify the efficacy and robustness of the suggested method, which is then calculated using a performance evaluation matrix consisting of SSIM, PSNR, RMSE, MSE, as well as BRISQUE. Below is a brief description of each process.

Data collection

The collection of data. Assemble the datasets for haze and dehaze first. There are 55 comparisons between images with and without haze in the collection. There are training and testing versions of this dataset.

Image pre-processing

Data pre-processing serves as a common and useful technique in the deep learning process. This is because it has the potential to both expand the original database's size and enhance the data that is hidden within the dataset. As a result, the efficiency of the way the subsequent procedures has been carried out is significantly influenced by how well the pre-processing was done. Image processing's main objective is to improve the picture data by eliminating distorted noise and enhancing image pixels. Numerous techniques are used to achieve this. In this project, we gather the unprocessed dehaze images and convert them to RGB. The next step is to normalize the images, which modifies the pixel's range of intensity. Next, create a NumPy vector with three images, each with a unique height, width, and color channel. Before merging the channels of the image, the next step is to make all of them the same.

Proposed model (AlexNet with functional neural network)

Applying a neural network to data [14, 15] a collection of methods that mimic the accuracy and processing speed of the brain in an effort to uncover hidden patterns. "Neural networks" are any systems, whether artificial or real, that are made up of neurons. Since neural networks are adaptable, they still can deliver superior outcomes even when the output requirements are essentially unchanged. More and more often, when creating new trading systems, neural networks, an idea derived from AI. In order to successfully classify pictures using ImageNet, AlexNet is the first significant neural network with a convolutional architecture. Only the older models which weren't deep learning-based were capable of outperforming AlexNet, which was joined in the competition.

Convolutional layers are followed by normalization layers, pooling layers, convolutional-pool-norm layers, a few additional convolutional layers, a max-pooling layer, and finally a number of fully connected layers in many ways resembles the LeNet network. In general, there really are simply more layers. The final fully connected layer, which connects to an output class, comes before These convolutional layers have five actual layers, two of which are fully connected.

AlexNet is a very reliable model which can deliver high levels of accuracy—even when applied to datasets that are exceedingly difficult. The performance of AlexNet would suffer significantly if any one of the convolution layers was removed. An established object-detection architecture with great potential for computer vision tasks is AlexNet. In the near future, it's possible that CNNs [16] will be replaced by AlexNet as the go-to source for image jobs.

AlexNet architecture AlexNet is a straightforward CNN architecture that performs well. As a part of the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC-2012), Alex Krizhevsky et al. made the initial suggestion [17]. Stages built on top of one another make up the majority of it. Convolution, pooling, rectified linear unit (ReLU), and fully connected layers are some of these stages. The first, second, third, and fourth layers of AlexNet are convolutional layers. Following the fifth and pooling layers, there are 3 fully connected layers. AlexNet is a fundamental, straightforward, and successful CNN architecture that's been initially proposed by Alex Krizhevsky et al. in the ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC-2012) [17]. The majority of its components are layered on top of one another. These steps are the pooling layer, the rectified linear unit (ReLU), the fully connected layer, and the convolution layer. Alex Net's first, second, third, and fourth convolutional layers are all present. The following two fully connected layers are the pooling layer and the fifth layer. Equation (1) illustrates how the ReLU, a form of half-wave rectifier, can be utilized to accelerate training and reduce over fitting. When paired with the fully connected layers of the AlexNet design, the dropout approach can be viewed as a sort of regularisation.

$$f(x) = \max(x, 0) \tag{1}$$

Figure 1 depicts the pre-trained AlexNet network model.

Data splitting

The data have been converted into a 90:10 ratio. 90% of the time is spent on teaching, with 10% going towards assessment. Overfitting can be avoided by splitting data using a machine learning method (ML). Overfitting is the process by which machine learning happens to fit the training data so well that it is unable to reliably fit any new data. That category includes this situation. Before entering this initial data into an ML model, it is frequently split into three to four different subgroups. Common examples of datasets are the testing and training datasets.

Proposed algorithm

Input: Haze and Dehaze Dataset

Output: Predicted Results

Step1—Dataset gathering and information

The gathering of information. Create the sets of data for haze as well as dehaze first. In the collection, there really are 55 comparisons among pictures with and without haze.

Step2—Data preprocessing

This preprocessing of the data from BRG to RGB lowers the contrast of the images. Creating a NumPy vector from just a single image, where each element has a height, width,

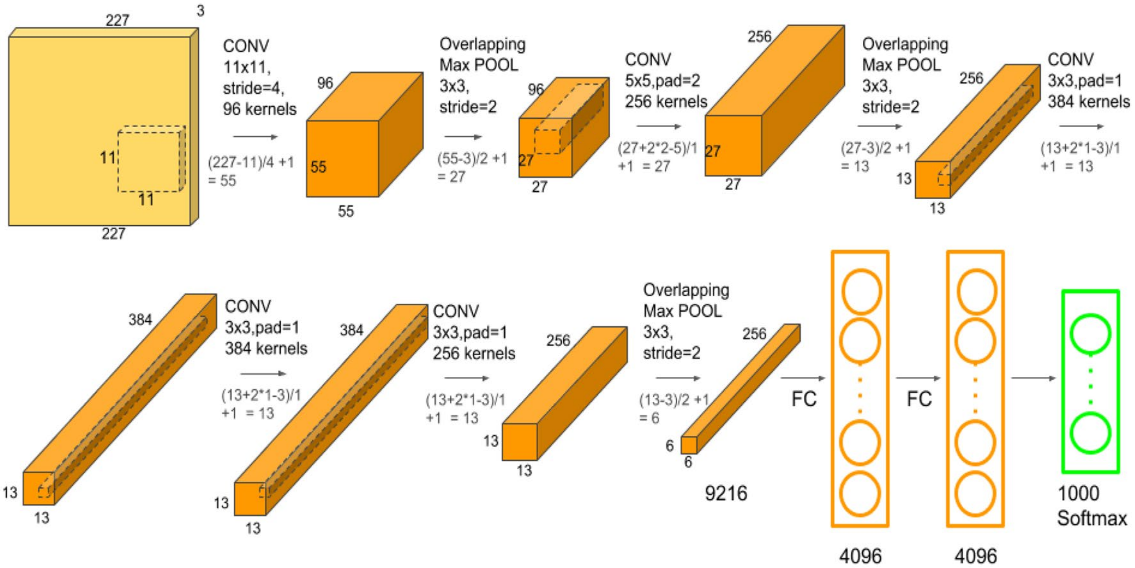


Fig. 1 The AlexNet architecture

and color channel. Combining the image after each channel has been adjusted into a single unit.

Step3—Exploratory data analysis (EDA)

Histogram maps of a predicted image and the raw image are plotted to show the differences. Likewise for data visualization.

Step4—Neural network model to dehaze images

For ground truth and dehazed images, prepare and test samples. 90% of the data are for training, and 10% are for testing. Parameters of a neural network. Used activation function (RELU, Sigmoid). Hyper-Training Conditions Functional neural network built on AlexNet that generates images.

Step5—Performance evaluation metrics

SSIM
PSNR
RMSE
MSE
BRISQUE

Step6—Predicted outcome

Proposed flowchart

The process flow of our work is shown in Fig. 2, below. Upon closer inspection, a structure can be seen inside the picture; this structure is made up of fundamental steps, and within each fundamental step is a sub-step. The study project's flowchart. The graph shows the steps as data collection, preprocessing, information splitting during the testing and training phases, implementing the suggested deep-learning model, and calculating the proposed model's performance evaluation.

Figure 2 above is a diagram of the study project's suggested flowchart. The flow of events is shown in the graph as starting with data collection, then pre—processing, data splitting during testing and training, application of the suggested deep-learning model, as well as calculation of the recommended model's performance evaluation.

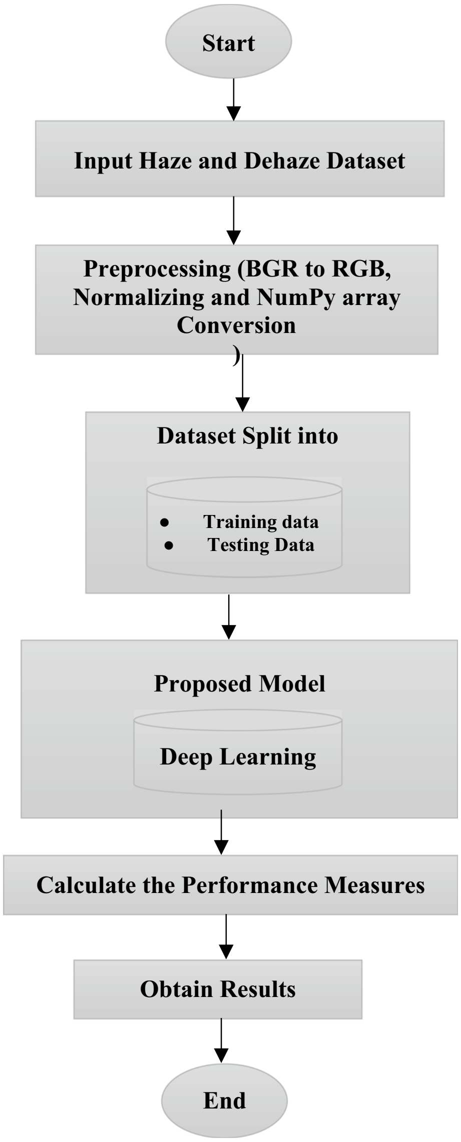

Start
Input Haze and Dehaze Dataset
Preprocessing (BGR to RGB, Normalizing and NumPy array Conversion
)
Dataset Split into
Training data
Testing Data
Proposed Model
Deep Learning
Calculate the Performance Measures
Obtain Results
End

Fig. 2 Proposed flowchart

Results and discussion

In this part, the specifics of the implementation are followed by the outcomes of the model are described. This part discusses the dataset that is used for image dehazing, and its visualization, and brings attention to the analysis of experiments that is included in the current study effort. During this research, the offered methods were applied using Python 3.0, and the dataset used was called "dehaze". In order to put the suggested idea into action, the computer language Python was used. Procedures for evaluation are carried out one after the other in order to verify that the selection of training and test datasets is completely at random. It has been determined that a selection rate of 90% of the data will be used for the training phase, and a selection rate of 10% will be used for the testing phase. In order to illustrate how well the recommended procedures worked, a number of different assessment markers were used. Several performance measures are used to figure out how well something worked.

Exploratory data analysis (EDA)

Expert data analysis (EDA) is a method that involves looking at multiple datasets to figure out how the data is organized. Usually, when people talk about EDA, they mean a way of thinking and a set of tools for adaptable data analysis that doesn't presuppose anything about how the data was originally created. There is a continuous increase in both the volume and the level of complexity of the data that are created by enterprises. EDA is a strategy for doing statistical data analysis.

Plotting some histogram maps of the raw image and the predicted image to clarify the difference.

Figure 3 shows the average columns and rows of every pixel of the haze and dehaze image. In the figure, the *x*-axis shows the rows and the *y*-axis shows the columns. Each picture is comprised of a grid of pixels, and each grid has its own width and height. The number of columns determines the width, while the number of rows determines the height.

Figure 4 shows the frequency of pixels of haze and dehaze images. Graph (a) and (b) shows the haze and dehaze image frequency. The frequency of the image shows on the *x*-axis and the range shows on the *y*-axis. The frequency range is 0–250. The numbers that are closer to zero indicate shades that are deeper, while the numbers that are closer to 255 describe shades that are lighter or whiter.

In Fig. 5 shows the color Intensity in haze and dehaze images. The graph (a) shows the intensity of color of the haze image and graph (b) shows the intensity of color of the dehaze images. The graph *x*-axis and the *y*-axis shows the range and frequencies of both types of data. The graph shows the RGB color performance.

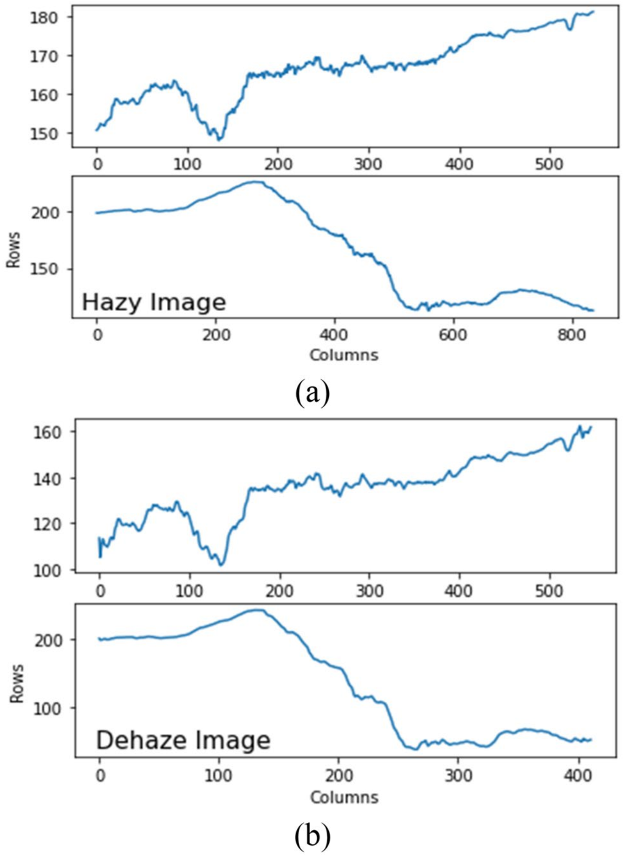


Fig. 3 Average columns and rows of every pixels of haze image and dehaze image

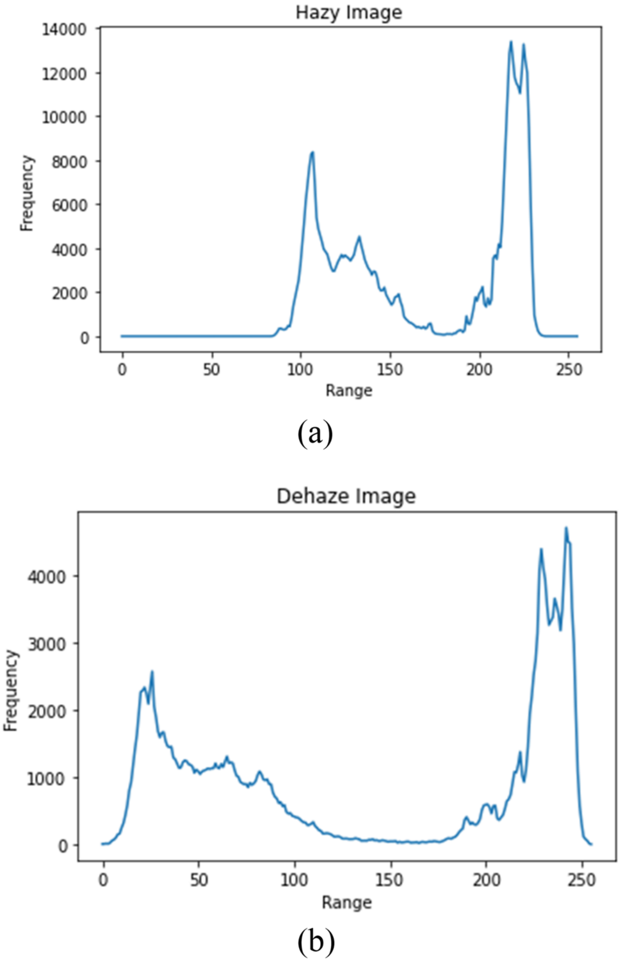


Fig. 4 Frequency of pixels in range 0–255 of haze image and dehaze Image

Performance evaluation measures

Measuring the performance of the trained DL [18] models require using performance assessment measures. This provides assistance in determining how much higher the DL model can execute on a dataset that it has never seen before. In this part, we provide an introduction to some of the most useful performance assessment measures that may be used in DL [7, 19, 20].

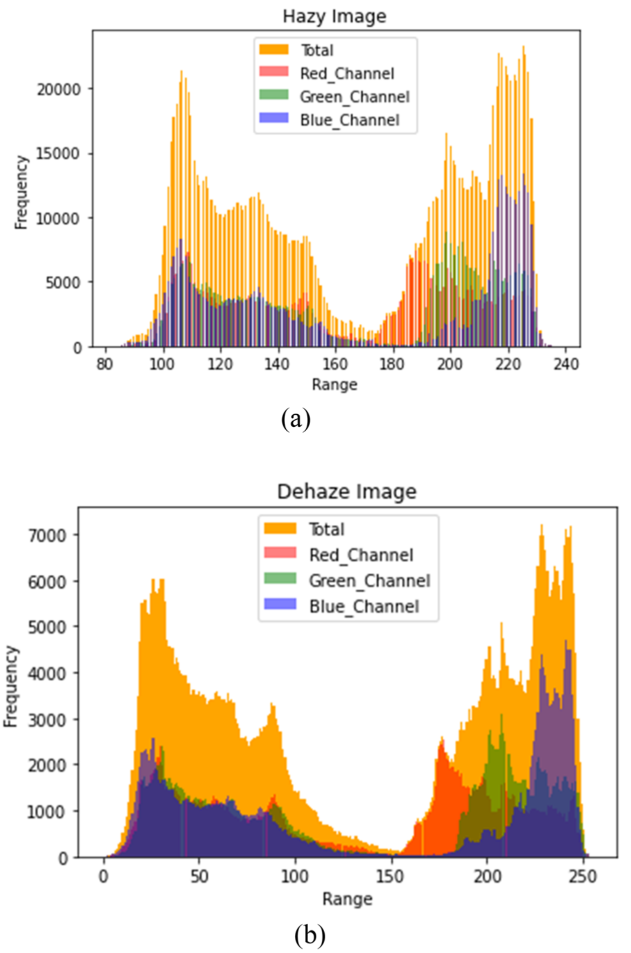


Fig. 5 Intensity of every color channel in haze image and dehaze Image (color figure online)

MSE (mean square error)

The most common way to measure the quality of an image is with the MSE. It is a full reference measure, and the numbers are better the closer they are to zero.

MSE among 2 images for example $g(x, y) and \hat{g}(x, y)$ is definite as:

$$\mathrm{MSE} = \frac{1}{MN} \sum_{n=0}^{M} \sum_{m=1}^{N} [\hat{g}(n, m) - g(n, m)]^2 \tag{2}$$

From Eq. (2), we can see that MSE is a representation of absolute error.

RMSE (root mean square error)

The root-mean-squared error (RMSE) is another type of error assessment approach commonly used to evaluate the gaps between an estimator's prediction and the actual result.

Table 1 Model performance between base and proposed model

Results	RMSE	SSIM	PSNR	BRISQUE	MSE
Base	–	0.90	27.81	22.32	–
Propose	0.012	0.99	66.5	15.42	3.21

This method of error analysis is similar to the concept of root-mean-square error. The error's significance is evaluated. It is the gold standard for measuring the precision with which different estimators forecast a given variable. It's the gold standard of precision, if you will.

Consider an estimator with respect to a specific estimated parameter, whereby the RMSE is defined as the square root of the MSE:

$$\mathrm{RMSE}(\hat{\theta}) = \sqrt{\mathrm{MSE}(\hat{\theta})} \tag{3}$$

PSNR (peak signal to noise ratio)

To determine the quality of a signal's representations, the PSNR is used to compute the ratio among the highest potential signal power as well as the power of the distorting noise. When comparing two photographs, the decibel ratio is used to calculate the difference between the two. The logarithm term of the decibel scale is often used to compute the PSNR because of the vast dynamic range of the signals being measured. Between the greatest and the lowest conceivable values, this dynamic range may be changed by their quality. In terms of PSNR:

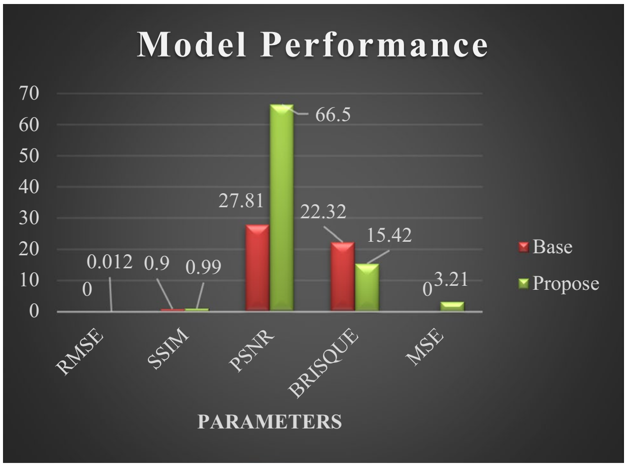
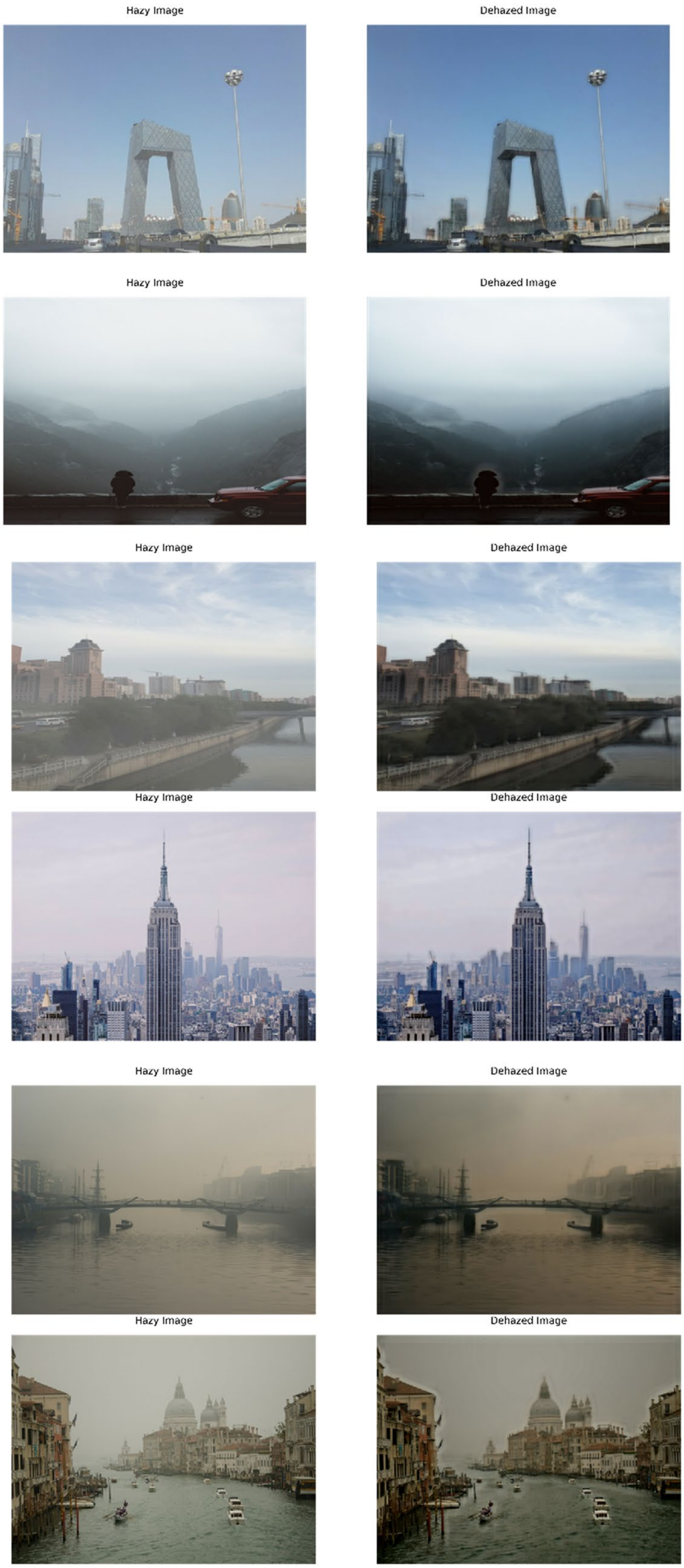


Fig. 6 Comparison graph of base and proposed model performance

Fig. 7 Output images of before and after haze and dehaze



$$\mathrm{PSNR} = 10\log_{10}(\mathrm{peakval}^2)/\mathrm{MSE} \tag{4}$$

Structure similarity index method (SSIM)

"SSIM is a technique that relies on people's subjective perceptions of similarity. Images are thought to be degraded when their structural information is altered. Other key perception-based facts such as luminance masking or contrast masking are also involved in this process. The phrase "structural information" refers to pixels that have a high degree of interdependence or are located in close proximity to one other". These intricately intertwined pixels point to more details about the visual items in the picture. It's called luminance masking when the distortion is reduced at the image's edges. Contrast masking, on the other hand, reduces the visibility of texture distortions in a picture. Image and video quality are assessed using SSIM. It compares two images: the original plus the one that was recovered.

Blind/reference less image spatial quality evaluator (BRISQUE)

"BRISQUE fits the mean subtracted contrast normalized (MSCN) coefficients plus their neighborhood coefficients using the generalized gaussian distribution (GGD) and the asymmetric generalized gaussian distribution (AGGD) models. The quality of a product is evaluated using these model parameters".

From the Table 1 and Fig. 6, shows the performance of base and proposed model, we can see in figure and table proposed model get RMSE is 0.012. SSIM is 0.99, PSNR is 66.5, BRISQUE is 15.22 and MSE is 3.21, respectively. While base SSIM PSNR and BRISQUE are 0.99, 27.81 and 22.32, respectively. The proposed model gets higher performance in comparison to existing model.

The above Fig. 6 shows the after and before haze and dehaze image of the predicated results. Image dehazing's primary goal is to make hazy pictures more clearly visible. The left side images of haze and right-side image of dehaze shows in above figure. First, a hazy picture is fed into the network model, which estimates the transmission map based on this image; next a ratio of foggy image to transmission map is fed into the network, which removes haze from the image. Improves dehazing performance by avoiding the estimate of ambient light (Fig. 7).

Conclusion and future work

The process of visually enhancing the vision that has been deteriorated as a result of atmospheric circumstances is referred to as image dehazing. The primary purpose of picture dehazing is to totally eliminate the haze or fog that is present in the image without causing any deterioration. This method has a wide range of potential applications, including video surveillance, imaging underwater, picture composting, image editing, interactive photomontage, and many more. Deep learning has been found to be an excellent way for picture dehazing in recent studies. In today's world, there has been development in the application of deep learning techniques to the process of picture dehazing. The research presents an image-dehazing technique that makes use of AlexNet in conjunction with a functional NN model. The findings demonstrate that the suggested model not only executes dehazing processing successfully for a variety of scenarios, but that it also does not exhibit any evident color distortion, picture blur, or other such issues. It is more comparable to the expected outcome. On the dataset consisting of both haze and its removal, the performance of the suggested method is assessed. We get good SSIM (0.99), PSNR (66.5), RMSE (0.012), MSE (3.21), and BRISQUE (15.42) scores on sets, and we also demonstrate how our technique produces superior visual results in comparison to previous learning-based approaches. In the not-too-distant future, one of our goals is to improve the structure of the network and find other applications for it. In addition to this, we are going to expand the data collection and make it more accurate. To further boost performance, we also need to raise the intensity of the training received by the network.

Funding Each author confirms that he or she has no vested interest (financial or otherwise) in the contents of this document and has no financial or other ties to any organisations that could.

References

1. S.G. Narasimhan, S.K. Nayar, Contrast restoration of weather degraded images. IEEE Trans. Pattern Anal. Mach. Intell. (2003). https://doi.org/10.1109/TPAMI.2003.1201821
2. J. Gui et al., A comprehensive survey on image dehazing based on deep learning. (2021). https://doi.org/10.24963/ijcai.2021/604
3. T. Guo, V. Monga, Reinforced depth-aware deep learning for single image dehazing. (2020). https://doi.org/10.1109/ICASSP40776.2020.9054504
4. Y.H. Lai, Y.L. Chen, C.J. Chiou, C.T. Hsu, Single-image dehazing via optimal transmission map under scene priors. IEEE Trans. Circuits Syst. Video Technol. (2015). https://doi.org/10.1109/TCSVT.2014.2329381
5. R.R. Choudhary, K.K. Jisnu, G. Meena, Image DeHazing using deep learning techniques Ravi Raj Choudhary. (2020). https://doi.org/10.1016/j.procs.2020.03.413
6. C.A. Hartanto, L. Rahadianti, Single image dehazing using deep learning. Int. J. Informatics Vis. (2021). https://doi.org/10.30630/joiv.5.1.431

7. Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. (2004). https://doi.org/10.1109/TIP.2003.819861
8. J. Li, G. Li, H. Fan, Image Dehazing using residual-based deep CNN. IEEE Access (2018). https://doi.org/10.1109/ACCESS.2018.2833888
9. J.L. Yin, Y.C. Huang, B.H. Chen, S.Z. Ye, Color transferred convolutional neural networks for image dehazing. IEEE Trans. Circuits Syst. Video Technol. (2020). https://doi.org/10.1109/TCSVT.2019.2917315
10. A. Golts, D. Freedman, M. Elad, Unsupervised single image dehazing using dark channel prior loss. IEEE Trans. Image Process. (2020). https://doi.org/10.1109/TIP.2019.2952032
11. X. Min et al., Quality evaluation of image dehazing methods using synthetic hazy images. IEEE Trans. Multimed. (2019). https://doi.org/10.1109/TMM.2019.2902097
12. L.Y. Huang, J.L. Yin, B.H. Chen, S.Z. Ye, Towards unsupervised single image dehazing with deep learning. (2019). https://doi.org/10.1109/ICIP.2019.8803316
13. Y. Du, X. Li, Recursive deep residual learning for single image dehazing. (2018). https://doi.org/10.1109/CVPRW.2018.00116
14. S. Kollmannsberger, D. D'Angella, M. Jokeit, L. Herrmann, Neural networks, in *Studies in Computational Intelligence* (2021)
15. L. Haripriya, M.A. Jabbar, M. Tech, A survey on neural networks and its applications. Int. J. Eng. Res. Comput. Sci. Eng. (2018)
16. J. Gu et al., Recent advances in convolutional neural networks. Pattern Recognit. (2018). https://doi.org/10.1016/j.patcog.2017.10.013
17. C. Szegedy et al., Going deeper with convolutions. (2015). https://doi.org/10.1109/CVPR.2015.7298594
18. C. Hodges, M. Bennamoun, H. Rahmani, Single image dehazing using deep neural networks. Pattern Recognit. Lett. (2019). https://doi.org/10.1016/j.patrec.2019.08.013
19. B. Sankur, Statistical evaluation of image quality measures. J. Electron. Imaging (2002). https://doi.org/10.1117/1.1455011
20. A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. (2012). https://doi.org/10.1109/TIP.2012.2214050

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

ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Design and Computational Analysis of an MMP9 Inhibitor in Hypoxia-Induced Glioblastoma Multiforme

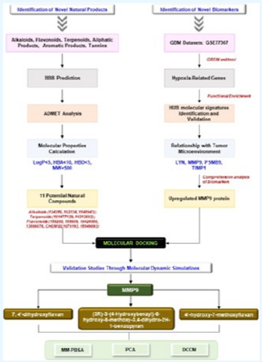
Smita Kumari and Pravir Kumar*

Cite This: https://doi.org/10.1021/acsomega.3c00441

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The main therapeutic difficulties in treating hypoxia-induced glioblastoma multiforme (GBM) are toxicity of current treatments and the resistance brought on by the microenvironment. More effective therapeutic alternatives are urgently needed to reduce tumor lethality. Hence, we screened plant-based natural product panels intending to identify novel drugs without elevating drug resistance. We explored GEO for the hypoxia GBM model and compared hypoxic genes to non-neoplastic brain cells. A total of 2429 differentially expressed genes expressed exclusively in hypoxia were identified. The functional enrichment analysis demonstrated genes associated with GBM, further PPI network was constructed, and biological pathways associated with them were explored. Seven webtools, including GEPIA2.0, TIMER2.0, TCGA-GBM, and GlioVis, were used to validate 32 hub genes discovered using Cytoscape tool in GBM patient samples. Four GBM-specific hypoxic hub genes, LYN, MMP9, PSMB9, and TIMP1, were connected to the tumor microenvironment using TIMER analysis. 11 promising hits demonstrated positive drug-likeness with nontoxic characteristics and successfully crossed blood−brain barrier and ADMET analyses. Top-ranking hits have stable intermolecular interactions with the MMP9 protein according to molecular docking, MD simulation, MM-PBSA, PCA, and DCCM analyses. Herein, we have reported flavonoids, 7,4′-dihydroxyflavan, (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran, and 4′-hydroxy-7-methoxyflavan, to inhibit MMP9, a novel hypoxia gene signature that could serve as a promising predictor in various clinical applications, including GBM diagnosis, prognosis, and targeted therapy.


Identification of Novel Natural Products
Identification of Novel Biomarkers
Alkaloids, Flavonoids, Terpenoids, Aliphatic Products, Aromatic Products, Tannins
GBM Datasets: GSE77307
GREIN webtool
BBB Prediction
Hypoxia Related Genes
Functional Enrichment
ADMET Analysis
HUB molecular signatures Identification and Validation
Molecular Properties Calculation
LogP<5, HBA<10, HBD<5, MW<500
Relationship with Tumor Microenvironment
LYN, MMP9, PSMB9, TIMP1
Comprehensive analysis of biomarkers
11 Potential Natural Compounds
Upregulated MMP9 protein
MOLECULAR DOCKING
Validation Studies Through Molecular Dynamic Simulations
MMP9
7, 4'-dihydroxyflavan
(3R)-3-(4-Hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2H-1-benzopyran
4'-hydroxy-7-methoxyflavan
MM-PBSA
PCA
DCCM

1. INTRODUCTION

According to CBTRUS (Central Brain Tumor Registry of the United States), 2021 recent research, glioblastoma multiforme (GBM) accounts for 48.6% of primary malignant brain tumors. Individuals aged 20−39 years experienced the most significant increases in survival, with 5 year survival increasing from 44 to 73%. In contrast, the failure to enhance survival in older age groups was primarily due to the inability to improve GBM therapy.[1] Currently, GBM is being treated with a combination of surgery, radiation therapy, and chemotherapeutics [alkylating drug temozolomide (TMZ) and antiangiogenic agent bevacizumab]. Furthermore, novel treatments such as tumor-treating fields and immunotherapy offer promise for a better prognosis.[2] Despite these treatment options, GBM patients' overall survival and quality of life remain dismal. The plethora of research mentioned numerous obstacles to GBM treatment, including tumor heterogeneity, acidic microenvironment, and immunosuppression, all of which are linked to the hypoxic environment to some degree.[3]

GBM, being a highly vascularized human tumor, its microcirculation is poor, resulting in the hypoxia region inside the tumor. In the tumor microenvironment (TME), unregulated cell proliferation in the tumor (tumor size exceeds the diameter of >1 mm) often surpasses the capacity of the pre-existing blood capillaries to meet the oxygen demand.[4] This results in a condition known as hypoxia, which impairs the availability of nutrients and promotes genetic instability because of an increase in the generation of reactive oxygen species making it a crucial factor for tumorigenesis. As the master regulator orchestrating cellular responses to hypoxia, hypoxia-inducible factor 1 (HIF-1) plays an essential role in GBM aggressiveness. This modulates the expression of angiogenic factors, such as vascular endothelial growth factor (VEGF), insulin-like growth factor II, and platelet-derived growth factor B (PDGF), and several glucose and fatty acid metabolism factors, the tumor-immune microenvironment, and stimulation of the epithelial−mesenchymal transition (EMT), suppressing apoptosis and promoting autophagy.[5,6] In addition, hypoxia also serves as a niche environment for the aggregation of cancer stem cells, which promotes carcinogenesis and resistance. Tumor cells use a variety of strategies in response to hypoxia, including the expulsion of cytotoxic anticancer drug by ABC-transporters, manifesting a dormant state and exhibiting pluripotency (stemness) traits, which can lead to the failure of existing therapy.[7] Studies showed that hypoxia promotes secretion of cytokines and chemokines which affects immunosurveillance by affecting CD8+ T cell infiltration and disrupting

Received: January 21, 2023
Accepted: February 28, 2023

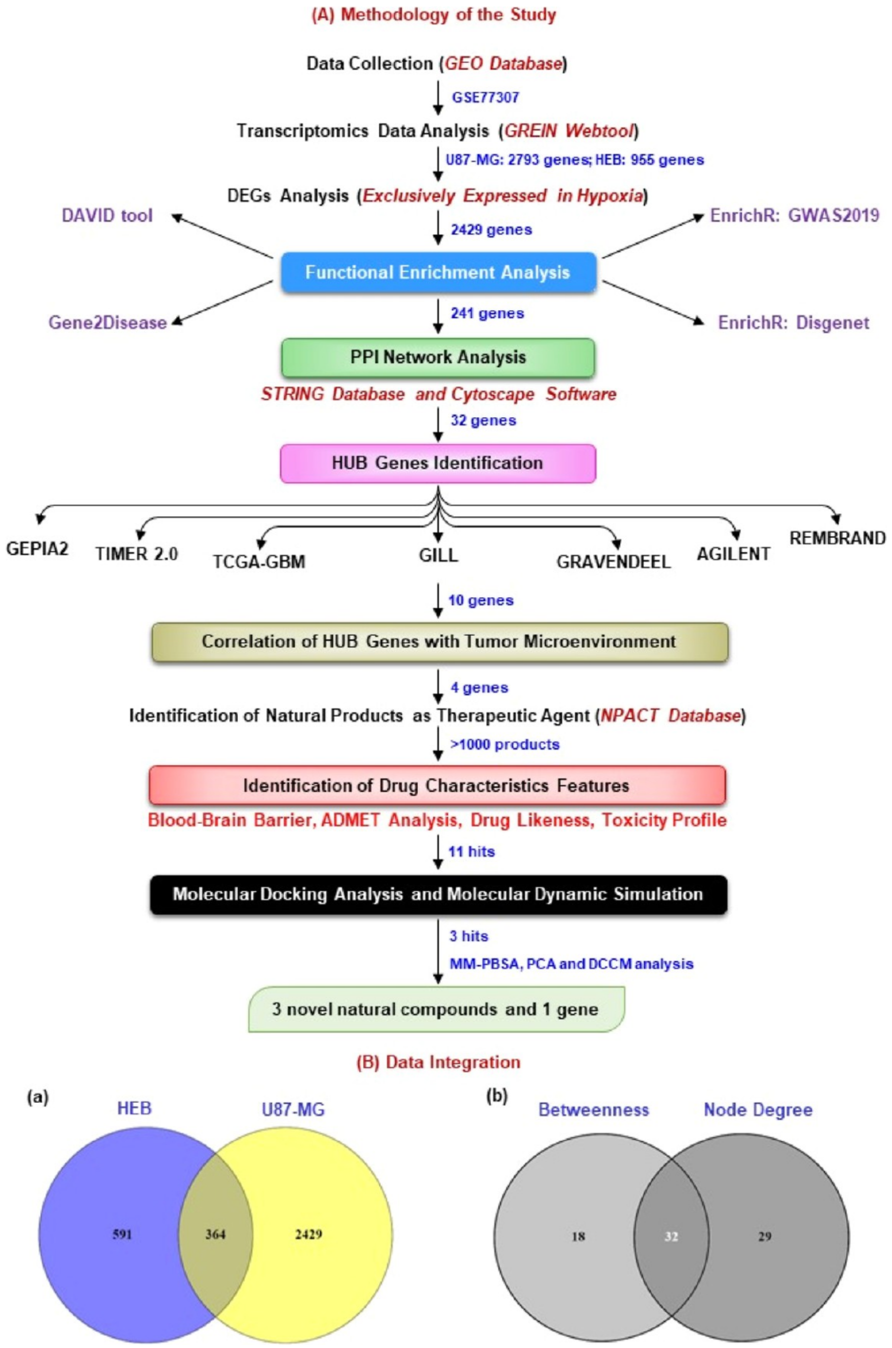

(A) Methodology of the Study
Data Collection (GEO Database)
GSE77307
Transcriptomics Data Analysis (GREIN Webtool)
U87-MG: 2793 genes; HEB: 955 genes
DEGs Analysis (Exclusively Expressed in Hypoxia)
2429 genes
DAVID tool
EnrichR: GWAS2019
Functional Enrichment Analysis
Gene2Disease
EnrichR: Disgenet
241 genes
PPI Network Analysis
STRING Database and Cytoscape Software
32 genes
HUB Genes Identification
GEPIA2
TIMER 2.0
TCGA-GBM
GILL
GRAVENDEEL
AGILENT
REMBRAND
10 genes
Correlation of HUB Genes with Tumor Microenvironment
4 genes
Identification of Natural Products as Therapeutic Agent (NPACT Database)
>1000 products
Identification of Drug Characteristics Features
Blood-Brain Barrier, ADMET Analysis, Drug Likeness, Toxicity Profile
11 hits
Molecular Docking Analysis and Molecular Dynamic Simulation
3 hits
MM-PBSA, PCA and DCCM analysis
3 novel natural compounds and 1 gene
(B) Data Integration
(a)
HEB
U87-MG
591
364
2429
(b)
Betweenness
Node Degree
18
32
29

Figure 1. (A) Workflow scheme for identification of novel natural compounds (target) against GBM-hypoxia microenvironment. (B) Interactive Venn analysis: (a) identification of DEGs in the GBM-hypoxia microenvironment. A total of 2429 altered DEGs exclusively expressed in hypoxia were identified from the GSE77307 data set using the GREIN tool. The "cross areas" are common DEGs in both cell lines. The cutoff criteria were p value ≤ 0.05 and [log fold change] ≥ ±1.5. (b) A total of 32 hub genes among topology parameters (betweenness and degree) were identified from Cytoscape software. The "cross areas" are common hub genes. HEB (purple): non-neoplastic brain cell; U87-MG (yellow): human GBM cell model.

the cytotoxicity of natural killer cells cells. In addition, hypoxic tumor-associated macrophages reduce T cell responses and encourage tumor proliferation and angiogenesis.[8,9] Another essential piece of research emphasizes the role of $\gamma\delta$ T cells as they do not require antigen presentation for activation compared to conventional T cells and are thus an excellent therapeutic target for brain tumors. This pathway is also mediated by hypoxia.[10] So, given hypoxia's critical role in intratumoral interactions, identifying targets that induce adaptation to the hypoxic niche is crucial for a better understanding of GBM origin, development, and treatment resistance.[11] Indeed, "hypoxia" is an essential driving force of GBM and could be used as a novel treatment tool.[12]

Regardless of the fact that there have been few improvements in the progression of GBM therapies to boost patient survival, researchers and clinicians are indeed eager to study novel therapies and techniques for treating this disease.[13] Natural compounds and their structure analogues have been the source of most medicines' active ingredients for various indications, including cancer.[14] Some widely used plant-derived natural compounds are etoposide, irinotecan, paclitaxel, and vincristine, bacteria-derived anti-cancer therapeutics are mitomycin C and actinomycin D, and marine-derived anti-cancer therapeutics is bleomycin.[15] Numerous studies suggest that natural compounds are used as chemosensitizers (such as quercetin, resveratrol, withaferin A, etc.), radiosensitizers (such as tetrandrine, zataria, multiflora, and guduchi), and anti-proliferative (such as curcumin, oridonin, rutin, and cucurbitacin) alkaloids and flavonoid agents.[16,17] Identification of new drugs that can modify the BBB (blood−brain barrier), decrease the tumor growth, and prevent the development of recurring tumors is critical for improving overall patient prognosis. In vitro and/or in vivo, various natural compounds with well-established biological benefits have oncologic effects on GBM.[18] These include flavonoids, terpenoids, alkaloids, tannins, coumarins, curcuminoids, terpenes, lignans, natural steroids, and plant extracts.[19] Statistics show that over 60% of the approved anti-cancer agents are of natural origin (natural compounds or synthetic compounds based on natural product models).

The present study conducted transcriptomic analysis between hypoxia and normoxia (in both normal non-neoplastic brain cells and GBM tumor cells) samples to screen differentially expressed genes (DEGs) related to hypoxia effects. Comprehensive bioinformatics and computational methodologies were used to identify hub genes (LYN, MMP9, PSMB9, and TIMP1) and significant modules and pathways related to the TME. We found that matrix metalloproteinase 9 (MMP9) plays a vital role as a hypoxic gene signature, which has the potential to be used as a biomarker. Numerous studies have also shown the dysregulation of MMP9 in the microenvironment associated with hypoxia and cancer.[20] MMP9 can cleave and remodel extracellular matrix (ECM) proteins such as collagens and elastin involved in invasion, metastasis, and angiogenesis.[21] MMP9 is produced de novo by monocytes and inflammatory macrophages, as well as most cancer cells, during stimulation induced by various extracellular signals present in TME, such as proinflammatory cytokines (such as TNF-α, IL-8, and IL-1β) and growth factors (such as TGF-β, PDGF, and bFGF), which can bind to their receptors and activate downstream signaling cascades involved in the activation of transcription factors including NF-κB, SP1, AP1, and HIF-1α. This affects various downstream biological processes, including matrix degradation, remodeling, EMT, enhanced tumoral invasion, metastasis, angiogenesis, inflammation, drug resistance, and so forth; hence, it acts as a challenging target for targeted therapy for cancer.[22]

Targeting TME has been a significant focus in recent years, and hence MMP inhibitors that will target a hypoxia condition in the microenvironment could be of great significance as a new antitumor agent. For this purpose, we have availed network pharmacology, structure-based drug design approach such as molecular docking, molecular dynamics (MD) simulation analysis, and molecular mechanics Poisson−Boltzmann surface area (MM-PBSA) approach to discover prospective classes of natural compounds with druggable and nontoxic properties from the plant-based natural compounds library. We identified 11 hits based on the particular interaction that satisfy the ADMET and LIPINSKI rule of five analyses, pass the toxicity profile, and have a significant affinity for the MMP9 binding site domain. The three best-docked compounds were further subjected to MDS for 50 ns to understand protein−ligand complex stability. Previously also, researchers have explored the potential of alkaloids and flavonoids for anti-cancer treatments.[23,24] Drugs, including natural compounds that target MMP9, have not been used in the clinical setting. Therefore, targeted MMP9 drugs must be screened for treating patients with GBM. Our results can potentially benefit from managing GBM malignancy caused by a hypoxia microenvironment. The findings of this study contribute to a better understanding of the role of the hypoxia microenvironment. Figure 1A depicts the process of the methodologies used in this investigation.

2. MATERIALS AND METHODS

2.1. Data set Acquisition and Processing. The NCBI-Gene Expression Omnibus (NCBI-GEO; https://www.ncbi.nlm.nih.gov/geo) database[25] is a publicly accessible library of next-generation sequencing, RNA sequencing, and microarray profiling used to gather GBM and non-neoplastic brain tissue gene expression profiles from GEO accession number, GSE77307. The transcriptome data in GSE77307 were derived from GPL11154, a platform using Illumina HiSeq 2000 (*Homo sapiens*). This included three replicates of each U87-MG cell line as a human GBM cancer cell model and the human brain HEB cell line as a non-neoplastic brain cell model cultured in 21% oxygen (normoxia) and 1% oxygen (hypoxia) for transcriptional profiling. This data set was chosen due to the availability of only one data set in the database based on the filter (glioblastoma; hypoxiaTME). High-throughput functional transcriptomic expression data from GSE data sets were analyzed through GEO RNA-seq Experiments Interactive Navigator online server (GREIN; https://shiny.ilincs.org/grein).[26] GREIN is provided by the backend compute pipeline for uniform processing of RNA-seq data and large numbers (>65,000) of processed data sets.

2.2. Enrichment Analysis of Identified DEGs. Transcriptomics data analysis was performed using the GREIN web tool. DEGs were determined by comparing their expression levels in hypoxia (1% oxygen) versus normoxia (21% oxygen) in GBM cells, U87-MG, and normal brain cells, HEB. Statistically significant DEGs were screened using cutoff filter criteria such as unpaired t-test and p-value ≤ 0.05, false discovery rate ≤ 0.05, and [log fold change] ≥ 1.5. DEGs only exclusively expressed in hypoxia conditions were considered for further analysis. In addition, enrichment analysis of DEGs, including both upregulated and downregulated genes associated with GBM, was performed by utilizing different omics approaches such as the Database for Annotation, Visualization and Integrated Discovery (DAVID) functional annotation tool (https://david.ncifcrf.gov/),[27] gene set to diseases (GS2D) tool (http://cbdm.uni-mainz.de/geneset2diseases),[28] and Enrichr-GWAS2019 and Enrichr-DisGeNET of Enrichr tool (https://amp.pharm.mssm.edu/Enrichr)[29,30] to identify and prioritize the most significant genes associated with GBM. Furthermore, the biological pathway and functional enrichment analyses of candidate DEGs and hub genes were determined through a freely available software known as the FunRichr tool (version 3.1.3) (http://www.funrich.org/)[31] to identify the biological pathways associated with them.

2.3. Integration of Protein−Protein Interaction Network and Hub Genes Identification. The selected enriched genes were then examined for designing Protein−Protein Interaction (PPI) using an online Search Tool for the Retrieval of Interacting Genes/Proteins (version 11.5) (STRING, https://string-db.org/) for *H. sapiens*[32] that covers known and predicted interactions for different organisms. The experimentally significant interactions (with high confidence scores ≥ 0.700) were chosen to build a network model, while the others were excluded from the analysis. Cytoscape software (version

https://doi.org/10.1021/acsomega.3c00441

3.8.1) (https://cytoscape.org/)[33] was implemented to analyze the PPI network and identify the hub protein. To calculate the topological parameters such as the node degree (the number of connections to the hub in the PPI network) and betweenness (which corresponds to the centrality index of a particular node), we used the CentiScaPe plugin (version 2.2). It denotes the shortest route between two nodes. Genes with higher values than the average score were chosen.

2.4. Hub Protein Shorting and Validation. To verify and validate the expression of the shortlisted hub proteins, we have utilized both transcriptomics and genomics data from GBM patients. Different databases were explored for RNA sequencing data, such as GEPIA2.0, TIMER2.0, TCGA-GBM, and GlioVis-GILL, and microarray data, such as GlioVis-REMBRANDT, GlioVis-AGILENT, and GlioVis-Gravendeel based on Cancer Genome Atlas (TCGA)-GBM.[34−36] GEPIA2.0 analyzed the RNA sequencing expression data of 9736 cancers and 8587 normal samples from the TCGA and GTEx projects using a standard processing pipeline. GlioVis is a user-friendly web tool that allows users to study brain tumor expression data sets through data visualization and analysis. For GlioVis-GILL, Gill et al. conducted RNA-seq and histological examination on radiographically labeled biopsies collected from different regions of GBM.[37] GlioVis-Repository of Molecular Brain Neoplasia Data (REMBRANDT), a cancer clinical genomics database and a web-based data mining and analysis platform, includes data produced from 874 glioma specimens with approximately 566 gene expression arrays and 834 copy number arrays generated through the Glioma Molecular Diagnostic Initiative.[38] In GlioVis-Gravendeel, gene expression profiling was carried out on a large cohort of glioma samples from all histologic subtypes and grades.[39] In TIMER2.0, multiple immune deconvolution algorithms were used to assess the quantity of immunological infiltrates. Its Gene DE module allows users to investigate the differential expression of any gene of interest in tumors and surrounding normal tissues across all TCGA tumors. All hub genes significantly expressed in all seven patient GBM databases were chosen for subsequent research. Finally, shortlisted genes were again subjected to Tumor IMmune Estimation Resource (TIMER) (https://cistrome.shinyapps.io/timer)[40] analysis. Here, we utilized this database to link hub gene expression with tumor purity and estimate the infiltration levels of six immune cell types [CD4+ T cells, CD8+ T cells, B cells, macrophages, neutrophils, and dendritic cells (DCs)] in GBM data sets. This tool calculates immune infiltration based on immune subsets' preset characteristic gene matrix.

2.5. Localization Study and Construction of Transcription Factor-Gene Network. CELLO (http://cello.life.nctu.edu.tw/cello.html): subcellular localization predictor combines a two-level support vector machine system and the homology search method-based tool to predict the subcellular localization of the protein.[41] Regulatory transcription factors (TFs) that control the expression of genes at the transcriptional level were obtained using the JASPAR database, containing curated and nonredundant experimentally defined TF binding sites.[42] The TF-gene interaction networks were constructed and analyzed with NetworkAnalyst (version3.0) (https://www.networkanalyst.ca/).[43]

2.6. Identification of Natural Compounds and Blood−Brain Permeability Prediction. The plant-derived natural compounds with known anti-cancer bioactivity information were obtained from a literature survey through PubMed and the central resource Naturally Occurring Plant-based Anti-cancer Compound-Activity-Target database (NPACT, http://crdd.osdd.net/raghava/npact/).[44] This database, which presently has 1574 compound entries, collects information on experimentally confirmed plant-derived natural compounds with anti-cancer action (in vitro and in vivo). We have chosen terpenoids (513 entries), flavonoids (329 entries), alkaloids (110 entries), polycyclic aromatic natural compounds (63 entries), aliphatic natural compounds (20 entries), and tannin (6 entries).BBB obstructions make it difficult to create drugs to treat brain cancer. The BBB blocks the uptake of necessary therapeutic drugs into the brain. The epithelial-like tight connections seen in the brain capillary endothelium are the source of this characteristic. For the treatment of GBM, it is crucial to screen drugs that have the ability to cross the BBB.[45] While designing a drug for brain diseases, physicochemical properties and brain permeation properties should be optimized. In consideration of this challenge, we analyzed our candidate natural compounds for physicochemical properties using the SwissADME (http://www.swissadme.ch/)[46] analysis tool and the CBLigand (version 0.90) online BBB predictor (https://www.cbligand.org/BBB/).[47]

2.7. Prediction of Molecular Properties and Drug Toxicity. Each natural compound's molecular formula (MF), molecular weight (MW), hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), log *P* value, and SMILES were retrieved using the PubChem chemical database (https://pubchem.ncbi.nlm.nih.gov/). The Lipinski rule of five was used to estimate the druggability of each phytocompound using the SMILES data of individual compounds on the MolSoft web server (https://molsoft.com/mprop/).[48] The server includes structural data such as MF, MW, HBA, HBD, and logP and a drug-likeness score prediction (DLS). The toxicity and pharmacokinetics of natural compounds with positive DLS were also predicted using the ADMETlab 2.0 (https://admetmesh.scbdd.com/) webserver.[49]

2.8. Molecular Docking Studies. *2.8.1. Preparation of Ligand.* Based on the network analysis and pharmacology approach, 11 natural compounds, viz., 6 flavonoids, 3 alkaloids, and 2 terpenoids, were qualified for all criteria required for being used as a drug candidate. Thus, the three-dimensional (3D) structures of 11 natural compounds along with 2 reference drugs (one natural compound and one conventional standard molecule) were retrieved from the PubChem database (https://pubchem.ncbi.nlm.nih.gov/) in the structure data file (.sdf) format. These structures additionally went through the dock prep section of Discovery Studio Visualizer[50] (BIOVIA Discovery Studio Visualizer; https://discover.3ds.com/discovery-studio-visualizer-download) 2019. The conjugate gradients algorithm was used to minimize the ligand structures using the "uff" forcefield.[51] The polar hydrogens and Gasteiger charges were added to the ligands to convert them into the ".pdbqt" format.

2.8.2. Preparation of Protein. Based on the network analysis and TIMER analysis, the overexpressed MMP9 gene associated with the TME was prioritized for future investigation. The Research Collaboratory for Structural Bioinformatics (RCSB; https://www.rcsb.org/) protein data bank was used to retrieve the X-ray crystallographic structure of MMP9 (PDB: 4HMA). Further, the PrankWeb (https://prankweb.cz/) server based on P2Rank, a machine learning method, was used to retrieve the information on the target active site and binding pockets, and the ligand was docked within the predicted site. Functional characteristics of protein structures were validated using

https://doi.org/10.1021/acsomega.3c00441

Ramachandran plot, ERRAT, and VERIFY3D.[52−54] For a good quality model, the ERRAT quality factor should be greater than 50, and the number of residues having a score ≥ 0.2 in the 3D/1D profile, as predicted by the VERIFY3D server, should be more than 80%.

2.8.3. Protein−Ligand Docking. All ligands were docked against protein using AutoDock vina 4.0 executed through the POAP pipeline.[55] The intermolecular interaction compounds showing the least binding energy and maximum intermolecular interaction with the active site residues were selected to visualize protein−ligand interactions using BIOVIA Discovery Studio Visualizer 2019 and further subjected for MD simulation.

2.9. MD Simulation of Best-Docked Protein−Ligand Complex. In order to infer the stability of docked complexes, we prioritized five complexes (three test and two standard complexes) and subjected to all-atoms explicit MD simulation for 50 ns production run using GROMACS version 2021.3 software package (GNU, General Public License; http://www.gromacs.org).[56] The ligand and protein topology were generated using Amber ff99SB-ildn force field (https://ambermd.org/AmberTools.php) via antechamber x-leap tool. The system was solvated using the TIP3P water model in an orthorhombic box with a boundary condition of 10.0 Å from the edges of the protein in all directions. The system was neutralized by adding necessary amounts of counterions. The conjugate gradient approach was employed to obtain the near-global state least-energy conformations after the steepest descent. Canonical (constant temperature, constant volume, *NVT*) and isobaric (constant temperature, constant pressure, *NPT*) equilibrations were performed on the systems for 1 ns. A modified Berendsen thermostat method was used in *NVT* equilibration to keep both the volume and temperature constant (300 K). Similarly, a Parrinello−Rahman barostat was used during *NPT* equilibration to keep the pressure at 1 bar constant. The particle mesh Ewald approximation was used with a 1 nm cutoff to calculate the long-range electrostatic interactions, van der Waals interactions, and coulomb interactions. In order to control the bond length, the LINCS algorithm (LINear Constraint Solver algorithm) was utilized. The coordinates were recorded every two fs during each complex's production run of 50 ns. In-built GROMACS utilities were used to evaluate the generated trajectories, and other software packages were incorporated where necessary for a more specialized analysis. MD trajectories were analyzed to determine the c-alpha root-mean-square fluctuation (RMSF) and root-mean-square deviation (RMSD) of the backbone and complex, the protein radius of gyration (R_g), the protein solvent-accessible surface area (SASA), and the number of hydrogen bonds between the protein and the ligand.

2.10. Investigation of Binding Affinity Using MM-PBSA. It is standard procedure to use the relative binding energy of a protein−ligand complex in MD simulations and thermodynamic calculations. MM-PBSA was performed by "g_mmpbsa" tool.[57] The total free energy of each of the three entities (ligand, protein receptor, and complex) mentioned can be calculated by adding the potential energy of the molecular mechanics and the energy of solvation. Early research work[58,59] was used to obtain the parameter that was used to determine the binding energy.

$$\Delta G_{(\text{binding})} = G_{(\text{complex})} - G_{(\text{protein})} - G_{(\text{ligand})} \quad (1)$$

where $G_{(\text{complex})}$ is the total free energy of the ligand−protein complex and $G_{(\text{protein})}$ and $G_{(\text{ligand})}$ are the total free energies of the isolated protein and ligand in the solvent, respectively.

The binding energy was calculated over the stable trajectory observed between 50 ns using 50 representative snapshots.

2.11. PCA and DCCM Analyses. Principal component analysis (PCA) was used in the current work to analyze the main types of molecular motions utilizing MD trajectories. It is employed to study the eigenvectors, which are crucial to understanding the overall movements of proteins during ligand binding. The "least square fit" to the reference structure is used to eliminate the molecule's translational and rotational mobilities. The "time-dependent movements" that the components carry out in a specific vibrational mode are demonstrated by projecting the trajectory onto a particular eigenvector. The average of the projection's time signifies the involvement of atomic vibration components in this form of synchronized motion. Using the "g_covar" and "g_anaeig" tools, which are already included in the GROMACS software package, the PCA was performed by first creating the covariance matrix of the Cα-atoms of the protein and then diagonalizing it. The *xmgrace* tool was used to plot the graphs.[60−62]

To determine if the motion between atom pairs is correlated (positive or negative), the dynamic cross-correlation matrix (DCCM) measures the magnitude of all pairwise cross-correlation coefficients. Herein, we investigated each element of DCCM, where $C_{ij} = 1$ representing the case of positively correlated fluctuations of atoms *i* and *j* have the same period and same phase, while $C_{ij} = -1$ and $C_{ij} = 0$, respectively, represent negatively or not correlated.[63,64]

2.12. Statistical Analysis. This study investigated the expression of hub genes in the GEPIA2.0 database and their connection with GBM using ANOVA. $|\log_2 \text{fold change}|$ cutoff ≤ 1.5 and *Q*-value ≤ 0.05 were considered significant. Tukey's Honest Significant Difference statistics were employed in the GlioVis database, where the *p*-value of the pairwise comparisons was used (***$p \leq 0.001$; **$p \leq 0.01$; *$p \leq 0.05$; ns, not significant). In TIMER2.0, the Wilcoxon test's statistical significance was indicated by the number of stars (***$p \leq 0.001$; **$p \leq 0.01$; *$p \leq 0.05$; ns, not significant). In the TIMER database analysis, a partial Spearman's correlation was applied. When $|\rho| > 0.1$, it indicated a correlation between the genes and immune cells. Significant data in the biological and KEGG pathway enrichment were screened according to *p*-value ≤ 0.05 with the Students' *t*-test.

3. RESULTS

3.1. Omics Data Mining and Identification of DEGs in GBM Hypoxia Condition. This study used the expression profile (GSE77307) from the NCBI-GEO database to identify DEGs exclusively expressed in hypoxia-induced GBM because targeting the hypoxic microenvironment could be a new tool for treatment.[7] Cells derived from GBM patient tumors and normal brain tissue were grown in hypoxic and normoxic conditions. GEO's raw RNA sequence (RNA-seq) data were processed and uploaded to GREIN using the GEO RNA-seq experiments processing (GREP2) pipeline. GREIN workflows with a graphical user interface provide complete interpretation, visualization, and analysis of processed data sets.[65] A normalized MA plot has been shown in Supporting Information Figure S1. GBM cancer cell model (U87-MG) and the human non-neoplastic brain cell model (HEB) were analyzed separately by comparing hypoxia with normoxia conditions to find dysregulated genes in hypoxia conditions. Subsequently, Venn's analysis demonstrated the involvement of 364 genes that were common in hypoxia conditions in both cell lines. 591 and 2429 genes

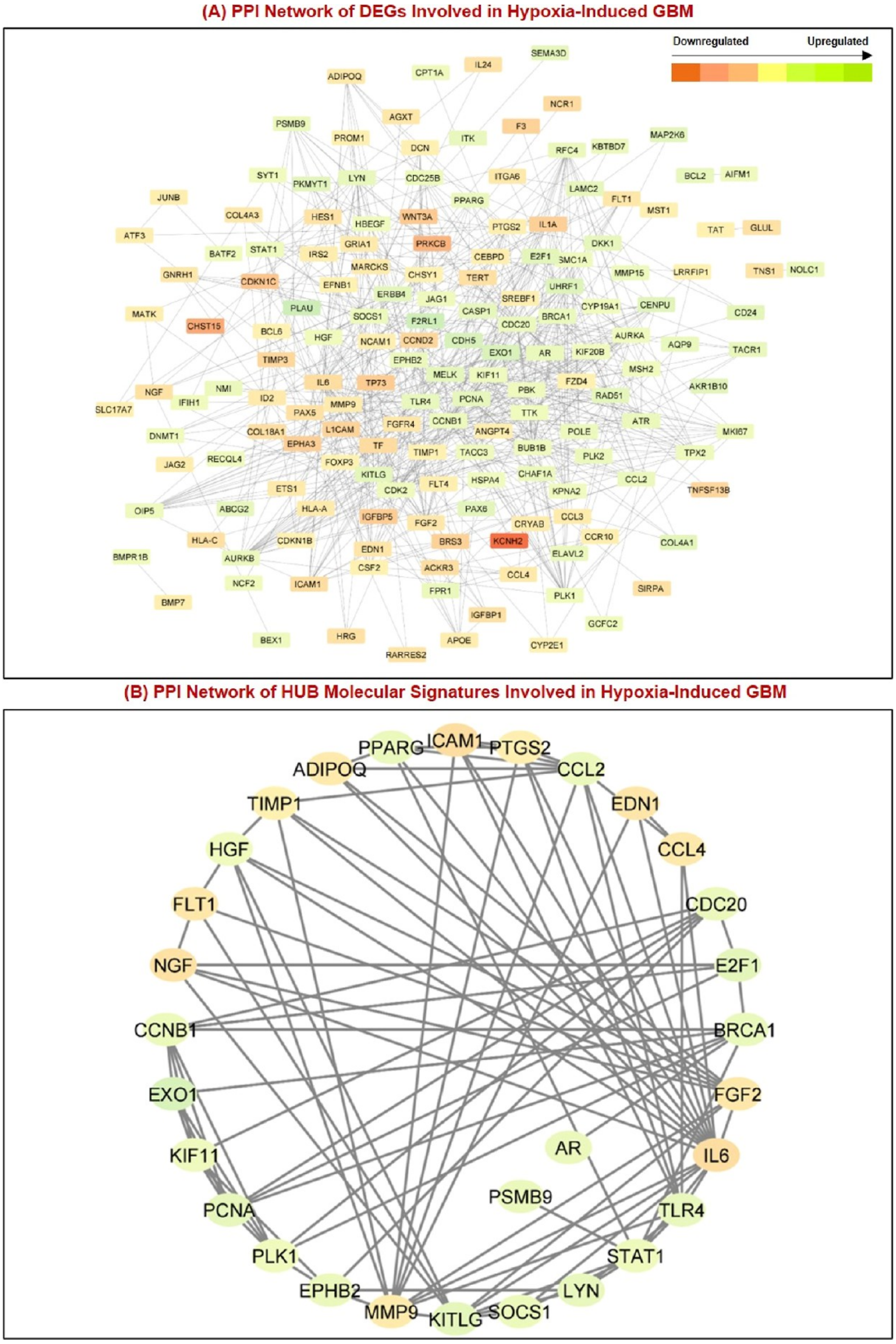


Figure 2. PPI network complex and modular analysis. (A) Module 1: a total of 241 DEGs (129 upregulated genes and 112 downregulated genes) were filtered into the DEG PPI network complex using STRING and Cytoscape software. It was composed of 163 nodes and 592 edges. (B) Module 2 showed a PPI network of 32 hub genes. Nodes in green signify upregulation and nodes in red signify downregulation. The colors from red to green represent the intensities of expression ($\log_2$ fold change, value: −6 to +14; cutoff value ±1.5), where red represents downregulation and green represents upregulation. In the presented figure, varying shades of red (from dark to light) show a decrease in the expression of downregulated genes, while shades of green (from light to dark) show increase in the expression of upregulated genes. Upregulated genes with $\log_2$ fold change ≥ 1.5 and downregulated genes with $\log_2$ fold change ≤ 1.5. STRING: Search Tool for the Retrieval of Interacting Genes/Proteins database.

expressed exclusively in hypoxia conditions in HEB and U87-MG cell lines, respectively.[66] Among them, we were interested in 2429 hypoxia-related DEGs exclusively expressed in hypoxia conditions and hence were considered for further analysis (Figure 1B,a). DAVID enrichment analysis of 2429 genes revealed that 30 genes have a significant association with GBM. In addition, G2SD enrichment (default cutoff parameter) showed 25 genes related to GBM. Similarly, GWAS-2019 and DisGeNET of Enrichr webtool enrichment analysis showed 3 and 242 genes linked with GBM, respectively. When we

Table 1. In Silico Expression Analysis and Validation of all 32 HUB Signatures Using Various Databases Containing Data from GBM Patient Samples[a]

Gene Name	RNA sequence dataset				Microarray datasets		
	GEPIA2	TIMER2.0	GlioVis				
			TCGA_GBM	GILL	REMBRANDT	AGILENT-4502a	Gravendeel
ADIPOQ	ns	*	**	ns	ns	ns	ns
AR	ns	**	**	**	***	***	**
BRCA1	*	***	***	***	***	***	***
CCL2	*	ns	ns	***	ns	*	***
CCL4	*	ns	ns	***	ns	ns	**
CCNB1	*	***	***	***	***	***	***
CDC20	*	***	***	***	***	***	***
E2F1	ns	*	**	ns	ns	**	ns
EDN1	ns	ns	ns	ns	ns	ns	ns
EPHB2	ns	ns	ns	ns	**	*	ns
EXO1	*	***	***	***	*	***	***
FGF2	ns	ns	ns	***	*	ns	ns
FLT1	ns	ns	ns	ns	ns	ns	ns
HGF	ns	ns	ns	ns	ns	ns	ns
ICAM1	*	ns	ns	***	*	*	ns
IL6	ns	ns	ns	***	ns	ns	ns
KIF11	*	***	***	***	***	***	***
KITLG	ns	ns	ns	ns	**	ns	ns
LYN	*	***	***	***	***	**	***
MMP9	*	***	***	***	***	***	**
	ns	***	**	**	***	***	ns
NGF	ns	**	*	***	***	ns	ns
PCNA	*	***	***	***	***	***	***
PLK1	*	***	***	***	***	***	ns
PPARG	ns	ns	ns	ns	ns	ns	ns
PSMB9	*	***	***	***	***	***	***
PTGS2	ns	ns	ns	ns	ns	ns	ns
SOCS1	*	***	***	***	***	*	ns
STAT1	*	*	ns	***	***	***	***
TF	ns	**	ns	***	***	***	***
TIMP1	*	***	***	***	***	***	***
TLR4	ns	ns	ns	ns	ns	ns	ns
SAMPLE SIZE							
GBM TUMOR	163	156	75	153	219	489	159
NORMAL TISSUES	207	4	17	5	28	10	8

[a] Dark green color = ***$p \leq 0.001$; medium green color = **$p \leq 0.01$; light green color = *$p \leq 0.05$; gray color = ns, not significant. In all seven GBM patient databases, including four RNA sequence data sets and three microarray data sets; the gene name printed in blue is among the top 10 hub genes that are significantly dysregulated.

(A) Correlation Analysis of 10 HUB Molecular Signatures with GBM Tumor Microenvironment

Gene Name	Variable	Purity	B Cell	CD8+ T Cell	CD4+ T Cell	Macrophage	Neutrophil	Dendritic Cell
BRCA1	partial.correlation	0.312	-0.132	0.042	0.090	0.048	0.149	0.090
	p-value	0.000	0.007	0.396	0.066	0.327	0.002	0.065
CCNB1	partial.correlation	0.347	-0.069	0.011	-0.161	-0.069	-0.038	0.070
	p-value	0.000	0.159	0.823	0.001	0.161	0.441	0.154
CDC20	partial.correlation	0.413	-0.135	-0.056	-0.091	-0.073	-0.086	0.054
	p-value	0.000	0.006	0.257	0.062	0.136	0.078	0.267
EXO1	partial.correlation	0.487	-0.067	-0.059	-0.072	-0.052	-0.054	-0.022
	p-value	0.000	0.174	0.225	0.141	0.293	0.268	0.656
KIF11	partial.correlation	0.404	-0.098	-0.018	0.004	-0.008	0.058	0.073
	p-value	0.000	0.046	0.721	0.932	0.863	0.234	0.138
LYN	partial.correlation	-0.439	0.277	-0.358	0.232	0.241	0.399	0.500
	p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MMP9	partial.correlation	-0.173	-0.095	-0.186	-0.059	-0.086	0.037	0.333
	p-value	0.000	0.052	0.000	0.228	0.079	0.449	0.000
PCNA	partial.correlation	0.382	0.061	0.054	-0.108	-0.016	-0.028	0.077
	p-value	0.000	0.211	0.267	0.027	0.750	0.564	0.117
PSMB9	partial.correlation	-0.172	0.324	-0.214	-0.031	0.099	0.153	0.229
	p-value	0.000	0.000	0.000	0.527	0.042	0.002	0.000
TIMP1	partial.correlation	-0.392	-0.119	0.002	0.009	0.050	-0.120	0.547
	p-value	0.000	0.015	0.975	0.847	0.310	0.014	0.000

Spearman positive correlation (ρ>0, p<0.05)

Spearman negative correlation (ρ<0, p<0.05)

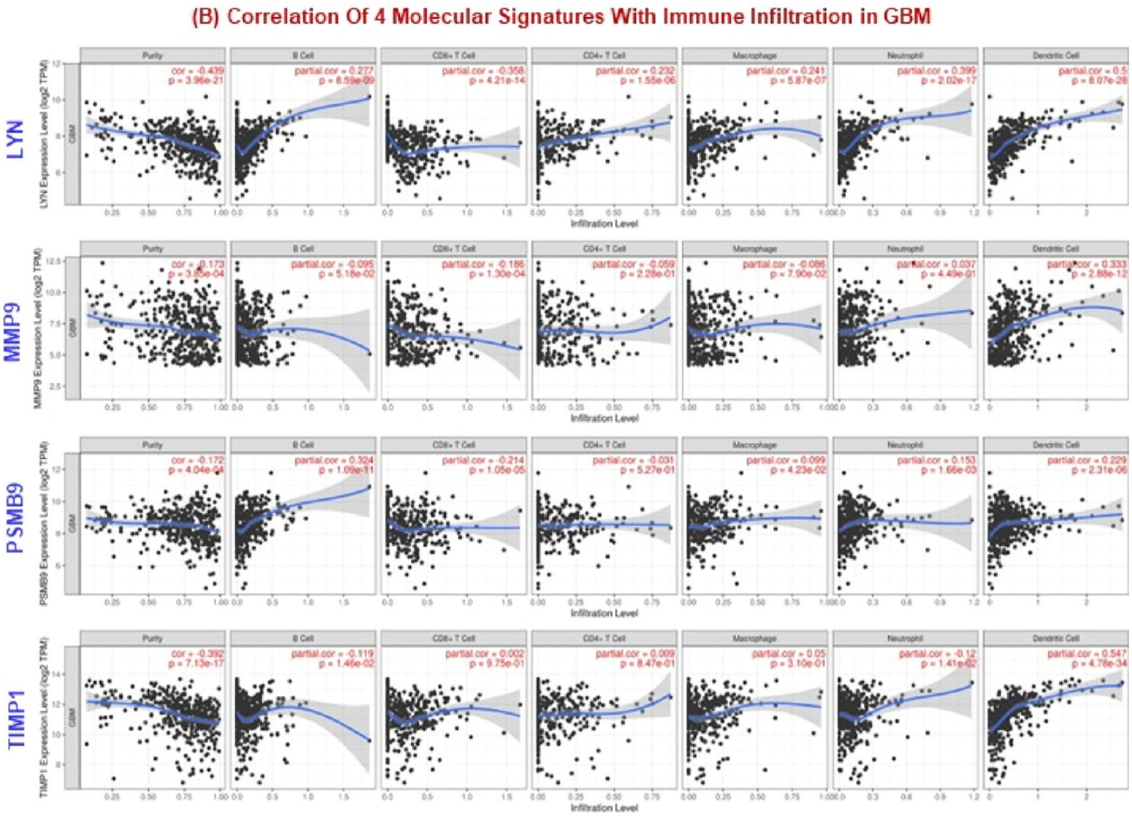

(B) Correlation Of 4 Molecular Signatures With Immune Infiltration in GBM

Figure 3. (A) Correlation analysis of 10 validated hub genes in GBM patient's data sets with tumor purity and six tumor infiltrating immune cells (B-cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophiles, and DCs). Genes highlighted in blue show negative tumor purity and hence shortlisted for further analysis. (B) Scatterplots from the TCGA-GBM data set illustrating the relationship between LYN, MMP9, PSMB9, and TIMP1 gene expressions and tumor purity and six key tumor infiltrating immune cell types in GBM. On the left-most panel, gene expression levels are compared to tumor purity, and genes that are highly expressed in the microenvironment are expected to have negative associations with tumor purity. In the TIMER database analysis, partial Spearman's correlation was applied. When $|\rho| > 0.1$ and p-value ≤ 0.05, it indicated that there was a link between the genes and immune cells. In general, the smaller the ρ value, the smoother the curve; the larger the ρ value, the fuller the curve; when $\rho < 0.5$, the curve is ellipse; when $\rho = 0.5$, the curve is parabola; when $\rho \geq 0.5$, the curve is hyperbola.

integrated the 3 enrichment analysis methods, a total of 241 GBM-related DEGs were documented, including 129 upregulated genes and 112 downregulated genes (Supporting Information Table S1).

3.2. PPI Analysis and Exploration of HUB Signatures in Hypoxia-Induced GBM. With the help of the STRING database on Cytoscape software, we evaluated the PPI network comprising 241 DEGs based on coexpression to explore the possibility of hub genes. The network consists of 163 nodes and 592 edges with a high confidence score of $\geq$0.700. Molecular signatures in the network were displayed based on their expression (green for upregulation, red for downregulation) and intensity based on fold change (log fold change, value: −6 to +14). To evaluate the importance of nodes in the PPI network, the topological parameters, including degree centrality and betweenness centrality, were calculated and utilized in the present study using the CentiScaPe plugin in Cytoscape software to find hub genes. We observed degree with a range of 1−14 and betweenness with a range of 0−684. Using the online Venny 2.0 tool, we observed the exchange and generated a Venn plot between "degree" and "betweenness" (Figure 1B,b). The 32 hub genes, a small number of critical nodes for the protein interactions in the PPI network, were chosen with a degree centrality > 7.00 (average value) and betweenness centrality > 342 (average value). PPI networks for DEGs and hub genes are shown in Figure 2A,B, respectively.

3.3. Validation of HUB Signatures in GBM Patients. We conducted the expression analysis of all 32 HUB signatures using various online web servers for RNA sequencing data, such as GEPIA2.0, TIMER2.0, TCGA-GBM, and GlioVis-GILL, and microarray data, such as GlioVis-REMBRAND, GlioVis-AGILENT, and GlioVis-Gravendeel. These web servers from the TCGA project provide extensive information concerning GBM patients. The expression of all 32 genes was examined using the databases described above as described in Table 1. Based on the selection criteria (***$p \leq 0.001$; **$p \leq 0.01$; *$p \leq 0.05$; ns, not significant), 10 genes out of 32 exhibited significant expression levels in both RNA and microarray databases of GBM patient samples. This also explains that these 10 molecular signatures, namel,y BRCA1, CCNB1, CDC20, EXO1, KIF11, LYN, MMP9, PCNA, PSMB9, and TIMP1, were expressed in GBM tumor samples. Molecular function of these signatures and their role in various malignancies have been briefly explained here. Breast cancer gene 1 (BRCA1) is a tumor suppressor protein that is essential for DNA damage repair, chromatin remodeling, and cell cycle regulation. Mutations in BRCA1 cause genetic changes, cancer, and a failure to repair DNA damage. Patients with BRCA1 germ line mutations have been associated with sporadic instances of GBM.[67] Cyclin B1 (CCNB1) and cell division cycle protein 20 (CDC20), both of which are associated with cell progression, demonstrated that their increased expression was substantially correlated with poor survival in GBM.[68] Exonuclease 1 (EXO1) is a member of the DNA damage repair enzyme family that is particularly active in homologous recombination (HR) and nonhomologous end-joining following DNA double-strand breaks. It increases cell proliferation, invasion, and metastasis in glioma and hepatocellular carcinoma.[69] According to Liu et al., increased Kinesin family member 11 (KIF11) enhances cell cycle development and chemoresistance, negatively correlates with the TP53 expression, and is a major cause of malignancy in GBM.[70] Lck/yes-related protein tyrosine kinase (LYN) showed a substantial positive connection with PD-L1, was connected to the control of carcinogenic genes, and was engaged in tumor mutation. In gliomas, LYN may serve as both a potential diagnostic and immunotherapy marker.[71] Likewise, the proliferative capacity of cells is impacted by high MMP9 expression in gliomas, which is also linked to patient survival rates.[72] Proteasome 20S subunit beta 9 (PSMB9), along with PSMB8 and PSMB10 genes that encode catalytic subunits of the immunoproteasome, was overexpressed in GBM and was reported by Liu et al. as a novel biomarker for lower-grade glioma prognosis and can be exploited as an immunotherapy target.[73] Similarly, a study by Smith et al., demonstrated that proliferating cell nuclear antigen (PCNA), a nuclear DNA replication and repair protein, has increased expression and poor prognosis in pancreatic ductal adenocarcinoma.[74] Last but not least, tissue inhibitor of metalloproteinases-1 (TIMP-1) is known to control the proteolytic activity of the MMPs that break down the extracellular matrix. High tumor TIMP-1 protein expression in GBM has been linked to irinotecan resistance and anticipated to predict lower overall survival in GBM.[75]

Thus, only 10 molecular signatures were selected for further analysis, which were significantly expressed in all seven patient GBM databases.

3.4. Correlation between HUB Signatures and GBM TME. Here, in this study, to filter out molecular signatures involved in TME, we used the TIMER database to investigate the connection and correlation of 10 molecular signatures (BRCA1, CCNB1, CDC20, EXO1, KIF11, LYN, MMP9, PCNA, PSMB9, and TIMP1) expression with tumor purity and immune cell infiltration in patients with hypoxia-induced GBM. Data have been compiled in Figure 3A. In addition, we used GBM data sets to estimate the amounts of infiltration of six immune cell types [(CD4+ T cells, CD8+ T cells, B cells, macrophages, neutrophils, and DCs). Tumor purity normalized spearman correlation analyses revealed a positive and negative correlation expression of hub genes with B cells, CD4+ T cells, CD8+ T cells, macrophages, neutrophils, and DCs in GBM cancer. After the inputs are successfully entered, scatterplots will be created and displayed, displaying the purity-corrected partial Spearman's rho value (ρ) and statistical significance. Genes with negative associations with tumor purity are highly expressed in TME, and positive associations are highly expressed in the tumor cells. Finally, we discovered four molecular signatures (LYN, MMP9, PSMB1, and TIMP1) with negative tumor purity, and it implicated in the GBM's hypoxic microenvironment. Figure 3B illustrates the scatterplot showing the relationship between LYN, MMP9, PSMB9, and TIMP1 gene expressions and tumor purity and six key tumor-infiltrating immune cell types in GBM.

LYN expression shown positive correlation with B cells (ρ = 0.28, p < 0.001), CD8+ T cells (ρ = 0.23, p < 0.001), macrophages (ρ = 0.24, p < 0.001), neutrophils (ρ = 0.39, p < 0.001), and DCs (ρ = 0.49, p < 0.001) and negative correlation with $CD8^+$ T Cells (ρ = −0.35, p < 0.001) in GBM. MMP9 shows positive correlation with DCs (ρ = 0.33, p < 0.001) and negative correlation with CD8+ T Cells (ρ = −0.18, p < 0.001). PSMB9 showed positive correlation with B cells (ρ = 0.32, p < 0.001), macrophages (ρ = 0.99, p < 0.001), neutrophils (ρ = 0.15, p < 0.001), and DCs (ρ = 0.22, p < 0.001) and negative correlation with $CD8^+$ T Cells (ρ = −0.21, p < 0.001).

A study by Wang et al., showed that cancer-derived MMP9 plays a crucial role in the development of tolerogenic DCs which further affects regulatory T cells (T_{reg}) in the case of laryngeal cancer.[76] Similarly, mounting evidence suggested that MMP9

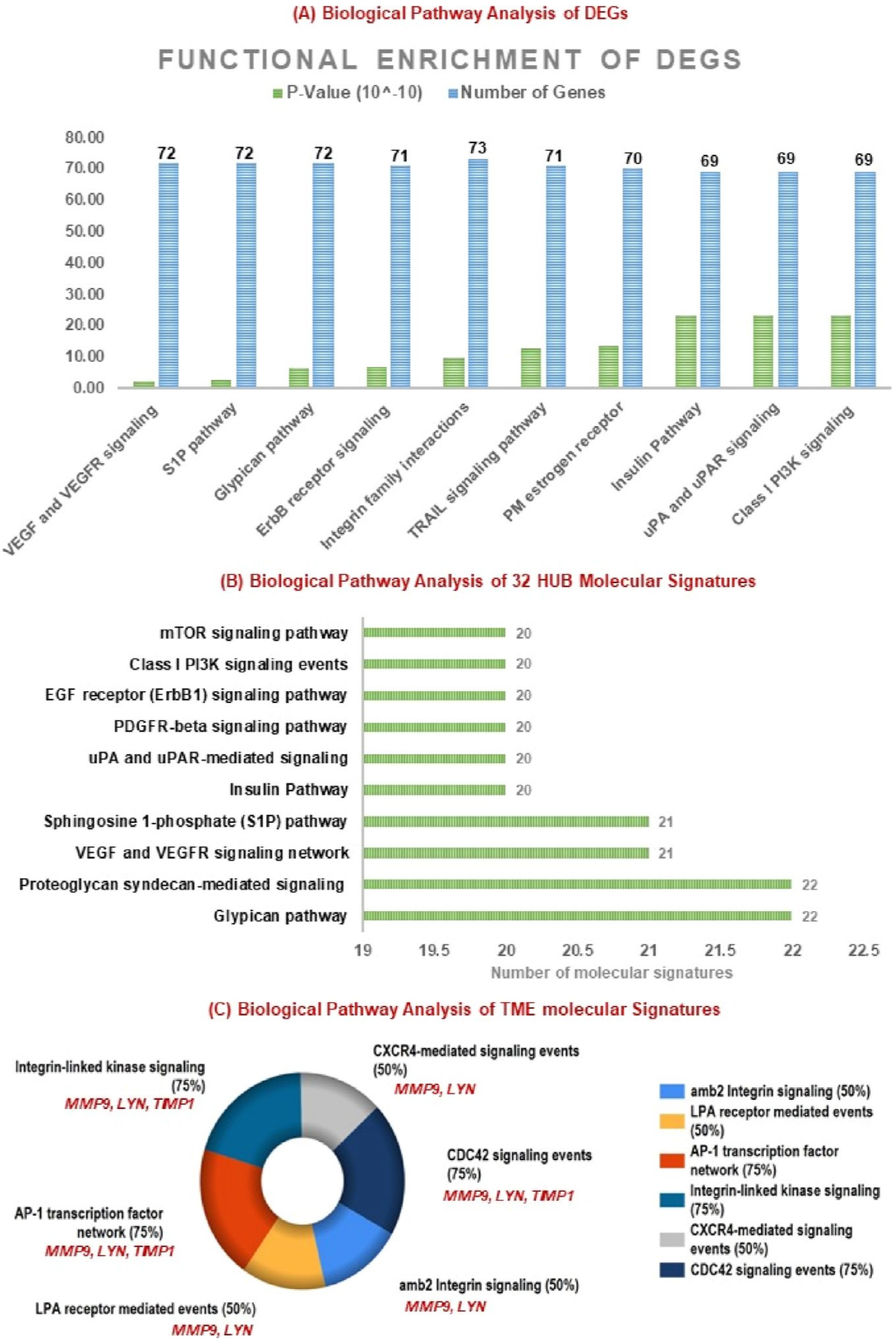


Figure 4. Significantly enriched biological pathway analysis: (A) Top 10 significantly functional enriched biological pathway terms of 241 DEGs associated with hypoxia-GBM. (B) Top 10 significantly functional enriched biological pathway terms of 32 hub signatures associated with hypoxia-GBM. (C) Top six enriched pathways of four molecular signatures (LYN, MMP9, PSMB9, and TIMP1) linked with the GBM microenvironment. Functional and signaling pathway enrichments were conducted using the KEGG pathway (http://www.genome.jp/kegg) and FunRich tool.

was involved in cancer-related inflammation by proteolyzing extracellular signal proteins, primarily those belonging to the CXC (C-X-C motif) chemokine family. As a result, MMP9 is regarded as a key architect and organizer of the tumor immune microenvironment.[77] Last TIMP1 expression linked positively with DCs ($\rho = 0.54$, $p < 0.001$) and negatively with B cells ($\rho = -0.11$, $p < 0.001$) and neutrophils ($\rho = -0.11$, $p < 0.001$). In contrast, BRCA1, CCNB1, CDC20, EXO1, KIF11, and PCNA showed positive correlations with tumor purity, attributed to their predominant expression and functions in tumor cells. Further, we identified the relationship between somatic cell number alteration and the presence of immune infiltrates of four genes (Supporting Information Figure S2A). Additionally, we have examined the connection between these molecular signatures and immune checkpoint inhibitors (ICIs), including PDCD1(PD1), CD274(PDL1), CTLA4, LAG-3, and HAVCR2(TIM-3) (Supporting Information Figure S2B). According to data, the genes LYN, PSMB9, and TIMP1 were all positively correlated with ICIs except for LAG3, while TIMP1 was negatively correlated with LAG3. MMP9 only had positive correlation with PD1 and TIM-3.

Therefore, we have discovered four molecular signatures, LYN, MMP9, PSMB9, and TIMP1, to target the microenvironment of GBM and to further research whether they are therapeutic targets or not. The study concluded that LYN and PSMB9 were downregulated in hypoxia-induced GBM with $\log_2$ fold change values of −2.247 and −2.096, whereas MMP9 and TIMP1 were upregulated with $\log_2$ fold change values of 2.144 and 1.647, respectively. Thus, TIPM1 and MMP9 were selected for the identification of novel natural compounds in hypoxia-induced GBM therapeutics. However, TIMP1 lacks the approved control drug in terms of chemical compound and hence discarded for further analysis. Thus, the current study aims to identify the novel natural compound against MMP9 in hypoxia-induced GBM.

3.5. Biological Pathway Analysis of DEGs, HUB Molecular Signatures, and TME-Related Signatures. Biological pathway analysis using FunRich software was performed on 241 DEGs, 32 hub genes, and 4 genes involved in TME. As shown in Figure 4A, DEGs involved in the top 10 significant biological pathways were (a) VEGF and VEGFR signaling, (b) sphingosine 1-phosphate (S1P) pathways, (c) glypican pathway, (d) ErbB receptor signaling pathway, (e) integrin family cell surface interactions, (f) TRAIL signaling pathway, (g) plasma membrane estrogen receptor signaling, (h) insulin Pathway, (i) urokinase-type plasminogen activator (uPA) and uPAR-mediated signaling, and (j) class I phosphatidylinositol-3-kinase (PI3K) signaling. Similarly, analysis of 32 hub genes enhanced in biological pathways were (Figure 4B) (a) glypican pathway, (b) proteoglycan syndecan-mediated signaling, (c) VEGF and VEGFR signaling, (d) S1P pathway, (e) insulin pathway, (f) uPA and uPAR-mediated signaling, (g) PDGFR-beta signaling, (h) ErbB1 signaling pathway, (i) class I PI3K signaling, and (j) mTOR signaling pathway. In addition, we have also analyzed four shortlisted molecular signatures involved in TME in Figure 4C to understand the major pathways involved, which were (a) integrin-linked kinase (ILK) signaling, (b) activating protein-1 (AP-1) transcription factor network, (c) CDC42 signaling events, (d) CXCR4-mediated signaling, (e) Amb2 integrin signaling, and (f) lysophosphatidic acid (LPA) receptor-mediated. Biological pathways with p-value ≤ 0.05 and count > 2 were measured as statistically significant.

3.6. Localization Study and Construction of Target Signature−Regulatory Transcription Factor Network. Based on the CELLO localization predictor, we have predicted the localization of four genes using their amino acid protein sequences. Results showed that MMP9 and TIMP1 were majorly localized in the extracellular space, followed by the plasma membrane. At the same time, LYN and PSMB9 were localized in the cytoplasm and chloroplast, respectively (Supporting Information Figure S3A). Further, we have predicted target genes (LYN, PSMB9, MMP9, and TIMP1) related to TFs and their expression in GBM patient samples using JASPAR and GEPIA2.0 databases, respectively. The main transcription factor and its targets are listed in (Supporting Information Figure S3B.1). TIMP1, MMP9, and PSMB9 all share the Yin Yang 1 (YY1) TF with the highest degree (3) and betweenness (109.00), but the expression in the GBM patient sample is not statistically significant. In contrast, TIMP1 and PSMB9 shared the RELA (degree: 2; betweenness: 33.83), but TFAP2A and NFKB1 were elevated against PSMB9 with $\log_2$ fold change ≥ 1.4 (p-value ≤ 0.05) in GBM. However, TFs against the MMP9 gene were FOS, JUN, and TP53. These TFs were upregulated in GBM ($\log_2$ fold change ≥ 1.5, p-value ≤ 0.05), whereas STAT3 was only upregulated TF against the LYN gene. Supporting Information Figure S3B.2 demonstrates the network showing the associated transcription factor with molecular signatures in GBM.

3.7. Screening of Natural Compounds Based on BBB and ADMET Analyses. We received plant-derived naturals compounds from the NPACT database, including terpenoids, flavonoids, alkaloids, polycyclic aromatic natural compounds, aliphatic natural compounds, tannin, and PubMed database. We carried out BBB permeability of all-natural compounds using the SwissADME and CBLigand online tool with a cutoff value of 0.02 as we know that protein associated with GBM will be found in the particular region of the brain; thus, for a drug to be effective, it must pass the BBB.[78] In addition, these were checked for positive DLS based on drug-likeness score prediction.[79] Also, compounds were studied for Lipinski rule (MW ≤ 500; log P ≤ 5; HBA ≤ 10; HBD ≤ 5) and PAINS alert.[80] Sixty-five novel natural compounds had passed the criteria of BBB, Lipinski rule, PAINS, and drug-likeness, which went under ADMET (absorption, distribution, metabolism, excretion, and toxicity) analysis.[81] ADMET analysis of nominated compounds was carried out to check the pharmacokinetics and pharmacodynamics properties. This server was selected to assess whether a ligand (drug) is hepatotoxic, nephrotoxic, arrhythmogenic, carcinogenic, or respiratory toxic because poor pharmacokinetics and toxicity of candidate compounds are the significant reasons for drug development failure. Our study predicts 18 ADMET properties of selected compounds out of the 3 of absorption, 2 of distribution and excretion, 1 of metabolism, and 10 of toxicity properties.

For each compound to be an effective drug, it must fulfill these parameters which have their own range values such as (a) *Absorption*: Caco2 permeability > −5.15 log cm/s, MDCK permeability (Papp) > 20 × 10^{-6} cm/s, intestinal absorption > 30%; (b) *Distribution*: plasma protein binding ≤ 90%, volume distributionVD: 0.04−20 L/kg; (c) *Metabolism*: CYP1A2 inhibitor a cytochrome P450 enzymes. Inhibitors of CYP1A2 will boost the medication's plasma concentrations, and in some situations, this will result in negative consequences;[82] (d) *Excretion*: clearance of a drug ≥ 5, the half-life of a drug ($T_{1/2}$): 0−0.3; (e) *Toxicology*: human ether-a-go-go related gene (hERG blockers), human hepatotoxicity (H-HT), Drug-induced liver injury, AMES Toxicity, Rat Oral Acute Toxicity, toxic dose threshold of chemicals in humans (FDAMDD), skin sensitization, carcinogenicity, eye corrosion/irritation, and respiratory toxicity range between 0 and 0.3 (—): excellent (green); 0.3−0.7 (+)/(−): medium (yellow); 0.7−1.0 (++): poor (red).

Table 2. List of Identified 11 Natural Compounds and Their Toxicity Profiles[a]

PubChem CID	**158280**	**185609**	**10424988**	**13886678**	**44479222**	**15549893**	**124256**	**162334**	**1548943**	**101477139**	**14313693**
Natural Compounds	7,4'-dihydroxyflavan	4'-hydroxy-7-methoxyflavan	4,4'-dihydroxy-2,6-dimethoxydihydrochalcone	7-Hydroxy-2',4'-dimethoxyisoflavanone	(3R)-3-(4-Hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2H-1-benzopyran	4'-hydroxy-2,4-dimethoxydihydrochalcone	N-(4-hydroxyundecanoyl)anabasine	N-n-octanoylnornicotine	8-Methyl-N-Vanillyl-6-Nonenamide	Multidione	Naviculol
Molecular formula	$C_{15}H_{14}O_3$	$C_{16}H_{16}O_3$	$C_{17}H_{18}O_5$	$C_{17}H_{16}O_5$	$C_{17}H_{18}O_4$	$C_{17}H_{18}O_4$	$C_{21}H_{34}N_2O_2$	$C_{17}H_{26}N_2O$	$C_{18}H_{27}NO_3$	$C_{20}H_{28}O_3$	$C_{15}H_{26}O$
hERG Blockers	(---)	(--)	(---)	(---)	(--)	(--)	(---)	(---)	(---)	(---)	(---)
H-HT	(---)	(---)	(---)	(--)	(--)	(--)	(+)	(-)	(--)	(--)	(---)
DILI	(---)	(--)	(--)	(+)	(---)	(+)	(---)	(---)	(---)	(--)	(---)
AMES Toxicity	(-)	(+)	(---)	(+)	(---)	(--)	(---)	(---)	(---)	(---)	(---)
Rat Oral Acute Toxicity	(--)	(--)	(--)	(--)	(---)	(--)	(---)	(---)	(---)	(---)	(---)
FDAMDD	(+)	(+)	(-)	(+)	(++)	(-)	(+++)	(++)	(---)	(--)	(---)
Carcinogencity	(+)	(+)	(--)	(-)	(+)	(+)	(---)	(---)	(---)	(---)	(++)
Eye Corrosion	(+)	(---)	(---)	(---)	(---)	(---)	(---)	(---)	(---)	(---)	(---)
Eye Irritation	(+++)	(+++)	(+)	(---)	(++)	(++)	(---)	(---)	(---)	(+)	(+)
Respiratory Toxicity	(--)	(--)	(--)	(--)	(--)	(--)	(---)	(--)	(---)	(--)	(+)
Caco2 permeability (> -5.15log cm/s)	-4.691	-4.7	-4.695	-4.796	-4.663	-4.747	-4.68	-4.494	-4.476	-4.657	-4.205
MDCK Permeability (> 20X10-6 cm/s)	1.10E-05	1.40E-05	1.70E-05	3.40E-05	1.60E-05	2.10E-05	2.8E-05	1.90E-05	2.70E-05	2.10E-05	1.70E-05
Intestinal absorption	(---)	(---)	(---)	(---)	(---)	(---)	(---)	(---)	(---)	(---)	(---)
PPB (≤ 90%)	96.63%	97.48%	86.48%	98.13%	96.01%	91.47%	88.48%	86.43%	96.49%	98.34%	95.56%
VD (0.04-20L/kg)	1.111	1.194	0.595	0.55	1.044	0.574	0.956	0.867	1.098	0.316	1.553
CYP1A2 inhibitor	(+++)	(+++)	(+++)	(+++)	(+++)	(+++)	(---)	(-)	(++)	(-)	(--)
CL(≥ 5)	16.437	12.53	11.71	9.771	14.822	12.32	9.359	6.442	11.309	9.861	12.763
T1/2	0.757	0.335	0.914	0.384	0.813	0.818	0.3	0.281	0.892	0.465	0.22

[a] Color code: green/(—): signifies excellent with score range between 0 and 0.3; yellow/(+)/(−): signifies medium with score ranging between 0.3 and 0.7; red/(++/+++) signifies poor with score range between 0.7 and 1.0.

https://doi.org/10.1021/acsomega.3c00441

Papp is extensively considered to be the in vitro point of reference for estimating the uptake efficiency of compounds into the body. Papp values of MDCK cell lines were also used to estimate the effect of the BBB. hERG-(Category 0) compounds had an IC_{50} > 10 μM or <50% inhibition at 10 μM, whereas hERG + (Category 1) molecules will have the opposite of this. The voltage-gated potassium channel encoded by hERG genes plays a key function in controlling the exchange of cardiac action potential and resting potential during cardiac depolarization and repolarization. Long QT syndrome, arrhythmia, and Torsade de Pointes are all possible side effects of hERG blocking and can result in palpitations, fainting, or even death. Hepatotoxicity predicts the action of a compound on normal liver function. Furthermore, if the given compound is AMES positive, it will be considered mutagenic. Similarly, compounds have positive carcinogenicity because of their ability to damage the genome or disrupt the cellular metabolic processes. Recently, respiratory toxicity has become the leading cause of drug withdrawal. Drug-induced respiratory toxicity is frequently underdiagnosed due to the lack of recognizable early signs or symptoms in commonly used drugs, resulting in severe morbidity and mortality. As a result, thorough monitoring and treating respiratory toxicity are critical.[83,84] Our study indicates that all 11 predicted compounds, alkaloids (PubChem CID:124256, 162334, and 1548943), terpenoids (PubChem CID: 101477139 and 14313693), and flavonoids (PubChem CID: 158280, 185609, 10424988, 13886678, 44479222, and 15549893) fulfill the eligibility criteria and show favorable results. Therefore, we summarize in Table 2 that all 11 natural compounds meet the ADMET criteria for being a novel compound to target GBM. The detailed methodology used to screen natural compounds are shown in Supporting Information Figure S4, and the characteristics and physiochemical of natural compounds are mentioned in Supporting Information Table S2.

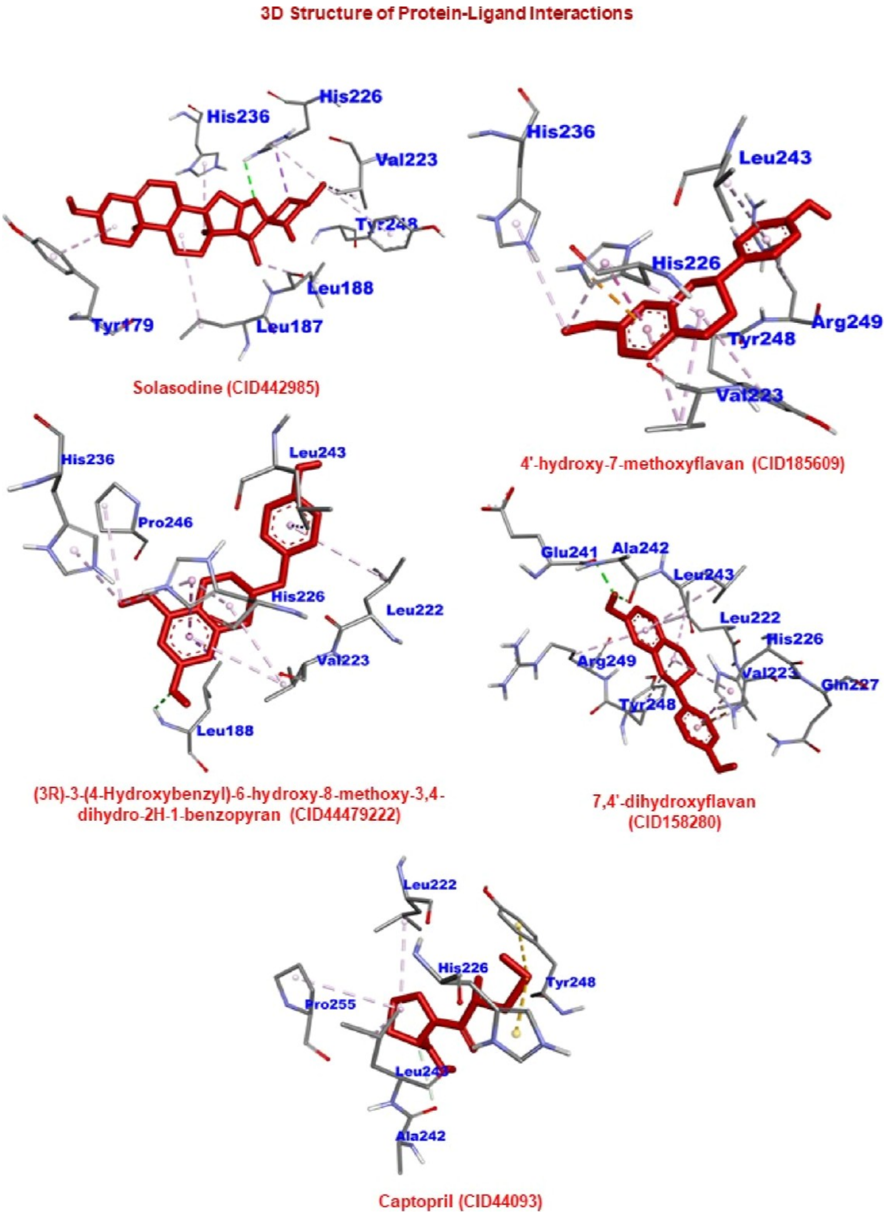
3.8. 7,4′-Dihydroxyflavan, (3*R*)-3-(4-Hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran), and 4′-Hydroxy-7-methoxyflavan) as Promising Natural Flavonoids Against MMP9: a Molecular Docking Approach.

To find effective drugs against the MMP9 gene, 11 natural compounds satisfied the filter criteria, and one reference drug, Captopril (FDA approved retrieved from the DrugBank database; https://www.drugbank.ca/) and one natural compound (Solasodine) from previous studies[85,86] were chosen. Autodock Vina 4.0 was used to perform blind molecular docking experiments of all prioritized natural compounds with MMP9 (PDB id: 4HMA) using default parameters. The docking or binding free energy screens the most effective chemicals and conformations. Table 3 depicts the particular docking binding energy [$-\Delta G$ value (kcal/mol)] and the detailed information regarding intermolecular interactions between ligands and proteins. In addition, we have predicted the binding residues for ligand binding using the PrankWeb tool. Pocket 1 with highest probability (0.99) was chosen whose residues for alpha chain were 179, 180, 186−193, 222, 223, 226, 227, 230, 233−238, 240, 242, 243, and 245−249.

Table 3. Binding Affinity and Binding Energy of Prioritized Natural Compounds along with the Reference Drug

Group	Reference Drug		Experimental Natural Compounds										
PubChem CID	*442985*	*44093*	*158280*	*44479222*	*185609*	*13886678*	*101477139*	*10424988*	*124256*	*15549893*	*1548943*	*162334*	*14313693*
Class of compounds	*Alkaloid*	*Small molecules*	*Flavonoid*	*Flavonoid*	*Flavonoid*	*Flavonoid*	*Terpenoid*	*Flavonoid*	*Alkaloid*	*Flavonoid*	*Alkaloid*	*Alkaloid*	*Terpenoid*
Ligand Name	Solasodine	Captopril	7,4'-dihydroxyflavan	(3R)-3-(4-Hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2H-1-benzopyran	4'-hydroxy-7-methoxyflavan	7-Hydroxy-2',4'-dimethoxyisoflavanone	Multidione	4,4'-dihydroxy-2,6-dimethoxydihydrochalcone	N-(4-hydroxyundecanoyl)anabasine	4'-hydroxy-2,4-dimethoxydihydrochalcone	8-Methyl-N-Vanillyl-6-Nonenamide	N-n-octanoylnornicotine	Naviculol
Total No. of interactions	15	14	17	17	16	16	17	16	22	15	17	14	9
No of interaction with active site residues	15	11	15	14	13	15	16	15	19	14	14	13	4
Binding Energy (kcal/mol)	-10.3	-6.6	-10.3	-10.3	-10	-8.5	-8.2	-8.2	-8.2	-8.2	-8.1	-7.1	-6.4
Conventional H-bond	HIS226	-	GLU241; ALA242	LEU188; HIS226	-	-	TYR248	HIS226	-	HIS226; GLN227; HIS236	GLN227; ARG249	TYR248	-
Carbon H-bond	-	ALA242	-	-	-	ALA189; HIS226	-	-	-	-	TYR245	PRO246	-
Van der waals	GLY186; ALA189; HIS190; ALA191; GLN227; HIS230; PRO246; MET247;	LEU188; VAL223; PRO240; GLU241; TYR245; MET247; ARG249; THR251	LEU188; HIS230; HIS236; PRO240; TYR245 PRO246; MET247; TYR251	ALA189; GLN227; GLU241; ALA242; TYR245; MET247; TYR248; ARG249; THR251; HIS257	HIS257; THR251; ALA242; LEU222; LEU188; GLN227; PRO246; MET247; TYR245; GLU241	LEU222; VAL223; GLN227; HIS236; LEU243; TYR245; PRO246 MET247; TYR248; ARG249; THR251	ALA189; GLY186; LEU188; TYR218; LEU222; GLN227; ALA242; TYR245; PRO246; MET247; ARG249; HIS236	GLY186; LEU187; LEU222; VAL223; GLN227; LEU243; TYR245; PRO246; ARG249; THR251	GLY186; LEU187; ALA189; HIS190 ALA191; GLN227 GLU241; ALA242; TYR245; PRO246; MET247; ARG249; THR251	ALA189; LEU243; TYR245; MET247; TYR248; ARG249; THR251	ALA189; HIS230; GLU241; PRO246; MET247 TYR248; THR251; HIS257	GLY186; LEU187; ALA189; GLN227; HIS236; TYR245; LEU243; MET247;	GLY233; ASN262; GLY263; LEU267
Alkyl/PI-alkyl	TYR179; LEU187; LEU188; VAL223; HIS226; HIS236; TYR248	LEU222; PRO225; LEU243	LEU222; VAL223; HIS226; LEU243; TYR248; ARG249	LEU222; VAL223; HIS226 HIS236; PRO246	VAL223; HIS236; LEU243; TYR248; ARG249;	LEU187; LEU188	LEU187; VAL223; HIS; 226; LEU243	LEU188; HIS226; HIS230; HIS236; MET247	LEU188; LEU222; VAL223; HIS226; HIS230; HIS236; LEU243; TYR248; PRO255	LEU188; LEU222; HIS226; HIS230; HIS236; PRO246	LEU188; LEU222; VAL223; HIS236; LEU243; ARG249	LEU188; TYR218; VAL223 HIS226	PHE110; LEU234; HIS266
Pi cation	-	-	HIS226	-	HIS226	-	HIS226		-	-	HIS226	-	-
Pi-Pi Stacked	-	-	-	HIS226	-	HIS226	TYR248	TYR248	-	HIS226		-	-
Pi-sigma	HIS226	-	-	LEU188; LEU243	-	-	-	-	-	-	-	-	-
Pi-sulphur		HIS226;	-	-	-	-	-	-	-	-		-	-
		TYR248											
Unfavorable donor-donor	-	-	GLN227	-	-	-	-	-	-	-	LEU188	-	ASP235

The MMP9 3D structure revealed that 88.6% of the residues were in the highly favored region and 0.4% were in the disallowed region respectively. Further structures were validated by ERRAT and VERIFY3D. The quality factor predicted by the ERRAT server for both alpha and beta chains of MMP9 was 76.17. VERIFY3D server predicted that 100% of residues had averaged a 3D−1D score ≥ 0.2respectively. Moreover, the docking energy of reference drugs, Captopril and Solasodine, were −6.6, −10.3 kcal/mol, respectively. Among 11 natural compounds, flavonoid 7,4′-dihydroxyflavan) and (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran) scored the highest binding energy −10.3 kcal/mol with 2 H-bond interaction with GLU241, ALA242, Leu188, and HIS226 than both the reference drugs, whereas 4′-hydroxy-7-methoxyflavan scored −10 kcal/mol binding energy with no H-bond interaction. Supporting Information Figure S5 shows two-



Solasodine (CID442985)

4'-hydroxy-7-methoxyflavan (CID185609)

(3R)-3-(4-Hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2H-1-benzopyran (CID44479222)

7,4'-dihydroxyflavan (CID158280)

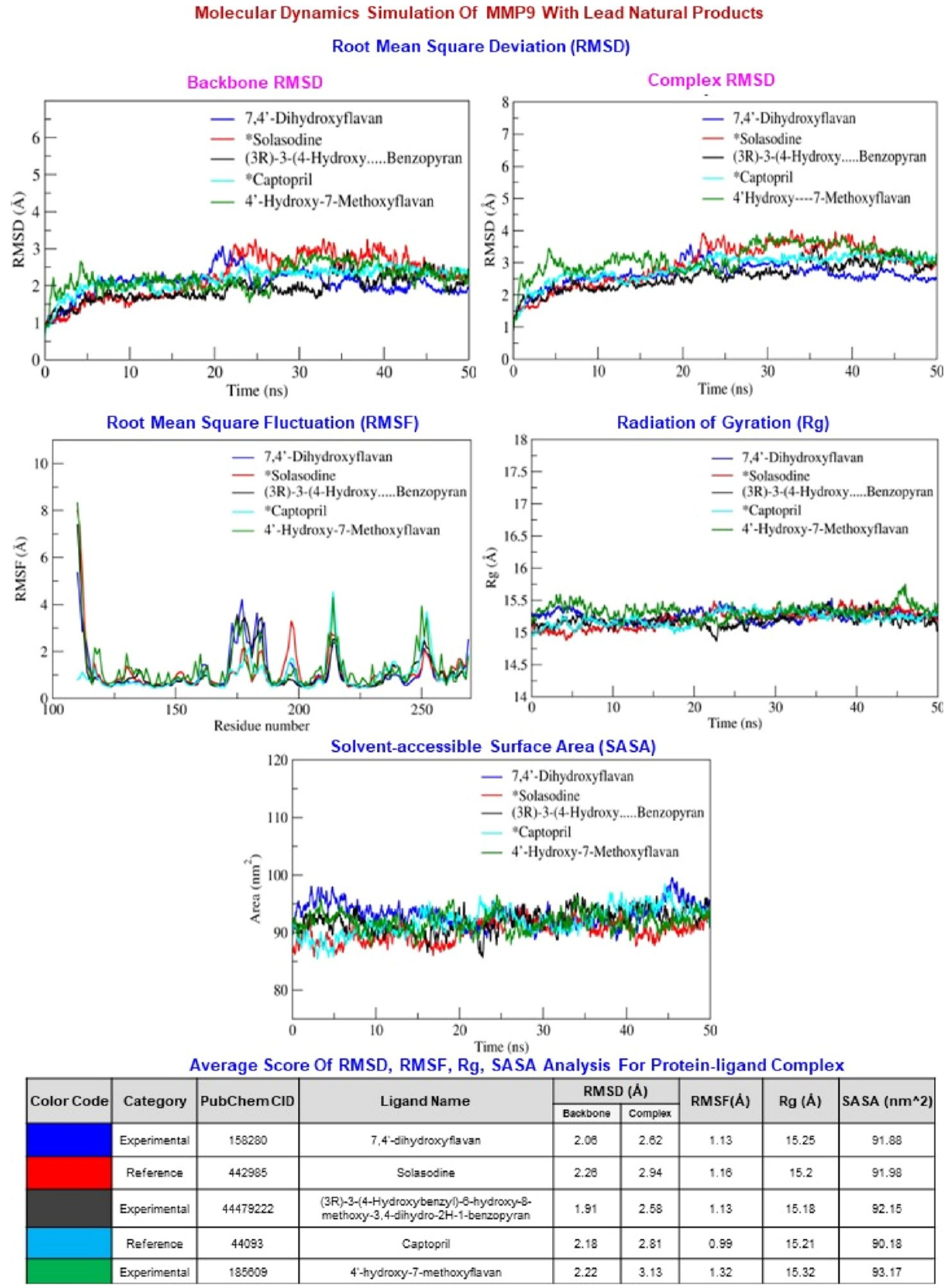
Captopril (CID44093)

Figure 5. 3D interaction diagrams for the docked complexes between MMP9 and ligands obtained in this study.

dimensional (2D) interaction diagrams for the docked complexes between MMP9 and ligand which includes all interactions such as H-bond and other interactions such as the van der waals force, π-alkyl, π-sigma, and so forth. Shortlisted natural compounds' binding energy and H-bond interaction have been tabulated in detail in Table 3. Three natural compounds 7,4′-dihydroxyflavan and (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran), and 4′-hydroxy-7-methoxyflavan) scoring the lowest binding energy and forming interaction with the active site were shortlisted for further studies along with Captopril and Solasodine. It was intriguing to note that all best-identified natural compounds showed stable and conserved intermolecular interactions as demonstrated in Figure 5.

3.9. Assessment of the Most Promising Protein–Ligand Complex by MD Simulation Run. MD simulation (RMSD, RMSF, R_g, and SASA) results of all mentioned protein–ligand complexes have been mentioned in Figure 6 along with the average score values of each parameter of three best-docked compounds and two reference drugs.

3.9.1. Stability of MMP9-7,4′-Dihydroxyflavan Complex. The time evolution of the RMSD was determined to check the structural stability of the protein in complex ligands during the simulation. The average RMSD values for the backbone and complex were ~2.06 and ~2.62 Å, respectively. The complex slightly deviated as RMSD > ~3 Å between 19 and 24 ns. At the binding site, a loop formed by the residues Pro240 and Arg249 that connects two helices displayed only slight residual fluctuations up to 0.9 Å. Flexible loops in the N-terminal region of the protein were extremely dynamic and exhibited RMSF > 2.5 Å. It was intriguing to observe that residues actively contributed to the stable interaction and exhibited significantly less fluctuation. The complex's overall average RMSF value was ~1.13 Å. The R_g value was determined for investigating the compactness and structural changes in the MMP9-7,4′-dihydroxyflavan complex. The root-mean-square distance of a protein atom in relation to the protein's center of mass is used to



Color Code	Category	PubChem CID	Ligand Name	RMSD (Å)		RMSF(Å)	Rg (Å)	SASA (nm^2)
				Backbone	Complex			
🟦	Experimental	158280	7,4'-dihydroxyflavan	2.06	2.62	1.13	15.25	91.88
🟥	Reference	442985	Solasodine	2.26	2.94	1.16	15.2	91.98
⬛	Experimental	44479222	(3R)-3-(4-Hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2H-1-benzopyran	1.91	2.58	1.13	15.18	92.15
🟦	Reference	44093	Captopril	2.18	2.81	0.99	15.21	90.18
🟩	Experimental	185609	4'-hydroxy-7-methoxyflavan	2.22	3.13	1.32	15.32	93.17

Figure 6. MD simulation analysis of MMP9 upon binding of the ligand as a function of time throughout 50 ns. Graph showing RMSD, RMSF, radius of gyration (R_g), and SASA for MMP9 with three best-docked compounds and two reference drugs.

compute the R_g value of the protein. The average value of R_g for the complex is ∼15.25 Å. The SASA was examined to study the protein compactness behavior. The initial and final surface areas occupied by the docked MMP9-7,4′-dihydroxyflavan complex are 91.40 and 92.90 nm^2, respectively, with an average surface area of ∼91.88 nm^2. This complex constructed two stable H-bonds, and both remained stagnant over the course of the simulations. The stable H-bond interactions were thought to be the primary factor that encouraged the stable complex formation. In addition, according to MM-PBSA calculation, the complex also demonstrated a binding energy of −85.24 kJ/mol. Moreover, the residues that contributed the most to the binding energy were found by computing the residue decomposition energy. The analysis suggested that five residues,

https://doi.org/10.1021/acsomega.3c00441

Table 4. MM-PBSA Calculations of Top Hit Complexes' Binding Free Energy and Interaction Energies[a]

complex	MM-PBSA (kJ/mol)				
	ΔE_{VDW}	ΔE_{ELE}	ΔG_{Sol}	ΔG_{Surf}	ΔG_{bind}
MMP9-7,4′-dihydroxyflavan	−167.19 ± 7.82	−14.98 ± 4.06	111.60 ± 9.88	−14.68 ± 0.78	−85.24 ± 11.81
MMP9-Solasodine	−148.31 ± 11.20	−777.73 ± 18.62	353.45 ± 15.04	−15.55 ± 0.91	−588.15 ± 17.82
MMP9-(3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran	−141.43 ± 13.78	−79.73 ± 8.29	142.50 ± 8.45	−15.49 ± 0.72	−94.16 ± 11.65
MMP9-4′-hydroxy-7-methoxyflavan	−154.50 ± 16.07	−27.86 ± 8.96	119.80 ± 20.33	−15.87 ± 0.90	−78.44 ± 16.16
MMP9 - Captopril	−83.65 ± 13.94	−622.30 ± 35.47	198.05 ± 38.01	−10.59 ± 1.54	−518.50 ± 22.39

[a] ΔE_{VDW}—van der Waal energy, ΔE_{ELE}—electrostatic energy, ΔG_{Sol}—polar solvation energy, ΔG_{Surf}—SASA energy, and ΔG_{bind}—binding energy.

namely, Leu222, Val223, Ala242, Met247, and Tyr248, contributed considerably to the creation of the stable complex. Most importantly, the residue Tyr248 showed significant contributions to the binding affinity by scoring the lowest contribution energy of −5.41 kJ/mol, followed by Leu222 (−4.71 kJ/mol), Met247 (−3.96 kJ/mol), Val223 (−2.67 kJ/mol), and Ala242 (−2.01 kJ/mol). However, residues Gln241and Pro255 did not favor the interactions.

3.9.2. Stability of MMP9-(3R)-3-(4-Hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2H-1-benzopyran Complex. This complex showed consistent structural stability during the simulation run for the 50 ns production run. Protein backbone and complex were found to have average RMSD values of ∼1.91 and ∼2.58 Å, respectively. The complex was a little unstable as RMSD > ∼3 Å between 33 and 37 ns and 39 to 47 ns. The maximum residual fluctuations in the N-terminal residues were >3.0 Å. However, the residues at the binding site from Leu222 to His230 (helix) and residues from Ala242 to Arg249 (loop) engaged in the stable and conserved nonbonded interactions and showed significantly much fewer variations of ∼0.5 and ∼1.13 Å, respectively. The complex has an average RMSF value of 1.13 Å. The average R_g value of 15.18 Å showed stable complex formation during the MD simulation by forming a compact structure. Meanwhile, the initial and final surface areas employed by the complex were 92.17 and 93.16 nm^2, with the average SASA score of the complex being 92.15 nm^2. During the simulation, this complex created five H-bonds, of which four were stable. The estimated binding affinity of the compound to MMP9 protein was −94.16 kJ/mol. Additionally, the residues Leu188, Leu222, Val223, His226, and Tyr248 encouraged stable complex formation. Most importantly, the decreasing order of binding affinity followed Leu222, Tyr248 and His226, Val223, and Leu188 with the lowest contribution energy of −5.74, −5.08, −4.58, −4.22, and −3.40 kJ/mol, respectively. However, the interactions were not favored by the residues Gln227 and Arg249.

3.9.3. Stability of MMP9-185609 (4′-Hydroxy-7-methoxyflavan) Complexes. The complex showed similar RMSD values of 50 ns and was stable. The complex's RMSD value ranged from 0.97 to 3.39 Å, whereas the backbone's RMSD value ranged from 0.85 to 2.5 Å. According to the residual fluctuations plotted for the C*α*, binding pockets encompassing residues between Leu222 and Gly229 (helix) and Ala242 and Arg249 (loop) showed the establishment of stable nonbonded contacts in residues with lower fluctuations. Residues at the N-terminal and residues adjacent to binding pockets, including Phe250 and Glu252, show higher residual fluctuation >3 Å due to increased local flexibility and ligand interaction observed during simulation. The overall average RMSF of the complex was 1.32 Å. Moreover, the R_g value demonstrated steady complex formation for 50 ns. In addition, the initial and final surface areas occupied by complexes were 91.63 and 96.49 nm^{2,} with the average SASA score of complexes being 93.17 nm^2. Two of the three H-bonds the complex created during the simulated period were consistent. The compound also had a binding energy of about −78.44 kJ/mol. Furthermore, the per-residue contribution energy showed six residues from the binding pocket, Leu188, Leu222, Val223, Leu243, Met247, and Tyr248, which had a considerable impact on the creation of a stable complex. The residues Leu188, Leu222, Val223, Leu243, Met247, and Tyr248 from the binding pocket showed significant contributions to the binding affinity by scoring the least residue decomposition/contribution energy of −2.36, −4.25, −6.22, −3.44, −2.22, and −4.23 kJ/mol, respectively. Arg249 residues do not favor the interaction.

3.9.4. Stability of MMP9-Captopril and MMP9-Solasodine Complexes. MMP9-Captopril and MMP9-Solasodine complexes showed stable interaction during the simulation run. The average RMSD value of the backbone and MMP9-Captopril complex was ∼2.18 and ∼2.81 Å, whereas the RMSD value with Solasodine was ∼2.26 and ∼2.94 Å. Moreover, the average RMSF values for the MMP9-Captopril complex and MMP9 and MMP9-Solasodine were 0.99 and 1.16 Å, respectively. Solasodine causes the N-terminal to fluctuate more than 3 Å, whereas Captopril did not cause this variation. Also, MMP9-Captopril and MMP9-Solasodine complexes have average R_g values of 15.21 and 15.2 Å, respectively. Meanwhile, MMP9-Captopril's initial and final surface areas were 88.85 and 91.54 nm^2, respectively, with an average SASA score of 90.18 nm^2. Comparatively, the MMP9-Solasodine complex had initial and final surface areas of 89.94 and 93.58 nm^2, with an average SASA score of 91.98 nm^2. Moreover, out of the three H-bonds formed, only two were stable during simulation for the Captopril complex and Solasodine complex. In addition, the the binding energy of MMP9-Captopril and MMP9-Solasodine was −518.50 and −588.15 kJ/mol, respectively. Furthermore, the MMP9-Captopril complex also showed 10 residues from the binding pocket, including Asp201, Asp205, Asp206, Asp207, Glu208, Asp235, Glu241, Glu252, Asp259, and Asp260, and significantly contributed to the stable complex formation. Likewise, 12residues, Asp177, Asp182, Asp201, Asp205, Asp206, Glu208, Asp235, Glu241, Pro246, Glu252, Asp259, and Asp260, helped create the stable MMP9-Solasodine complex.

Thus, data confirmed that the binding energies of MMP9 with ligands 7,4′-dihydroxyflavan, (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran, and 4′-hydroxy-7-methoxyflavan were similar (−10 kcal/mol) to that of the reference drug Solasodine and better than Captopril. All three natural compounds interact within the binding domain of the MMP9 pocket, and this interaction was stable for 50 ns with less deviation and fluctuations. The RMSD value difference

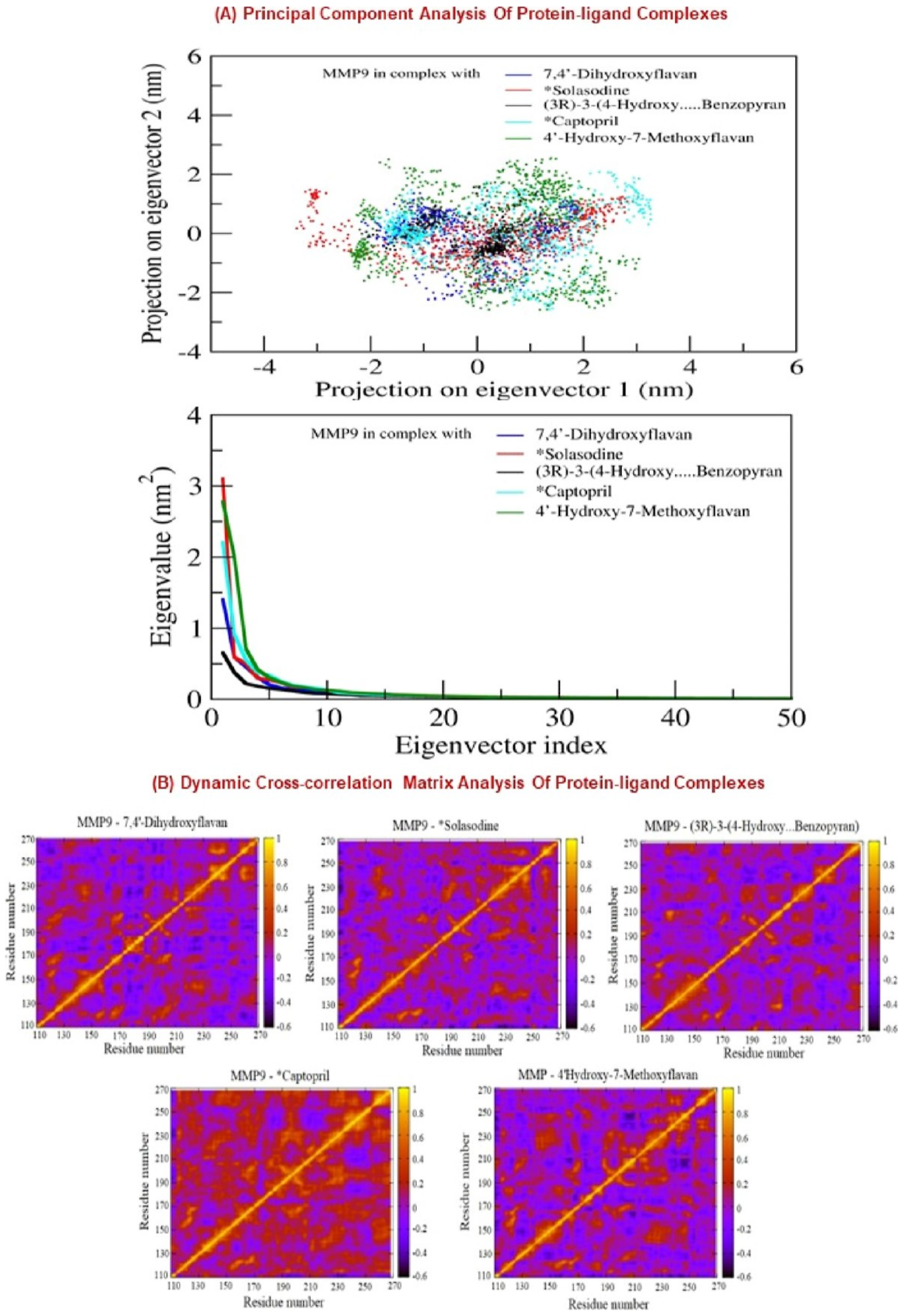


Figure 7. (A) PCA of protein−ligand complexes: In the scatterplot, the first two principal components (PC1, PC2) were plotted to analyze the collective motion of ligand-bound protein complexes during the simulations. The dots with different colors (blue, red, black, aqua, and green) represent the collective motion of MMP9 residue after ligand binding. Dots with smaller regions represent the higher structural stability and conformation flexibility and vice versa. The collective motion of MMP9 in the presence of ligands is depicted in the second graph using projections of MD trajectories onto two eigenvectors corresponding to the first two principal components. The first 50 eigenvectors were plotted versus eigenvalue for 5 ligands including 3 hit natural compounds and 2 reference drugs. Color code used in the scatterplot and graph: blue: 7,4′-dihydroxyflavan; red: Solasodine; black: (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran),;aqua: Captopril; Green: 4′-hydroxy-7-methoxyflavan. (B) DCCM of Cα atoms observed in complexes for 7,4′-dihydroxyflavan, Solasodine, (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran), Captopril, and 4′-hydroxy-7-methoxyflavan. The positive regions, colored amber, represent strongly correlated motions of Cα atoms ($C_{ij} = 1$), whereas the negative regions, colored blue, represent anticorrelated motions ($C_{ij} = -1$).

between the backbone and the complex was <3 Å. RMSF, R_g, and SASA also showed steady complex formation.

The g_mmpbsa tool computed the binding affinity of the protein−ligand complex using the MM-PBSA method. The free energy (kJ/mol) contribution of lead hits and standard molecules in relation to their respective targets is summarized in Table 4. In addition, detailed description of the total number of H-bond interactions in the protein−ligand complex has been shown in Supporting Information Figure S6A. Similarly, the contribution energy plot illustrated in Supporting Information Figure S6B exhibits the importance of the binding pocket residues in stable complex formation.

https://doi.org/10.1021/acsomega.3c00441

3.10. PCA and DCCM Analysis of Complexes. We employ PCA analysis to explore the dynamics of protein−ligand conformation for five complexes (two complexes with the reference drug and three complexes with the natural compound ligand) obtained from an MD simulation run of 50 ns. A PCA produces a matrix of eigenvectors and a list of related eigenvalues, which together represent the principal components and amplitudes of the internal movements of a protein. The first two eigenvectors/principal components (eigenvector 1 and eigenvector 2) are used to calculate the concerted motions of the past 50 ns trajectory since they can best describe the majority of the internal movements within a protein. The first two eigenvectors' 2D projection as well as the scatterplot are shown in Figure 7A. Captopril and Solasodine, two of the reference drugs employed in this study and directed at the MMP9 protein, were seen to have a greater range of conformations during the simulations (shown as a red and aqua line, respectively, in Figure 7A. Moreover, during simulation, the shortlisted MMP9-targeting ligands 7,4′-dihydroxyflavan, (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran, and MMP9-4′-hydroxy-7-methoxyflavan displayed less diversity than the reference drug (shown in blue, black, and green lines, respectively). Both the reference drugs demonstrated increased conformational flexibility with the maximum number of diverse conformations. Intriguingly, the MMP9 inhibitors 7,4′-dihydroxyflavan, (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran, and 4′-hydroxy-7-methoxyflavan took up substantially less conformational space than the Captopril reference drug. In contrast, only 7,4′-dihydroxyflavan, (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran performed better compared to the Solasodine reference drug as shown in the scatterplot (less dispersed plot). Therefore, we suggest that three lead-hit natural compounds could be more effective than the reference drugs.

The DCCM of Cα atoms in complexes provides a deeper structural understanding of the collective motion of the ligand-binding regions. The coordinated residual motion of the Cα atoms in each of the simulated complexes is shown in Figure 7B. Each residue exhibits a significant self-correlation with itself, as evidenced by the diagonal amber line. Scaling from amber to blue, respectively, is the strength of correlation (C_{ij} = 1) and anticorrelation (C_{ij} = −1). In complex MMP9-7,4′-dihydroxyflavan, the binding site residues show a positive correlation with the N-terminal domain of the MMP9. The scale of this correlation's amplitude goes from blue to amber color in smaller steps. Similarly, MMP9-MMP9-4′-hydroxy-7-methoxyflavan also showed a positive correlation with higher amplitude near binding site residues 222−249. In contrast, complex MMP9-44479222 showed anticorrelation, and its amplitude scaled from amber to blue color. The relevance of the active site residues in stabilizing the complexes was demonstrated by the coordinated motion displayed by the binding pocket residues spanning from 220 to 249 with the N-terminal region. The N-terminal residues of the MMP9 protein revealed a high association with the binding site residues of the reference ligands, such as Captopril and Solasodine. Comparing Captopril to the Solasodine ligand, the correlation magnitude was larger. The results showed that MMP9-containing natural compounds complexes and the reference ligand exhibited similar correlations near binding residues. In light of this, the DCCM displayed cooperative and anticooperative motion in the protein, indicating the conformational flexibility of the investigated complexes and stable connections mediated by noncooperative motion on the opposite side, which triggered the opening and shutting of the binding pocket residues and enabled the stable complex formation during the MD simulation.

4. DISCUSSION

The present study analyzed hypoxia, a critical microenvironmental condition of GBM, to identify potential biomarkers and establish treatment strategies for GBM treatment. In recent years, TME gained the attention of researchers as it regulates tumor growth and significantly influences treatment response. Hypoxia condition and immune cell infiltration in TME promote and antagonize tumor growth. Herein, we identify hypoxia-related molecular signatures involved in GBM pathogenesis. Based on the functional enrichment analysis, we have found 32 HUB signatures whose expressions were validated through microarray and RNA sequence data sets obtained from TCGA data sets of GBM patients. Indeed, we subjected 10 shortlisted molecular signatures to the RNA deconvolution-based TIMER analysis. From the gene expression profiles, TIMER employs an algorithm to determine the abundance of tumor-infiltrating immune cells. The proportion of cancer cells in the tumor tissue is described as tumor purity (also known as tumor cell fraction), which indicates the characteristics of TME. Recent studies have shown that tumor purity is linked to prognosis, mutation burden, and a robust immunological phenotype.[87,88] Our results indicate that LYN, MMP9, PSMB9, and TIMP1 were linked with the GBM microenvironment. Zhao et al. demonstrated a high expression of the PLOD family with negative tumor purity and high immune infiltration.[89] In our study, LYN was downregulated in the hypoxic condition in GBM. According to a study by Dai and Siemann, hypoxia has little to no impact on the expression of phosphorylated LYN.[90] However, the elevated MMP9 expression in hypoxic TME enhances DC infiltration and reduces the infiltration of cytotoxic T cells (CD8+ T cells).[91] In contrast, increased CD8+ T-cell infiltration had been linked to a better predictive factor for long-term survival in glioblastoma patients.[92] Additionally, PSMB8 and PSMB9 immunoproteasome subunits are overexpressed in melanoma cell lines, and their reduced expression is linked to a poor prognosis in nonsmall-cell lung carcinoma.[93] Herein, in this study, the reduced PSMB9 expression is linked to increased immune cell infiltration, with the exception of CD8+ T cells. Our findings are backed up by the fact that all members of the TIMP family had significantly higher levels of expression in GBM.[94] TIMP1 expression levels in hypoxic-GBM are exclusively correlated with DC infiltration and are inversely related to B cells and neutrophils. Consistent with our results, previous studies have also identified the four molecular signatures (LYN, TIMP1, MMP9, and PSMB9) as potential biomarkers associated with TME in GBM and other cancers.[95−97] Herein, we briefly discussed the relevant pathways mentioned above by starting with the ILK pathway known to promote cell growth, cell cycle progression, and increase VEGF expression by stimulating HIF-1 via a phosphatidylinositol 3-kinase (PI3K)−dependent activation.[98] Another significant pathway that is involved in the TME of GBM is the AP-1 transcription factor (dimeric in nature), which is made up of proteins from the Jun (c-Jun, JunB, and JunD) and Fos (c-Fos, FosB, Fra1, and Fra2) families. Studies have concluded that different triggers, such as inflammatory cytokines, stress inducers, or pathogens, activate the AP-1 transcription factor family, resulting in innate and adaptive immunities.[99] In addition, active CDC42 (ρ-GTPase) has been shown to facilitate glioma cell migration and invasion and regulate cell polarity.[100] In GBM, HIF-1 and VEGF upregulate CXCR4, which is significant for angiogenesis and cell invasion.[101] Furthermore, another fascinating study showed that the interaction of microglia and GBM through the LPA pathway has important consequences for tumor progression. A deeper understanding of this interaction could lead to the development of new therapeutic techniques that target LPA as a possible GBM target.[102] Another study found that hypoxic TME stimulates invadopodia development (actin-rich protrusions of the plasma membrane that focus ECM breakdown through the secretion of MMPs), which are essential for metastasis.[103] In addition, our data showed that the localization of MMP9 was mainly the extracellular region, and FOS, JUN, and TP53 were only significantly overexpressed associated TFs in GBM patient's samples. MMP9 was overexpressed in different subtypes of GBM including classical, mesenchymal, neural, and proneural (shown in Supporting Information Figure S7A). It also has the potential to act as a poor prognostic biomarker (HR > 1) as it shows significant disease-free survival (shown in Supporting Information Figure S7B). This all together suggests the significance of targeting TME. LYN and PSMB9 being downregulated in hypoxic condition, and due to unavailability of the reported drug against TIMP1, these biomarkers were not explored in the current study in identifying the novel drug. Hence, MMP9 was selected for identifying natural compounds as inhibitors in order to reduce GBM pathogenesis.

MMP9, a member of the gelatinase family of MMPs that degrades and remodels ECM proteins, plays a vital role in cell migration and EMT and angiogenesis.[104] Other TME components, such as nonmalignant stromal cells, neutrophils, macrophages, and endothelial cells, release MMP9 in the microenvironment. MMPs are known to be induced by HIF-1.[105,106] MMP inhibitors can diminish tumor cells' invasive and migratory abilities in cancer. MMP9 inhibitors were previously discovered using a computational technique, indicating that MMP9 is a targetable protein.[107,108] Based on previous studies, we have selected Captopril and Solasodine as reference drugs against MMP9. Captopril is an MMP2 inhibitor for treating patients on continuous ambulatory peritoneal dialysis therapy.[109] Captopril inhibits MMP2 and MMP9 via chelating zinc ions at the enzyme's active site. It also utilized alongside other medicines like Disulfiram and Nelfinavir as adjuvant therapy for GBM.[110] Moreover, it can inhibit MMP2 and MMP9, suspected of having a role in GBM metastasis and invasion, since it is an angiotensin-converting enzyme inhibitor, which belongs to a family of metalloproteinases comparable to MMPs.[111] Similarly, Solasodine has been reported to inhibit MMP9 and induce cell apoptosis, particularly in human lung cancer. However, this drug's pharmacokinetics, safety, and effectiveness in clinical practice remain unclear.[85,112]

During identifying new agents for MMP9, we explored six classes of natural compounds, including alkaloids, flavonoids, terpenoids, aliphatic compounds, aromatic compounds, and tannins. Previous studies have also supported that multiple natural compounds have antitumor and apoptotic effects in TMZ and p53 resistance GBM cells. Various natural compounds such as chrysin, epigallocatechin-3-gallate, hispidulin, rutin, and silibinin were also used in combination with TMZ and other chemotherapeutic drugs due to their potential to act as chemosensitizers (such as icariin and quercetin), radiosensitizers (*Zataria multiflora*), inhibits proliferation (such as *Zingiber officinale* and *Rhazya stricta*) and migration, and induces apoptosis (Baicalein).[16,113,114] However, these were checked for BBB permeability, druglikeness, and LIPINSKI rules of 5, and ADMET analysis was performed. We performed in silico molecular docking and MD simulations with MMP9 protein (alpha chain) using Autodock Vina 4.0 and GROMACS to evaluate the inhibitory effect of shortlisted drugs. Ramachandran plot of MMP9 (PDB identifier: 4HMA) is shown in Supporting Information Figure S7C. The binding affinity of ligands (drugs) was calculated and compared with reference drugs. In this instance, we have picked three best-docked compounds with binding energies comparable to Solasodine and better than Captopril for MD simulations. Stability should be taken into careful consideration during drug testing in addition to safety. The software's MD simulation module examined the stability of these MMP9-compound complexes in the natural environment. Further compounds interacted with targets with a minimum of at least 2 H-bond interactions. Numerous studies have been conducted in the past to implement molecular docking and MD simulations and MM-PBSA assessment to record drug transport variability, identify protein allosteric inhibition, consider the impact of chirality in selective enzyme inhibition, investigate the irreversible style of the receptors, and evaluate ligand−protein interactions. Similarly, this study examined the intermolecular contact stability of identified prospective lead compounds and standard molecules with their respective targets using classical MD simulation for 50 ns of MMP9 protein with ligands.[115] Subsequently, the efficacy of molecules' molecular interactions can be examined using structural analysis, such as RMSD and RMSF.[116] Results revealed that the binding energy of MMP9 with ligands 7,4′-dihydroxyflavan, (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran, and 4′-hydroxy-7-methoxyflavan was similar (−10 kcal/mol) to that of the reference drug Solasodine and better than Captopril. All three ligands, flavonoids in nature, interact within the binding domain of the MMP9 pocket, and this interaction was stable for 50 ns with less deviation and fluctuations. RMSD value difference between the backbone and complex was <3 Å. The MMP9-7,4′-dihydroxyflavan complex findings suggest that five residues, Leu222, Val223, Ala242, Met247, and Tyr248, contributed significantly to the formation of the stable complex. Most importantly, the residues Tyr248 showed significant contributions to the binding affinity by scoring the lowest contribution energy of −5.41 kJ/mol. MMP9-(3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran had a 94.16 kJ/mol determining binding affinity. Leu188, Leu222, Val223, His226, and Tyr248 residues also facilitated stable compound formation. Leu222 scored the highest binding affinity of −5.74 kJ/mol. Similarly, the binding energy of MMP9-4′-hydroxy-7-methoxyflavan was around 78.44 kJ/mol. The per-residue contribution energy also revealed that the formation of a stable complex was significantly influenced by six residues from the binding pocket: Leu188, Leu222, Val223, Leu243, Met247, and Tyr248. The binding affinity of the residue Met247 is −6.22 kJ/mol. Further, PCA analysis revealed that the MMP9-targeting ligands, 4′-dihydroxyflavan, (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran, and 4′-hydroxy-7-methoxyflavan had less diversity than the reference drug during the simulation run. Both reference drugs demonstrated increased conformational flexibility with the maximum number of diverse conformations. Interestingly, compared to the Captopril reference drug, the MMP9 inhibitors, 7,4′-dihydroxy-

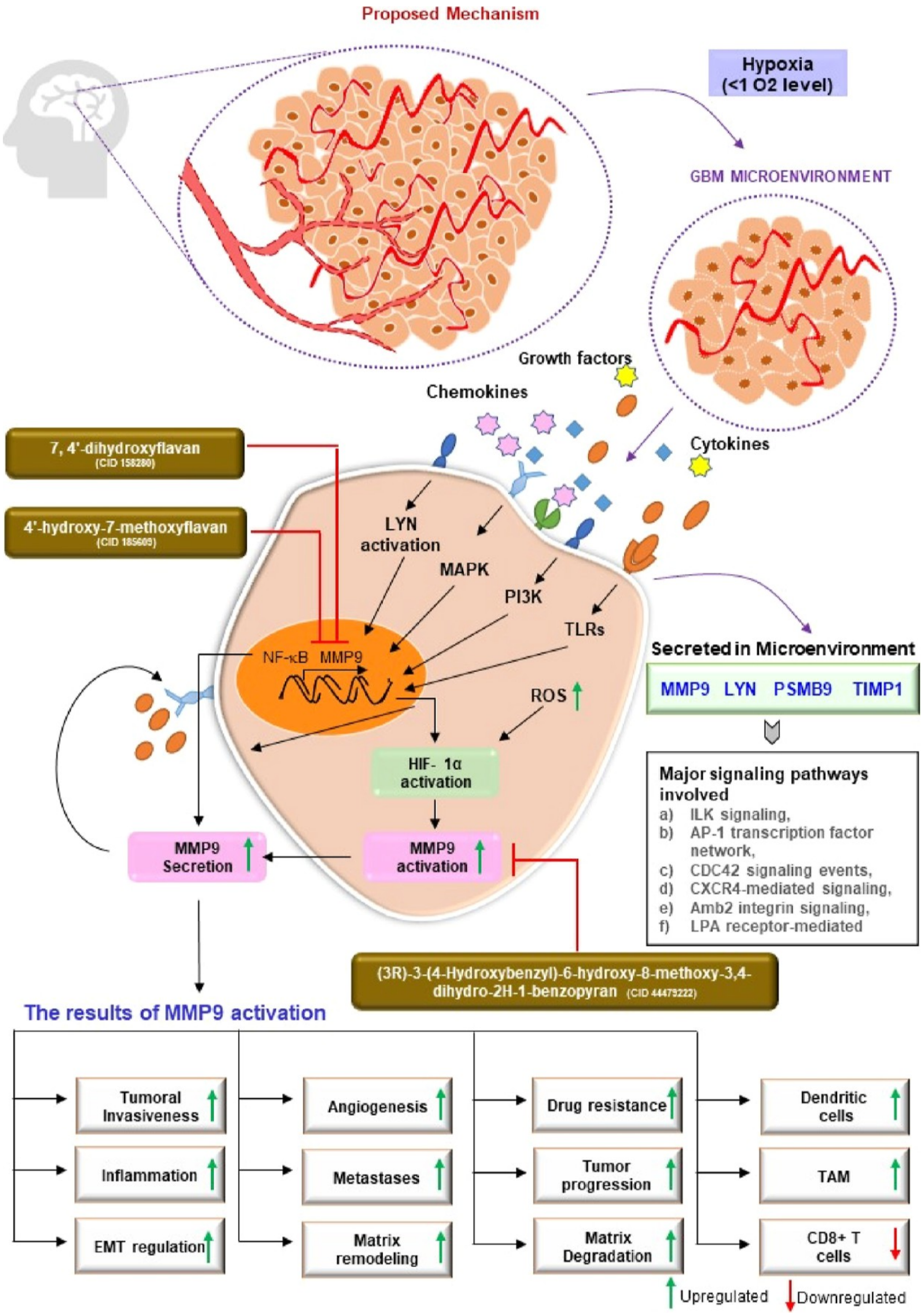


Figure 8. Potential of novel inhibitors 7,4′-dihydroxyflavan, (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran, and 4′-hydroxy-7-methoxyflavan in suppressing GBM pathogenesis by interacting with MMP9 protein produced in a hypoxic environment condition. MMP9 is synthesized de novo during stimulation induced with cytokines by activating various signaling pathways such as NF-*κ*B, HIF-1, MAPK, PI3K, etc. Cytokines (TNF-*α*, IL-8, and IL-1*β*) and growth factors (TGF-*β*, PDGF, and bFGF) bind to their receptors which regulate MMP9 activation and secretion. MMP9 is secreted by tumor cells, monocytes, inflammatory macrophages, and stromal cells in the extracellular environment. This affects various downstream biological processes, including matrix degradation, remodeling, EMT (enhanced tumoral invasion, metastases), angiogenesis, inflammation, drug resistance, etc. Novel inhibitors 7,4′-dihydroxyflavan, (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran, and 4′-hydroxy-7-methoxyflavan bind to MMP9 and suppress its activation and thus reduce the expression and regulation of downstream process involved in GBM pathogenesis in the above figure. Our approaches to GBM treatment are being reoriented by focusing on these features of MMPs.

flavan, (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran, and 4′-hydroxy-7-methoxyflavan, used significantly less conformational space. Contrarily, only 7,4′-dihydroxyflavan and (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran outperformed the Solasodine reference drug.

Furthermore, 7,4′-dihydroxyflavan, (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran, and 4′-hydroxy-7-methoxyflavan showed positive correlations with

https://doi.org/10.1021/acsomega.3c00441

the N-terminal domain of proteins, while (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran displayed an anticorrelation. As a result, we demonstrated how three lead flavonoids may be able to target the MMP9 protein. The fact that 7,4′-dihydroxyflavan was derived from the African forest tree *Guibourtia ehie* or Shedua, which has been utilized traditionally for tumor and wound healing, provided additional support for our findings in earlier investigations. It acts as a metabolite and shows anti-inflammatory and antioxidant effects in prostate cancer, breast cancer, and osteosarcoma by regulating Akt/Bad and MAPK signaling. In addition, (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran was found in *Soymida febrifuge* (Indian-redwood). Its fruits are therapeutic and have been used to treat cervical and colon cancer.[117] Interestingly, a study by Sowmyya and Vijaya Lakshmi discovered that extracts from these dried fruits contributed to the creation of silver nanoparticles by acting as reducing and stabilizing agents during the conversion of Ag^+ to nano-silver.[118] The last compound, 4′-hydroxy-7-methoxyflavan, was derived from the orchid tree *Bauhinia divaricate* and was formerly used to treat skin and colon cancer. These three flavonoids will inhibit MMP9 and lower its overexpression brought on by hypoxia in GBM. As a result of these inhibitions, the downstream effects of MMP9 activation will be diminished, which will minimize the pathogenesis of GBM. Cell proliferation, invasion, angiogenesis, drug resistance, matrix remodeling, and immune cell infiltration are significant pathways that will be impacted. The infiltration of DCs in response to MMP9 overexpression was also demonstrated by our data, which also indicated a positive correlation with immune checkpoints like PD-1 and TIM-3. Figure 8 illustrates the proposed mode of action for three novel flavonoids, including 7,4′-dihydroxyflavan (PubChem CID 158280), (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran (PubChem CID 44479222), and 4′-hydroxy-7-methoxyflavan (PubChem CID 185609). These will attenuate MMP9 activation's impact on GBM.

5. CONCLUSIONS AND FUTURE PERSPECTIVES

Despite recent advancements in chemotherapy, radiotherapy, and immunotherapy, there is currently no satisfactory therapy for GBM in clinics due to many reasons, being toxicity of chemotherapy, failure of the drug to cross BBB, involvement of TME, and less immune infiltration. For instance, immune checkpoint blockade targeting CD8+ T cells is ineffective for GBM.[119] There is an unmet need for novel approaches to treat GBM and other brain cancers. Here in our study, we have focused on a crucial TME parameter, that is, hypoxia caused due to intense cell respiration, excessive nutrient consumption by tumor cells, and abnormal vasculature. However, hypoxia is a hallmark of brain tumors, and if and how hypoxia affects antitumor immunity in the brain remains unclear. Our findings shed light on the potential of MMP9 as a therapeutic target and a robust biomarker in GBM's hypoxic microenvironment. In Figure 8, it is illustrated that in response to cytokine-induced stimulation, MMP9 is synthesized de novo by activating various signaling pathways including NF-κB, HIF-1, MAPK, PI3K, and so forth. Cytokines such as TNF-α, IL-8, and IL-1β and growth factors namely TGF-β, PDGF, and bFGF bind to their respective receptors and influence the activation and production of MMP9. This has an impact on a number of biological functions that come thereafter, such as drug resistance, remodeling of the matrix, EMT, increased tumoral invasion, metastases, angiogenesis, and remodeling.

Previous studies supported our results where researchers have shown that MMP9, a zinc-dependent endopeptidase, was upregulated in glioma tissues, and its expression was correlated with tumor grade and poor prognosis. Hypoxia condition increases the protein expression of HIF-α, MMP2, and MMP9 in cancer[120] and regulates tight junction rearrangement, leading to vascular leakage in the brain.[121] Majority of the ECM components are substrates of MMPs. MMP-9 can cleave many ECM proteins to regulate ECM remodeling and affects the alteration of cell–cell and cell–ECM interactions. It can also cleave many plasma surface proteins to release them from the cell surface. It has been implicated in the invasion and also implicated in BBB opening as part of the neuroinflammatory response, metastasis through proliferation, vasculogenesis, and angiogenesis.[72] MMP9 has been a potential biomarker for many cancers, including osteosarcoma, breast, cervical, ovarian, and pancreatic, giant cell tumor of bone, and non-small cell lung cancer.[21] Herein the current study, we have proposed MMP9 as a promising biomarker for hypoxic microenvironmental conditions in GBM. Other molecular signatures, such as LYN, PSMB9, and TIMP1, could be investigated further as druggable biomarkers or prognostic markers in addition to MMP9. Infiltration of immune cells such as neutrophils and DCs was linked to this gene's expression to varying degrees. This effect opens up new avenues for study into MMP9 and GBM. A negative correlation with B cells, CD4+ T cells, and CD8+ T cells supports the failure of current immune checkpoint inhibitors.

The current study used in silico techniques such as compound-protein-pathway enrichment analysis, network pharmacology, molecular docking, MD simulation, MM-PBSA, PCA, and DCCM investigations to identify a collection of druggable and nontoxic natural compounds from plants. The potential of natural compounds to be used as drugs was revealed by ADMET analysis of 11 novel hits. A chemical substance must have absorption, distribution, metabolism, excretion, and toxicity values to be utilized as a medication. The results obtained showed flavonoids named 7,4′-dihydroxyflavan, (3*R*)-3-(4-hydroxybenzyl)-6-hydroxy-8-methoxy-3,4-dihydro-2*H*-1-benzopyran, and 4′-hydroxy-7-methoxyflavan as potential inhibitors of MMP9 produced from the hypoxic condition in GBM. These inhibitors have comparable or better results compared to reference drugs Solasodine and Captopril. Our results indicate that MMP9 and drug interaction are stable, and proposed novel flavonoids can inhibit or reduce MMP9 expression in hypoxia conditions, which will further affect the downstream process involved in GBM pathogenesis. Hence, targeting an essential microenvironmental condition will improve therapeutic efficacy and expand the treatment drug library against GBM. Limiting to the present findings, we point out that the results presented in this work are based on processor simulations which need to be further validated with wet-lab experimental protocols.

In conclusion, the observations of this work suggest novel plant-based flavonoids inhibited the potential role of MMP9 as a biomarker factor and active MMP9 in GBM. Prior to synthesizing therapeutics, the results of this investigation could be helpful. Other natural compounds and plant-based natural compounds could be examined and studied to understand and explore whether they could be employed as future possibilities for GBM medicines. The results of this study are helpful for drug development. The findings may aid in the assisted screening of therapeutics for GBM. This study is novel in incorporating various computational methodologies for the virtual screening of natural compounds based on BBB, ADMET, PAINS, and Lipinski's rule. This study allows scientists to explore these molecules in vitro or in vivo as a medicinal approach. We have validated our results using different computational methodologies such as multiple-target validation, literature validation, TCGA databases (containing GBM samples data), cell culture, and animal model research which will fill in the gaps. We identified the common residues via which the inhibitor can potentially bind to the target using bioinformatics tools and in silico studies. However, the molecular mechanism underlying the reduction of target expression needs only to be validated through in vitro experiments. New leads are being discovered in several ongoing studies using advanced computational strategies and machine learning models to filter massive pharmaceutical libraries. The experimental screening strategy alone may not enhance lead productivity for the rapid development of viable medicines. Our findings will aid researchers in concentrating on TME components and their conditions in order to produce novel natural product-based anti-GBM therapies that address two major issues: toxicity and resistance and target of a major microenvironmental condition hypoxia.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c00441.

> MA plot; list of DEGs; mutational and correlation analyses; localization and transcription factor analysis; screening of natural compounds; physiochemical properties; 2D docked structure; H-bond and contribution energy plot; and MMP9 characteristics (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Pravir Kumar − *Molecular Neuroscience and Functional Genomics Laboratory, Department of Biotechnology, Delhi Technological University (Formerly DCE), Delhi 110042, India;* orcid.org/0000-0001-7444-2344; Phone: +91-9818898622; Email: pravirkumar@dtu.ac.in

Author

Smita Kumari − *Molecular Neuroscience and Functional Genomics Laboratory, Department of Biotechnology, Delhi Technological University (Formerly DCE), Delhi 110042, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c00441

Author Contributions

P.K. and S.K. conceived and designed the manuscript. S.K. collected, analyzed, and critically evaluated these data. S.K. prepared the figures and tables. P.K. and S.K. analyzed the entire data and wrote the manuscript.

Funding

S.K. has received Senior Research Fellowship (SRF) from the Department of Biotechnology (DBT), Govt. of India (Fellow ID: DBT/2019/DTU/1308).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We would like to thank the senior management of Delhi Technological University (DTU) and the Department of Biotechnology (DBT), Government of India, for their constant support and financial assistance.

■ LIST OF ABBREVIATIONS

2D, 2-dimensional; 3D, 3-dimensional; Caco-2, colon adenocarcinoma cell lines; DCCM, domain cross-correlation matrix; DEGs, differentially regulated genes; DCs, dendritic cells; EPO, erythropoietin; GBM, glioblastoma multiforme; GEO, gene expression omnibus; GEPIA2.0, Gene Expression Profiling Interactive Analysis; GMQE, Global Model Quality Estimate; GS2D, gene set to diseases; HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; H-bond, hydrogen bond; HIA, human intestinal absorption cells; HR, hazard ratio; KEGG, Kyoto Encyclopedia of Genes and Genomes; logCPM, $\log_2$-counts-per-millions; LYN, Lck/Yes-related novel protein tyrosine kinase; MB-PBSA, molecular mechanics Poisson−Boltzmann surface area; MD, molecular dynamics; MDCK, Madin−Darby canine kidney cells; MMP-9, matrix metalloproteinase 9; MW, molecular weight; PAINS, Pan Assay Interference Compounds; PCA, principal component analysis; PDGF, platelet-derived growth factor; PPI, protein−protein interaction; PSMB9, proteasome 20S subunit beta 9; RCSB, Research Collaboratory for Structural Bioinformatics; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation; R_g, radiation of gyration; SASA, solvent accessible surface area; SMILES, Simplified Molecular-Input Line-Entry System; STRING, Search Tool for the Retrieval of Interacting Genes/Proteins; TFs, transcription factors; TIMP1, tissue inhibitor of metalloproteinases 1; TME, tumor microenvironment; TPM, transcript per million; vdw, van der Waal force; VEGF, vascular endothelial growth factor

■ REFERENCES

(1) Miller, K. D.; Ostrom, Q. T.; Kruchko, C.; Patil, N.; Tihan, T.; Cioffi, G.; Fuchs, H. E.; Waite, K. A.; Jemal, A.; Siegel, R. L.; Barnholtz-Sloan, J. S. Brain and Other Central Nervous System Tumor Statistics, 2021. *Ca -Cancer J. Clin.* **2021**, *71*, 381−406.

(2) Stylli, S. S. Novel Treatment Strategies for Glioblastoma. *Cancers* **2020**, *12*, 2883.

(3) DeCordova, S.; Shastri, A.; Tsolaki, A. G.; Yasmin, H.; Klein, L.; Singh, S. K.; Kishore, U. Molecular Heterogeneity and Immunosuppressive Microenvironment in Glioblastoma. *Front. Immunol.* **2020**, *11*, 1402.

(4) Li, Y.; Zhao, L.; Li, X. F. Hypoxia and the Tumor Microenvironment. *Technol. Cancer Res. Treat.* **2021**, *20*, 153303382110363.

(5) Huang, W. J.; Chen, W. W.; Zhang, X. Glioblastoma Multiforme: Effect of Hypoxia and Hypoxia Inducible Factors on Therapeutic Approaches (Review). *Oncol. Lett.* **2016**, *12*, 2283−2288.

(6) Velásquez, C.; Mansouri, S.; Gutiérrez, O.; Mamatjan, Y.; Mollinedo, P.; Karimi, S.; Singh, O.; Terán, N.; Martino, J.; Zadeh, G.; Fernández-Luna, J. L. Hypoxia Can Induce Migration of Glioblastoma Cells through a Methylation-Dependent Control of ODZ1 Gene Expression. *Front. Oncol.* **2019**, *9*, 1036.

(7) Emami Nejad, A.; Najafgholian, S.; Rostami, A.; Sistani, A.; Shojaeifar, S.; Esparvarinha, M.; Nedaeinia, R.; Haghjooy Javanmard, S.; Taherian, M.; Ahmadlou, M.; Salehi, R.; Sadeghi, B.; Manian, M. The Role of Hypoxia in the Tumor Microenvironment and Development of Cancer Stem Cell: A Novel Approach to Developing Treatment. *Cancer Cell Int.* **2021**, *21*, 62.

(8) Zheng, X.; Qian, Y.; Fu, B.; Jiao, D.; Jiang, Y.; Chen, P.; Shen, Y.; Zhang, H.; Sun, R.; Tian, Z.; Wei, H. Mitochondrial Fragmentation Limits NK Cell-Based Tumor Immunosurveillance. *Nat. Immunol.* **2019**, *20*, 1656−1667.

(9) Henze, A. T.; Mazzone, M. The Impact of Hypoxia on Tumor-Associated Macrophages. *J. Clin. Invest.* **2016**, *126*, 3672.

(10) Park, J. H.; Kim, H. J.; Kim, C. W.; Kim, H. C.; Jung, Y.; Lee, H. S.; Lee, Y.; Ju, Y. S.; Oh, J. E.; Park, S. H.; Lee, J. H.; Lee, S. K.; Lee, H. K. Tumor Hypoxia Represses Γδ T Cell-Mediated Antitumor Immunity against Brain Tumors. *Nat. Immunol.* **2021**, *22*, 336−346.

(11) Bronisz, A.; Salińska, E.; Chiocca, E. A.; Godlewski, J. Hypoxic Roadmap of Glioblastoma-Learning about Directions and Distances in the Brain Tumor Environment. *Cancers* **2020**, *12*, 1213.

(12) Kalkan, R. Hypoxia Is the Driving Force Behind GBM and Could Be a New Tool in GBM Treatment. *Crit. Rev. Eukaryot. Gene Expr.* **2015**, *25*, 363−369.

(13) Tan, A. C.; Ashley, D. M.; López, G. Y.; Malinzak, M.; Friedman, H. S.; Khasraw, M. Management of Glioblastoma: State of the Art and Future Directions. *Ca -Cancer J. Clin.* **2020**, *70*, 299−312.

(14) Atanasov, A. G.; Zotchev, S. B.; Dirsch, V. M.; Orhan, I. E.; Supuran, M.; Rollinger, J. M.; Barreca, D.; Weckwerth, W.; Bauer, R.; Bayer, E. A.; Majeed, M.; Bishayee, A.; Bochkov, V.; Bonn, G. K.; Braidy, N.; Bucar, F.; Cifuentes, A.; D'Onofrio, G.; Bodkin, M.; Diederich, M.; Dinkova-Kostova, A. T.; Efferth, T.; El Bairi, K.; Arkells, N.; Fan, T. P.; Fiebich, B. L.; Freissmuth, M.; Georgiev, M. I.; Gibbons, S.; Godfrey, K. M.; Gruber, C. W.; Heer, J.; Huber, L. A.; Ibanez, E.; Kijjoa, A.; Kiss, A. K.; Lu, A.; Macias, F. A.; Miller, M. J. S.; Mocan, A.; Müller, R.; Nicoletti, F.; Perry, G.; Pittalà, V.; Rastrelli, L.; Ristow, M.; Russo, G. L.; Silva, A. S.; Schuster, D.; Sheridan, H.; Skalicka-Woźniak, K.; Skaltsounis, L.; Sobarzo-Sánchez, E.; Bredt, D. S.; Stuppner, H.; Sureda, A.; Tzvetkov, N. T.; Vacca, R. A.; Aggarwal, B. B.; Battino, M.; Giampieri, F.; Wink, M.; Wolfender, J. L.; Xiao, J.; Yeung, A. W. K.; Lizard, G.; Popp, M. A.; Heinrich, M.; Berindan-Neagoe, I.; Stadler, M.; Daglia, M.; Verpoorte, R.; Supuran, C. T. Natural Products in Drug Discovery: Advances and Opportunities. *Nat. Rev. Drug Discovery* **2021**, *20*, 200−216.

(15) Huang, M.; Lu, J. J.; Ding, J. Natural Products in Cancer Therapy: Past, Present and Future. *Nat. Prod. Bioprospect.* **2021**, *11*, 5−13.

(16) Vengoji, R.; Macha, M. A.; Batra, S. K.; Shonka, N. A. Natural Products: A Hope for Glioblastoma Patients. *Oncotarget* **2018**, *9*, 22194.

(17) Santos, B. L.; Oliveira, M. N.; Coelho, P. L. C.; Pitanga, B. P. S.; da Silva, A. B.; Adelita, T.; Silva, V. D. A.; Costa, M. D. F. D.; El-Bachá, R. S.; Tardy, M.; Chneiweiss, H.; Junier, M. P.; Moura-Neto, V.; Costa, S. L. Flavonoids Suppress Human Glioblastoma Cell Growth by Inhibiting Cell Metabolism, Migration, and by Regulating Extracellular Matrix Proteins and Metalloproteinases Expression. *Chem. Biol. Interact.* **2015**, *242*, 123−138.

(18) Soukhtanloo, M.; Mohtashami, E.; Maghrouni, A.; Mollazadeh, H.; Mousavi, S. H.; Roshan, M. K.; Tabatabaeizadeh, S. A.; Hosseini, A.; Vahedi, M. M.; Jalili-Nik, M.; Afshari, A. R. Natural Products as Promising Targets in Glioblastoma Multiforme: A Focus on NF-KB Signaling Pathway. *Pharmacol. Rep.* **2020**, *72*, 285−295.

(19) Zhai, K.; Siddiqui, M.; Abdellatif, B.; Liskova, A.; Kubatka, P.; Büsselberg, D. Natural Compounds in Glioblastoma Therapy: Preclinical Insights, Mechanistic Pathways, and Outlook. *Cancers* **2021**, *13*, 2317.

(20) Pujada, A.; Walter, L.; Patel, A.; Bui, T. A.; Zhang, Z.; Zhang, Y.; Denning, T. L.; Garg, P. Matrix Metalloproteinase MMP9 Maintains Epithelial Barrier Function and Preserves Mucosal Lining in Colitis Associated Cancer. *Oncotarget* **2017**, *8*, 94650.

(21) Huang, H. Matrix Metalloproteinase-9 (MMP-9) as a Cancer Biomarker and MMP-9 Biosensors: Recent Advances. *Sensors* **2018**, *18*, 3249.

(22) Augoff, K.; Hryniewicz-Jankowska, A.; Tabola, R.; Stach, K. MMP9: A Tough Target for Targeted Therapy for Cancer. *Cancers* **2022**, *14*, 1847.

(23) Atiq, A.; Parhar, I. Anti-Neoplastic Potential of Flavonoids and Polysaccharide Phytochemicals in Glioblastoma. *Molecules* **2020**, *25*, 4895.

(24) Mondal, A.; Gandhi, A.; Fimognari, C.; Atanasov, A. G.; Bishayee, A. Alkaloids for Cancer Prevention and Therapy: Current Progress and Future Perspectives. *Eur. J. Pharmacol.* **2019**, *858*, 172472.

(25) Edgar, R.; Domrachev, M.; Lash, A. E. Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository. *Nucleic Acids Res.* **2002**, *30*, 207−210.

(26) Mahi, N. A.; Najafabadi, M. F.; Pilarczyk, M.; Kouril, M.; Medvedovic, M. GREIN: An Interactive Web Platform for Re-Analyzing GEO RNA-Seq Data. *Sci. Rep.* **2019**, *9*, 7580.

(27) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. Bioinformatics Enrichment Tools: Paths toward the Comprehensive Functional Analysis of Large Gene Lists. *Nucleic Acids Res.* **2009**, *37*, 1−13.

(28) Fontaine, J. F.; Andrade-Navarro, M. A. Gene Set to Diseases (GS2D): Disease Enrichment Analysis on Human Gene Sets with Literature Data. *Genomics Comput. Biol.* **2016**, *2*, No. e33.

(29) Kuleshov, M. V.; Jones, M. R.; Rouillard, A. D.; Fernandez, N. F.; Duan, Q.; Wang, Z.; Koplev, S.; Jenkins, S. L.; Jagodnik, K. M.; Lachmann, A.; McDermott, M. G.; Monteiro, C. D.; Gundersen, G. W.; Ma'ayan, A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **2016**, *44*, W90−W97.

(30) Chen, E. Y.; Tan, C. M.; Kou, Y.; Duan, Q.; Wang, Z.; Meirelles, G. V.; Clark, N. R.; Ma'ayan, A. Enrichr: Interactive and Collaborative HTML5 Gene List Enrichment Analysis Tool. *BMC Bioinf.* **2013**, *14*, 128.

(31) Pathan, M.; Keerthikumar, S.; Ang, C. S.; Gangoda, L.; Quek, C. Y. J.; Williamson, N. A.; Mouradov, D.; Sieber, O. M.; Simpson, R. J.; Salim, A.; Bacic, A.; Hill, A. F.; Stroud, D. A.; Ryan, M. T.; Agbinya, J. I.; Mariadason, J. M.; Burgess, A. W.; Mathivanan, S. FunRich: An Open Access Standalone Functional Enrichment and Interaction Network Analysis Tool. *Proteomics* **2015**, *15*, 2597−2601.

(32) Szklarczyk, D.; Gable, A. L.; Nastou, K. C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N. T.; Legeay, M.; Fang, T.; Bork, P.; Jensen, L. J.; von Mering, C. The STRING Database in 2021: Customizable Protein-Protein Networks, and Functional Characterization of User-Uploaded Gene/Measurement Sets. *Nucleic Acids Res.* **2021**, *49*, D605−D612.

(33) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498−2504.

(34) Li, T.; Fu, J.; Zeng, Z.; Cohen, D.; Li, J.; Chen, Q.; Li, B.; Liu, X. S. TIMER2.0 for Analysis of Tumor-Infiltrating Immune Cells. *Nucleic Acids Res.* **2020**, *48*, W509−W514.

(35) Bowman, R. L.; Wang, Q.; Carro, A.; Verhaak, R. G. W.; Squatrito, M. GlioVis Data Portal for Visualization and Analysis of Brain Tumor Expression Data sets. *Neuro Oncol.* **2017**, *19*, 139−141.

(36) Tang, Z.; Kang, B.; Li, C.; Chen, T.; Zhang, Z. GEPIA2: An Enhanced Web Server for Large-Scale Expression Profiling and Interactive Analysis. *Nucleic Acids Res.* **2019**, *47*, W556−W560.

(37) Gill, B. J.; Pisapia, D. J.; Malone, H. R.; Goldstein, H.; Lei, L.; Sonabend, A.; Yun, J.; Samanamud, J.; Sims, J. S.; Banu, M.; Dovas, A.; Teich, A. F.; Sheth, S. A.; McKhann, G. M.; Sisti, M. B.; Bruce, J. N.; Sims, P. A.; Canoll, P. MRI-Localized Biopsies Reveal Subtype-Specific Differences in Molecular and Cellular Composition at the Margins of Glioblastoma. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 12550−12555.

(38) Madhavan, S.; Zenklusen, J. C.; Kotliarov, Y.; Sahni, H.; Fine, H. A.; Buetow, K. Rembrandt: Helping Personalized Medicine Become a Reality through Integrative Translational Research. *Mol. Cancer Res.* **2009**, *7*, 157−167.

(39) Gravendeel, L. A. M.; Kouwenhoven, M. C. M.; Gevaert, O.; de Rooi, J. J.; Stubbs, A. P.; Duijm, J. E.; Daemen, A.; Bleeker, F. E.; Bralten, L. B. C.; Kloosterhof, N. K.; De Moor, B.; Eilers, P. H. C.; van der Spek, P. J.; Kros, J. M.; Sillevis Smitt, P. A. E.; van den Bent, M. J.; French, P. J. Intrinsic Gene Expression Profiles of Gliomas Are a Better Predictor of Survival than Histology. *Cancer Res.* **2009**, *69*, 9065−9072.

(40) Li, T.; Fan, J.; Wang, B.; Traugh, N.; Chen, Q.; Liu, J. S.; Li, B.; Liu, X. S. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* **2017**, *77*, e108−e110.

(41) Yu, C. S.; Chen, Y. C.; Lu, C. H.; Hwang, J. K. Prediction of Protein Subcellular Localization. *Proteins: Struct., Funct., Bioinf.* **2006**, *64*, 643−651.

(42) Fornes, O.; Castro-Mondragon, J. A.; Khan, A.; van der Lee, R.; Zhang, X.; Richmond, P. A.; Modi, B. P.; Correard, S.; Gheorghe, M.; Baranašić, D.; Santana-Garcia, W.; Tan, G.; Chèneby, J.; Ballester, B.; Parcy, F.; Sandelin, A.; Lenhard, B.; Wasserman, W. W.; Mathelier, A. JASPAR 2020: Update of the Open-Access Database of Transcription Factor Binding Profiles. *Nucleic Acids Res.* **2020**, *48*, D87.

(43) Zhou, G.; Soufan, O.; Ewald, J.; Hancock, R. E. W.; Basu, N.; Xia, J. NetworkAnalyst 3.0: A Visual Analytics Platform for Comprehensive Gene Expression Profiling and Meta-Analysis. *Nucleic Acids Res.* **2019**, *47*, W234−W241.

(44) Mangal, M.; Sagar, P.; Singh, H.; Raghava, G. P. S.; Agarwal, S. M. NPACT: Naturally Occurring Plant-Based Anti-Cancer Compound-Activity-Target Database. *Nucleic Acids Res.* **2013**, *41*, D1124.

(45) Angeli, E.; Nguyen, T. T.; Janin, A.; Bousquet, G. How to Make Anticancer Drugs Cross the Blood-Brain Barrier to Treat Brain Metastases. *Int. J. Mol. Sci.* **2019**, *21*, 22.

(46) Daina, A.; Michielin, O.; Zoete, V. SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal Chemistry Friendliness of Small Molecules. *Sci. Rep.* **2017**, *7*, 42717.

(47) Liu, H.; Wang, L.; Lv, M.; Pei, R.; Li, P.; Pei, Z.; Wang, Y.; Su, W.; Xie, X. Q. AlzPlatform: An Alzheimer's Disease Domain-Specific Chemogenomics Knowledgebase for Polypharmacology and Target Identification Research. *J. Chem. Inf. Model.* **2014**, *54*, 1050−1060.

(48) Molsoft L. L. C. Drug-Likeness and molecular property prediction. https://molsoft.com/mprop/ (accessed Dec 26, 2021).

(49) Xiong, G.; Wu, Z.; Yi, J.; Fu, L.; Yang, Z.; Hsieh, C.; Yin, M.; Zeng, X.; Wu, C.; Lu, A.; Chen, X.; Hou, T.; Cao, D. ADMETlab 2.0: An Integrated Online Platform for Accurate and Comprehensive Predictions of ADMET Properties. *Nucleic Acids Res.* **2021**, *49*, W5−W14.

(50) BIOVIA Discovery Studio Visualizer-Dassault Systèmes. https://discover.3ds.com/discovery-studio-visualizer-download (accessed Sep 11, 2022). Free Download.

(51) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.

(52) Anderson, R. J.; Weng, Z.; Campbell, R. K.; Jiang, X. Main-Chain Conformational Tendencies of Amino Acids. *Proteins* **2005**, *60*, 679−689.

(53) Colovos, C.; Yeates, T. O. Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions. *Protein Sci.* **1993**, *2*, 1511−1519.

(54) Bowie, J. U.; Lüthy, R.; Eisenberg, D. A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Structure. *Science* **1991**, *253*, 164−170.

(55) Samdani, A.; Vetrivel, U. POAP: A GNU Parallel Based Multithreaded Pipeline of Open Babel and AutoDock Suite for Boosted High Throughput Virtual Screening. *Comput. Biol. Chem.* **2018**, *74*, 39−48.

(56) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(57) Kumari, R.; Kumar, R.; Lynn, A. G_mmpbsa–a GROMACS Tool for High-Throughput MM-PBSA Calculations. *J. Chem. Inf. Model.* **2014**, *54*, 1951−1962.

(58) Bhandare, V. V.; Kumbhar, B. V.; Kunwar, A. Differential Binding Affinity of Tau Repeat Region R2 with Neuronal-Specific β-Tubulin Isotypes. *Sci. Rep.* **2019**, *9*, 10795.

(59) Dwivedi, P. S. R.; Patil, V. S.; Khanal, P.; Bhandare, V. V.; Gurav, S.; Harish, D. R.; Patil, B. M.; Roy, S. System Biology-Based Investigation of Silymarin to Trace Hepatoprotective Effect. *Comput. Biol. Med.* **2022**, *142*, 105223.

(60) Taidi, L.; Maurady, A.; Britel, M. R. Molecular Docking Study and Molecular Dynamic Simulation of Human Cyclooxygenase-2 (COX-2) with Selected Eutypoids. *J. Biomol. Struct. Dyn.* **2020**, *40*, 1189−1204.

(61) Khanal, P.; Patil, V. S.; Bhandare, V. V.; Dwivedi, P. S. R.; Shastry, C. S.; Patil, B. M.; Gurav, S. S.; Harish, D. R.; Roy, S. Computational Investigation of Benzalacetophenone Derivatives against SARS-CoV-2 as Potential Multi-Target Bioactive Compounds. *Comput. Biol. Med.* **2022**, *146*, 105668.

(62) Bhandare, V. V.; Ramaswamy, A. The Proteinopathy of D169G and K263E Mutants at the RNA Recognition Motif (RRM) Domain of Tar DNA-Binding Protein (Tdp43) Causing Neurological Disorders: A Computational Study. *J. Biomol. Struct. Dyn.* **2018**, *36*, 1075−1093.

(63) Arnold, G. E.; Ornstein, R. L. Molecular Dynamics Study of Time-Correlated Protein Domain Motions and Molecular Flexibility: Cytochrome P450BM-3. *Biophys. J.* **1997**, *73*, 1147−1159.

(64) Khanal, P.; Zargari, F.; Far, B. F.; Kumar, D.; R, M.; Mahdi, Y. K.; Jubair, N. K.; Saraf, S. K.; Bansal, P.; Singh, R.; Selvaraja, M.; Dey, Y. N. Integration of System Biology Tools to Investigate Huperzine A as an Anti-Alzheimer Agent. *Front. Pharmacol.* **2021**, *12*, 785964.

(65) Mahi, N. A.; Najafabadi, M. F.; Pilarczyk, M.; Kouril, M.; Medvedovic, M. GREIN: An Interactive Web Platform for Reanalyzing GEO RNA-Seq Data. *Sci. Rep.* **2019**, *9*, 7580.

(66) Venny 2.1.0. https://bioinfogp.cnb.csic.es/tools/venny/index.html (accessed Nov 29, 2021).

(67) Vassilakopoulou, M.; Won, M.; Curran, W. J.; Souhami, L.; Prados, M. D.; Langer, C. J.; Rimm, D. L.; Hanna, J. A.; Neumeister, V. M.; Melian, E.; Diaz, A. Z.; Atkins, J. N.; Komarnicky, L. T.; Schultz, C. J.; Howard, S. P.; Zhang, P.; Dicker, A. P.; Knisely, J. P. S. BRCA1 Protein Expression Predicts Survival in Glioblastoma Patients from an NRG Oncology RTOG Cohort. *Oncology* **2021**, *99*, 580.

(68) Zhang, Y.; Xia, Q.; Lin, J. Identification of the Potential Oncogenes in Glioblastoma Based on Bioinformatic Analysis and Elucidation of the Underlying Mechanisms. *Oncol. Rep.* **2018**, *40*, 715−725.

(69) Yang, G.; Dong, K.; Zhang, Z.; Zhang, E.; Liang, B.; Chen, X.; Huang, Z. EXO1 Plays a Carcinogenic Role in Hepatocellular Carcinoma and Is Related to the Regulation of FOXP3. *J. Cancer* **2020**, *11*, 4917−4932.

(70) Liu, B.; Zhang, G.; Cui, S.; Du, G. Upregulation of KIF11 in TP53 Mutant Glioma Promotes Tumor Stemness and Drug Resistance. *Cell. Mol. Neurobiol.* **2022**, *42*, 1477−1485.

(71) Jiang, C.; Zhang, H.; Wu, W.; Wang, Z.; Dai, Z.; Zhang, L.; Liu, Z.; Cheng, Q. Immune Characteristics of LYN in Tumor Microenvironment of Gliomas. *Front. Cell Dev. Biol.* **2022**, *9*, 760929.

(72) Xue, Q.; Cao, C.; Chen, X. Y.; Zhao, J.; Gao, L.; Li, S. Z.; Fei, Z. High Expression of MMP9 in Glioma Affects Cell Proliferation and Is Associated with Patient Survival Rates. *Oncol. Lett.* **2017**, *13*, 1325.

(73) Liu, J.; Yang, X.; Ji, Q.; Yang, L.; Li, J.; Long, X.; Ye, M.; Huang, K.; Zhu, X. Immune Characteristics and Prognosis Analysis of the Proteasome 20S Subunit Beta 9 in Lower-Grade Gliomas. *Front. Oncol.* **2022**, *12*, 875131.

(74) Smith, S. J.; Li, C. M.; Lingeman, R. G.; Hickey, R. J.; Liu, Y.; Malkas, L. H.; Raoof, M. Molecular Targeting of Cancer-Associated PCNA Interactions in Pancreatic Ductal Adenocarcinoma Using a Cell-Penetrating Peptide. *Mol. Ther. Oncolytics* **2020**, *17*, 250−256.

(75) Aaberg-Jessen, C.; Fogh, L.; Sørensen, M. D.; Halle, B.; Brünner, N.; Kristensen, B. W. Overexpression of TIMP-1 and Sensitivity to Topoisomerase Inhibitors in Glioblastoma Cell Lines. *Pathol. Oncol. Res.* **2019**, *25*, 59−69.

(76) Wang, B. Q.; Zhang, C. M.; Gao, W.; Wang, X. F.; Zhang, H. L.; Yang, P. C. Cancer-Derived Matrix Metalloproteinase-9 Contributes to Tumor Tolerance. *J. Cancer Res. Clin. Oncol.* **2011**, *137*, 1525−1533.

(77) Juric, V.; O'Sullivan, C.; Stefanutti, E.; Kovalenko, M.; Greenstein, A.; Barry-Hamilton, V.; Mikaelian, I.; Degenhardt, J.; Yue, P.; Smith, V.; Mikels-Vigdal, A. MMP-9 Inhibition Promotes Anti-Tumor Immunity through Disruption of Biochemical and Physical Barriers to T-Cell Trafficking to Tumors. *PLoS One* **2018**, *13*, No. e0207255.

(78) van Tellingen, O.; Yetkin-Arik, B.; de Gooijer, M. C.; Wesseling, P.; Wurdinger, T.; de Vries, H. E. Overcoming the Blood-Brain Tumor Barrier for Effective Glioblastoma Treatment. *Drug Resist. Updates* **2015**, *19*, 1−12.

(79) Ursu, O.; Rayan, A.; Goldblum, A.; Oprea, T. I. Understanding Drug-Likeness. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 760−781.

(80) Yang, Z. Y.; Yang, Z. J.; He, J. H.; Lu, A. P.; Liu, S.; Hou, T. J.; Cao, D. S. Benchmarking the Mechanisms of Frequent Hitters: Limitation of PAINS Alerts. *Drug Discov. Today* **2021**, *26*, 1353−1358.

(81) Guan, L.; Yang, H.; Cai, Y.; Sun, L.; Di, P.; Li, W.; Liu, G.; Tang, Y. ADMET-Score − a Comprehensive Scoring Function for Evaluation of Chemical Drug-Likeness. *MedChemComm* **2019**, *10*, 148.

(82) Zhu, R.; Hu, L.; Li, H.; Su, J.; Cao, Z.; Zhang, W. Novel Natural Inhibitors of CYP1A2 Identified by in Silico and in Vitro Screening. *Int. J. Mol. Sci.* **2011**, *12*, 3250.

(83) Durán-Iturbide, N. A.; Díaz-Eufracio, B. I.; Medina-Franco, J. L. In Silico ADME/Tox Profiling of Natural Products: A Focus on BIOFACQUIM. *ACS Omega* **2020**, *5*, 16076−16084.

(84) Lei, T.; Chen, F.; Liu, H.; Sun, H.; Kang, Y.; Li, D.; Li, Y.; Hou, T. ADMET Evaluation in Drug Discovery. Part 17: Development of Quantitative and Qualitative Prediction Models for Chemical-Induced Respiratory Toxicity. *Mol. Pharm.* **2017**, *14*, 2407−2421.

(85) Shen, K. H.; Hung, J. H.; Chang, C. W.; Weng, Y. T.; Wu, M. J.; Chen, P. S. Solasodine Inhibits Invasion of Human Lung Cancer Cell through Downregulation of MiR-21 and MMPs Expression. *Chem. Biol. Interact.* **2017**, *268*, 129−135.

(86) Liu, N.; Wang, X.; Wu, H.; Lv, X.; Xie, H.; Guo, Z.; Wang, J.; Dou, G.; Zhang, C.; Sun, M. Computational Study of Effective Matrix Metalloproteinase 9 (MMP9) Targeting Natural Inhibitors. *Aging* **2021**, *13*, 22867−22882.

(87) Mao, Y.; Feng, Q.; Zheng, P.; Yang, L.; Liu, T.; Xu, Y.; Zhu, D.; Chang, W.; Ji, M.; Ren, L.; Wei, Y.; He, G.; Xu, J. Low Tumor Purity Is Associated with Poor Prognosis, Heavy Mutation Burden, and Intense Immune Phenotype in Colon Cancer. *Cancer Manage. Res.* **2018**, *10*, 3569.

(88) Gong, Z.; Zhang, J.; Guo, W. Tumor Purity as a Prognosis and Immunotherapy Relevant Feature in Gastric Cancer. *Cancer Med.* **2020**, *9*, 9052−9063.

(89) Zhao, Y.; Zhang, X.; Yao, J. Comprehensive Analysis of PLOD Family Members in Low-Grade Gliomas Using Bioinformatics Methods. *PLoS One* **2021**, *16*, No. e0246097.

(90) Dai, Y.; Siemann, D. C-Src Is Required for Hypoxia-Induced Metastasis-Associated Functions in Prostate Cancer Cells. *OncoTargets Ther.* **2019**, *12*, 3519.

(91) Baek, J.-H.; Birchmeier, C.; Zenke, M.; Hieronymus, T. The HGF Receptor/Met Tyrosine Kinase Is a Key Regulator of Dendritic Cell Migration in Skin Immunity. *J. Immunol.* **2012**, *189*, 1699−1707.

(92) Yang, I.; Tihan, T.; Han, S. J.; Wrensch, M. R.; Wiencke, J.; Sughrue, M. E.; Parsa, A. T. CD8+ T-Cell Infiltrate in Newly Diagnosed Glioblastoma Is Associated with Long-Term Survival. *J. Clin. Neurosci.* **2010**, *17*, 1381.

(93) Kalaora, S.; Lee, J. S.; Barnea, E.; Levy, R.; Greenberg, P.; Alon, M.; Yagel, G.; Bar Eli, G.; Oren, R.; Peri, A.; Patkar, S.; Bitton, L.; Rosenberg, S. A.; Lotem, M.; Levin, Y.; Admon, A.; Ruppin, E.; Samuels, Y. Immunoproteasome Expression Is Associated with Better Prognosis and Response to Checkpoint Therapies in Melanoma. *Nat. Commun.* **2020**, *11*, 896.

(94) Han, J.; Jing, Y.; Han, F.; Sun, P. Comprehensive Analysis of Expression, Prognosis and Immune Infiltration for TIMPs in Glioblastoma. *BMC Neurol.* **2021**, *21*, 447.

(95) Xu, B. Prediction and Analysis of Hub Genes between Glioblastoma and Low-Grade Glioma Using Bioinformatics Analysis. *Medicine* **2021**, *100*, No. e23513.

(96) Tornillo, G.; Knowlson, C.; Kendrick, H.; Cooke, J.; Mirza, H.; Aurrekoetxea-Rodríguez, I.; Vivanco, M. d. M.; Buckley, N. E.; Grigoriadis, A.; Smalley, M. J. Dual Mechanisms of LYN Kinase Dysregulation Drive Aggressive Behavior in Breast Cancer Cells. *Cell Rep.* **2018**, *25*, 3674−3692.

(97) Liu, H.; Chen, D.; Liu, P.; Xu, S.; Lin, X.; Zeng, R. Secondary Analysis of Existing Microarray Data Reveals Potential Gene Drivers of Cutaneous Squamous Cell Carcinoma. *J. Cell. Physiol.* **2019**, *234*, 15270−15278.

(98) Edwards, L. A.; Woo, J.; Huxham, L. A.; Verreault, M.; Dragowska, W. H.; Chiu, G.; Rajput, A.; Kyle, A. H.; Kalra, J.; Yapp, D.; Yan, H.; Minchinton, A. I.; Huntsman, D.; Daynard, T.; Waterhouse, D. N.; Thiessen, B.; Dedhar, S.; Bally, M. B. Suppression of VEGF Secretion and Changes in Glioblastoma Multiforme Microenvironment by Inhibition of Integrin-Linked Kinase (ILK). *Mol. Cancer Ther.* **2008**, *7*, 59−70.

(99) Gazon, H.; Barbeau, B.; Mesnard, J. M.; Peloponese, J. M. Hijacking of the AP-1 Signaling Pathway during Development of ATL. *Front. Microbiol.* **2018**, *8*, 2686.

(100) Okura, H.; Golbourn, B. J.; Shahzad, U.; Agnihotri, S.; Sabha, N.; Krieger, J. R.; Figueiredo, C. A.; Chalil, A.; Landon-Brace, N.; Riemenschneider, A.; Arai, H.; Smith, C. A.; Xu, S.; Kaluz, S.; Marcus, A. I.; Van Meir, E. G.; Rutka, J. T. A Role for Activated Cdc42 in Glioblastoma Multiforme Invasion. *Oncotarget* **2016**, *7*, 56958.

(101) Zagzag, D.; Lukyanov, Y.; Lan, L.; Ali, M. A.; Esencay, M.; Mendez, O.; Yee, H.; Voura, E. B.; Newcomb, E. W. Hypoxia-Inducible Factor 1 and VEGF Upregulate CXCR4 in Glioblastoma: Implications for Angiogenesis and Glioma Cell Invasion. *Lab. Invest.* **2006**, *86*, 1221−1232.

(102) Amaral, R. F.; Geraldo, L. H. M.; Einicker-Lamas, M.; e Spohr, T. C. L. d. S.; Mendes, F.; Lima, F. R. S. Microglial Lysophosphatidic Acid Promotes Glioblastoma Proliferation and Migration via LPA1 Receptor. *J. Neurochem.* **2021**, *156*, 499−512.

(103) Harper, K.; Lavoie, R. R.; Charbonneau, M.; Brochu-Gaudreau, K.; Dubois, C. M. The Hypoxic Tumor Microenvironment Promotes Invadopodia Formation and Metastasis through LPA1 Receptor and EGFR Cooperation. *Mol. Cancer Res.* **2018**, *16*, 1601−1613.

(104) Quintero-Fabián, S.; Arreola, R.; Becerril-Villanueva, E.; Torres-Romero, J. C.; Arana-Argáez, V.; Lara-Riegos, J.; Ramírez-Camacho, M. A.; Alvarez-Sánchez, M. E. Role of Matrix Metalloproteinases in Angiogenesis and Cancer. *Front. Oncol.* **2019**, *9*, 1370.

(105) Kessenbrock, K.; Plaks, V.; Werb, Z. Matrix Metalloproteinases: Regulators of the Tumor Microenvironment. *Cell* **2010**, *141*, 52.

(106) Petrova, V.; Annicchiarico-Petruzzelli, M.; Melino, G.; Amelio, I. The Hypoxic Tumour Microenvironment. *Oncogenesis* **2018**, *7*, 10.

(107) Jana, S.; Singh, S. K. Identification of Selective MMP-9 Inhibitors through Multiple e-Pharmacophore, Ligand-Based Pharmacophore, Molecular Docking, and Density Functional Theory Approaches. *J. Biomol. Struct. Dyn.* **2019**, *37*, 944−965.

(108) Yamamoto, D.; Takai, S.; Jin, D.; Inagaki, S.; Tanaka, K.; Miyazaki, M. Molecular Mechanism of Imidapril for Cardiovascular Protection via Inhibition of MMP-9. *J. Mol. Cell. Cardiol.* **2007**, *43*, 670−676.

(109) Yamamoto, D.; Takai, S.; Hirahara, I.; Kusano, E. Captopril Directly Inhibits Matrix Metalloproteinase-2 Activity in Continuous Ambulatory Peritoneal Dialysis Therapy. *Clin. Chim. Acta* **2010**, *411*, 762−764.

(110) Kast, R. E.; Halatsch, M. E. Matrix Metalloproteinase-2 and -9 in Glioblastoma: A Trio of Old Drugs—Captopril, Disulfiram and Nelfinavir—Are Inhibitors with Potential as Adjunctive Treatments in Glioblastoma. *Arch. Med. Res.* **2012**, *43*, 243−247.

(111) Lastakchi, S.; Olaloko, M. K.; McConville, C. A Potential New Treatment for High-Grade Glioma: A Study Assessing Repurposed Drug Combinations against Patient-Derived High-Grade Glioma Cells. *Cancers* **2022**, *14*, 2602.

(112) Jiang, Q. W.; Chen, M. W.; Cheng, K. J.; Yu, P. Z.; Wei, X.; Shi, Z. Therapeutic Potential of Steroidal Alkaloids in Cancer and Other Diseases. *Med. Res. Rev.* **2016**, *36*, 119−143.

(113) Jiang, G.; Zhang, L.; Wang, J.; Zhou, H. Baicalein Induces the Apoptosis of U251 Glioblastoma Cell Lines via the NF-KB-P65-Mediated Mechanism Baicalein Induces the Apoptosis of U251 Glioblastoma Cell Lines via the NF-KB-P65-Mediated Mechanism. *Anim. Cell Syst.* **2016**, *20*, 296.

https://doi.org/10.1021/acsomega.3c00441

(114) Zhai, K.; Mazurakova, A.; Koklesova, L.; Kubatka, P.; Büsselberg, D. Flavonoids Synergistically Enhance the Anti-Glioblastoma Effects of Chemotherapeutic Drugs. *Biomolecules* **2021**, *11*, 1841.

(115) Xue, W.; Wang, P.; Tu, G.; Yang, F.; Zheng, G.; Li, X.; Li, X.; Chen, Y.; Yao, X.; Zhu, F. Computational Identification of the Binding Mechanism of a Triple Reuptake Inhibitor Amitifadine for the Treatment of Major Depressive Disorder. *Phys. Chem. Chem. Phys.* **2018**, *20*, 6606−6616.

(116) Khanal, P.; Dey, Y. N.; Patil, R.; Chikhale, R.; Wanjari, M. M.; Gurav, S. S.; Patil, B. M.; Srivastava, B.; Gaidhani, S. N. Combination of System Biology to Probe the Anti-Viral Activity of Andrographolide and Its Derivative against COVID-19. *RSC Adv.* **2021**, *11*, 5065−5079.

(117) Awale, S.; Miyamoto, T.; Linn, T. Z.; Li, F.; Win, N. N.; Tezuka, Y.; Esumi, H.; Kadota, S. Cytotoxic Constituents of Soymida Febrifuga from Myanmar. *J. Nat. Prod.* **2009**, *72*, 1631−1636.

(118) Sowmyya, T.; Vijaya Lakshmi, G. Antimicrobial and Catalytic Potential of Soymida Febrifuga Aqueous Fruit Extract-Engineered Silver Nanoparticles. *Bionanoscience* **2018**, *8*, 179−195.

(119) Lim, M.; Xia, Y.; Bettegowda, C.; Weller, M. Current State of Immunotherapy for Glioblastoma. *Nat. Rev. Clin. Oncol.* **2018**, *15*, 422−442.

(120) Lee, Y.-L.; Cheng, W.-E.; Chen, S.-C.; Chen, C.; Shih, C.-M. The Effects of Hypoxia on the Expression of MMP-2, MMP-9 in Human Lung Adenocarcinoma A549 Cells. *Eur. Respir. J.* **2014**, *44*, P2699.

(121) Bauer, A. T.; Bürgers, H. F.; Rabie, T.; Marti, H. H. Matrix Metalloproteinase-9 Mediates Hypoxia-Induced Vascular Leakage in the Brain via Tight Junction Rearrangement. *J. Cerebr. Blood Flow Metabol.* **2010**, *30*, 837.

https://doi.org/10.1021/acsomega.3c00441

Detecting Fake Drugs using Blockchain

Abhinav Sanghi, Aayush, Ashutosh Katakwar, Anshul Arora, Aditya Kaushik

[1] ***Abstract***: ***The existing supply chain for the pharmaceutical industry is obsolete and lacks clear visibility over the entire system. Moreover, the circulation of counterfeit drugs in the market has increased over the years. According to the WHO report, around 10.5% of the medicinal drugs in lower / middle income countries are fake and such drugs may pose serious threats to public health, sometimes leading to death. Keeping these threats in mind, in this paper, we propose a blockchain-based model to track the movement of drugs from the industry to the patient and to minimize the chances of a drug being counterfeit. The reasons for using blockchain technology in our work include its immutability property and easy tracking of an entity in the blockchain. Through this proposed model, the manufacturer would be able to upload the details corresponding to a drug, after which it will be sent for approval to the Government. Thereafter, hospitals and pharmacies, based upon their requirements, can request the approved drugs. In the future, if a patient wants some medication, then he or she has to request it on the blockchain network. The request will be sent to the nearest hospital/pharmacy and thereafter, the patient can collect the medication. To implement this model, we have used Hyperledger fabric due to the presence of many auto-implemented features in it. Our implementation of the proposed blockchain based model highlights that the model can successfully detect any drug being counterfeit. This will be beneficial for the users getting affected with counterfeit drugs. Moreover, with the proposed model, we can also track the movement of the drug beginning from the manufacturer right up to the patient consuming that drug.***

Index Terms: ***Blockchain, Counterfeit Drugs, Drugs Tracking, Fake Medicines, Health Care.***

Keywords: The Reasons For Using Blockchain Technology In Our Work Include Its Immutability Property And Easy Tracking Of An Entity In The Blockchain.

I. INTRODUCTION

In this era, the world of piracy and counterfeiting has touched nearly every product including medicines and drugs. The challenge of counterfeit drugs in the pharmaceutical industry has been increasing across the globe over the past many years. According to a WHO report [1], around 10.5% of the pharmaceutical drugs in the markets of low or middle-income countries are fake. Hence, there is a need to develop a strong model to overcome the issue of counterfeiting drugs. Moreover, the current industry lacks clear visibility over the delivery of the drugs from the pharmaceutical company to the patients. Keeping these challenges in mind, we aim to develop a blockchain-based model that can prevent drug counterfeiting and keep track of drug movement from the industry to the patients.

Contributions: Such a problem of counterfeiting drugs and their tracking can be solved by applying QR codes on them during their manufacturing process. Thereafter, we can track their journey by scanning their QR codes. However, because one can make a copy of the QR code and this copied code can be applied to the counterfeit drug, this solution will not completely solve the problem of drugs tracking and counterfeiting. Hence, we came up with a model based on a decentralized system such as blockchain, using Hyperledger fabric, in which the manufacturer will create a drug and will upload the details on this blockchain. After that, the Government will approve these drugs. Thereafter, hospitals and pharmacies can request the available approved drugs as per their requirements. In the future, if any patient wants some medication, then he or she has to request it on the blockchain network, and then the request will be sent to the nearest hospital/pharmacy and after that, the patient can collect the required drugs. The main advantage of using such a blockchain network is that drug tracking is easy as the drug is visible on the network at every stage. Moreover, because this blockchain network is closed, no one from the outside can fraud the drugs.

Organization: The rest of the paper is structured as follows. We review the related works in the field of blockchain and the healthcare industry in Section 2. We discuss the proposed blockchain-based model in detail in Section 3, and we conclude in Section 4.

II. RELATED WORK

In this section, we review the related work in the field of healthcare and blockchain. We further divide this section into two subsections: 1. Blockchain-related works, and 2. Blockchain Applied in the healthcare field.

A. Blockchain Related Works

First, what we have done is, review the works that have discussed blockchain network various use cases. The study conducted in [2] proposed an approach based on a decentralized solution which is blockchain to creating a DT (Digital Twin) which ensured the data traceability, data authenticity, and immutability of information. They utilized the decentralized IPFS stockpiling workers to store the information identified with DT.

Manuscript received on April 27, 2021.
Revised Manuscript received on May 04, 2021.
Manuscript published on May 30, 2021.
* Correspondence Author

Abhinav Sanghi*, Mathematics and Computing, Delhi Technological University, Delhi, India. Email: sanghi.aabhu1@gmail.com

Aayush, Mathematics and Computing, Delhi Technological University, Delhi, India. Email: aayushsimple28@gmail.com

Ashutosh Katakwar, Mathematics and Computing, Delhi Technological University, Delhi, India Email: ashutoshkatakwar26@gmail.com

Anshul Arora, Faculty, Mathematics and Computing, Delhi Technological University, Delhi, India. Email: anshularora@dtu.ac.in

Aditya Kaushik, Faculty, Mathematics and Computing, Delhi Technological University, Delhi, India Email: akaushik@dtu.ac.in

It is very convenient if one can remotely control IOT devices but it comes with the cost of data exploitation. Lin et al. [3] designed a secure and efficient remote user confirmation system based on a blockchain model. They incorporated this decentralized method, group benchmark, and subject matter validation to give solid evaluation of the user's entrance history.

Parking vehicles in big cities can be a big challenge. Many savvy parking applications tend to solve these problems however, a large number people experience the ill effects of protection issues or they work in a centralized environment. Zhang et al. [4] proposed smart parking based on a decentralized approach which was reliable and the privacy was protected. Using the concept of smart contracts, they were able to accomplish fairness. The writer in [7] made a singular assent working model for the information sharing stages identified with wellbeing. Through this model, they successfully managed the accountability of all the members in the information sharing stages. This model also makes sure that a person’s will is kept on priority. Autonomous vehicles can sense and navigate in their environment without the interference of human hands. But in case of road accidents, mishap legal sciences must decide the obligations. The authors in [8] made an instrument of verification of occasion with a dynamic alliance agreement to accomplish undeniable mishap crime scene investigation by generating trustworthy data. The most challenging aspect in conducting elections is to gain the faith of all the electors in the counting cycle. Yang et al. [9] proposed a protocol for election which is based on decentralized approach i.e. blockchain-technology. Their model does not need a committee to count the votes. They successfully created an encryption system that guarantees that no one can decode the votes but each one of us can ensure the legitimacy of the votes. Yuan et al. [10] directed a study based on blockchain Intelligent Transportation Systems (ITS). They introduced a contextual investigation for blockchain based on going ride-sharing services.

B. Blockchain for Healthcare and Drugs

In this subsection, we discuss the works that have applied blockchain in any aspect of the medical industry. The authors in [14] highlighted various use cases of blockchain in the medical industry such as research on users that are on medication and management of the health sector. In this report, they mentioned strategies to remove middlemen in the medical sector and also highlighted the new way of doing medical transactions. McGhin et al. [15] addressed the future research side of the medical sector . They also addressed some of the unique requirements that are not addressed in earlier conducted blockchain experiments. Apart from that, they addressed research areas like scalability, block withholding attack, and blockchain mining incentives in which blockchain might lack for the healthcare sector. Siyal et al. [17] reviewed the work that is already existed in the medical sector. They also highlighted the recent research in this sector that is using a decentralized model such as blockchain. They highlighted the usability of a decentralized model for neural system. They have successfully stored a virtual digital brain on the decentralized network like blockchain.. They also highlighted some impactful factors that are creating hurdles of the blockchain in the medical sector. Every blockchain based model runs according to the smart contract. Kumar et al. [20] designed the same for the medical industry. Apart from that, they highlighted various challenges of blockchain for healthcare such as scalability restrictions, high development cost, standardization challenges, cultural resistance, regulatory uncertainty, etc.

Bell et al. [21] addressed the issues existing in the current healthcare field. They also highlighted fields in the healthcare system where blockchain can solve various problems. Some of those fields are Clinical Trials, Data Sharing between many entities (such as hospitals, manufacturers), Patient Records, etc. They also mentioned that the resistance of this system to adopt blockchain in their supply chain is the biggest reason for not using blockchain in healthcare. The authors in [22] addressed the use cases of blockchain and reviewed, assessed various publications and consequently proposed a methodology which is to integrate blockchain with the processes involved in the current healthcare system. They found that EHR (Electronic Health Records) and PHR (Personal Health Records) are the areas where blockchain is mainly used and Ethereum and Hyperledger fabric are the most preferred open-source frameworks for developing a blockchain based application. The authors in [23] highlighted the privacy related to data stores and the data sharing platforms. They used a permissioned blockchain to ensure the vulnerabilities are removed corresponding to data transfer and since the system is decentralized it also ensures the issue of single point of failure is resolved. They also developed a mobile application which collects user’s data through the means of manual input and medical devices. Dwivedi et al. [24] pointed out the issues related with privacy and security associated with data sharing and storing. They used blockchain to ensure secure handling of data within the network. They also tried to solve the problems associated with integrating blockchain with IoT devices and proposed a new structure of blockchain that can be integrated with IoT devices.

The authors in [33] highlighted the issue regarding drug safety and tried to solve the same issue using Blockchain technology which was integrated with QR code. They highlighted the irregularities present in the current supply chain of pharma industries and proposed a methodology that consisted of blockchain-based architecture for the supply chain. Their proposed methodology ensured the reliability aspect of the drug as well as well as the genuineness of the involved manufacturer. Haq et al. [34] specified the problems that are present in the current pharma supply chain and explained how blockchain can be used instead of the current supply chain to ensure traceability and transparency while transferring a particular entity from one level to another. They suggested a permissioned blockchain for storing all the data involved within the network and since it is a permissioned blockchain so it ensured that only trusted parties are becoming a part of the network.

The authors in [35] highlighted the use cases of a decentralized model such as blockchain in the medical sector. They discussed the use of blockchain in various fields such as EHR (Electronic Health Records), Medical Insurance, Bio-medical Field and Medical Supply Chain.In conclusion they stated that this technology has still not been adopted by healthcare systems where this is capable of solving various problems.People who are in-charge of making these decisions should become aware of the technology's potential and the revolutionary power that it carries with itself and should introduce it in the current healthcare system.

Debe et al. [36] addressed various issues that can impact the medical supply chain. He also mentioned some reasons for this such as wrong prescribed medicines to patients, ordering of too many medicines. To wipe out the above mentioned issue, they proposed a blockchain solution that can handle the easy return and exchange of medicines that can be used further.In their proposed system, medical stores and customers have power to give the fit drugs to needy one at a lower price.They have used Ethereum for implementing their idea and designed an architecture for the same. Shae et al. [37] proposed an architecture based on blockchain for solving problems such as improper prescription of medicines for patients .They also highlighted various issues such as technical problems on implementing this and shared some insights for solving them. In the paper,the authors briefly talked about the medical sector and described the architecture design,barriers and usability of blockchain in medical sector. Saxena et al. [38] briefly talked about the medical supply chain and highlighted various issues that are making the supply chain worse.They also discussed the current implemented strategy for solving the counterfeiting problem.They had researched with people that are from the medical industry and developed a blockchain based tool "Pharma Crypt" for solving some of the problems.

III. METHODOLOGY AND IMPLEMENTATION

In this section, we explain our proposed blockchain-based model for drugs tracking and counterfeiting. We begin our discussion with the introduction to Smart Contracts and after that we discuss the concept of transactions in blockchain.We show how a transaction will execute in blockhain network with diagrammatic representation.

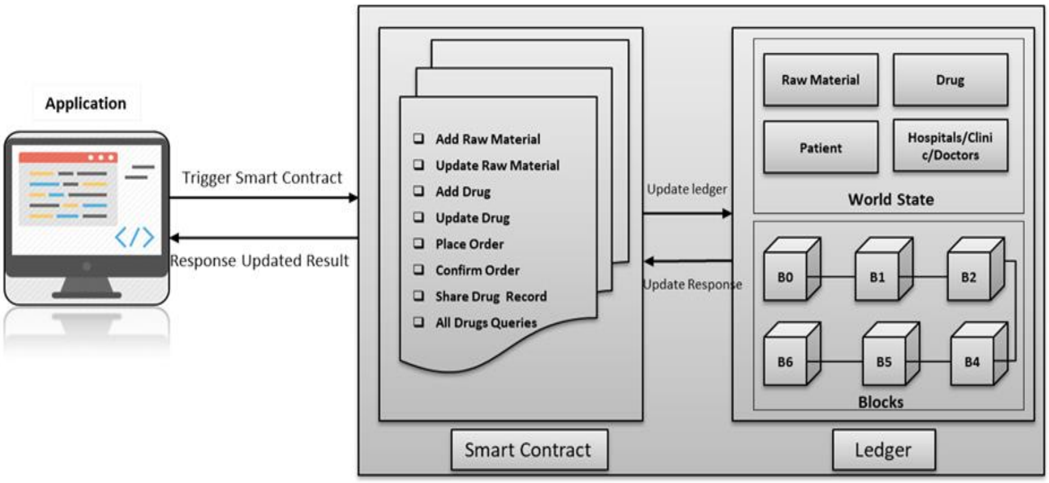


Figure 1: Smart Contract Architecture

A. Smart Contracts

A Smart Contract is a few lines of code which is automatically executed whenever some terms and conditions(that are already set) are satisfied in a blockchain network. It can include the transfer of assets from one level to another or some kind of update in the network.

Basically, it is a piece of code that enforces the agreement done between two parties without paying any amount to a third person. They also enable the users to manage their access rights and their assets among different parties. They are stored on ledgers and are secured from any kind of tampering. The time complexity for the execution of transactions is high because the transactions are executed among all the peer nodes of the network periodically or in sequence. Moreover, the data corresponding to that is written on all the ledgers hence giving rise to space complexity. This issue is addressed in our model by deploying the contracts only for certain nodes of the network and not all the nodes. Hence, some nodes of this network can validate the transactions, and hence space and time complexity is improved. We used Java and Node.JS programming languages for writing smart contracts. The figure 1 represents an overall architecture of smart contract execution for our proposed model.

B. Transactions Execution Procedure

The roadmap for a transaction in the network is summarized in figure 2. Users are shown the front-end of the application where the credentials are required so that the user can enter the blockchain network. Enrollment of all the participants in the network is done by the administrator which gives the credentials along with an enrollment certificate to all the users. After the user has gone through the login process, the user initiates a transaction using his/her credentials. Then the request is sent to all the peer nodes (divided into two sub-categories: committers and endorsers).

Endorsers execute the transaction if it is valid. Committers validate the result obtained by executing those requests before it is written to the ledger. We can also say that endorsers are similar to committers who hold the predefined smart contract. These endorsers execute the contract of requested transactions in their simulated environment before writing to the ledger. The endorser fetches the read/write data while executing the request in their environment which is RW set. The read in the RW set contains information about the world state before the transaction is executed and the writing part of the RW set contains information about what is written in the world state after the request is executed in the environment. Then endorsers return the executed transaction to the client application along with RW sets. The user again submits the signed transaction with all RW sets to the consensus node. The consensus node sends the transaction to the committers. The committers validate by matching the current world state and if matched then it is written into the ledger. Finally, committers will send an alert message regarding the transaction status. Communication between application and blockchain network is achieved with the help of REST API and SDK.

C. Drug Supply Chain with Hyperledger

Now, we will demonstrate the working of the Drug Supply Chain using Hyperledger fabric. The first step that is involved in this application is to first start the blockchain network. The network consists of the peers, the ORDERER, and the chain code which is installed on all the peers. The chain code is nothing but the smart contract, which is installed in all the peers. Moving on, we will see how the flow of the application works through a network diagram as shown in figure 4.

Figure 4 summarizes an architectural view of our proposed model. We identified six active peers, i.e., the manufacturer, Government, Drugs Administration Organization (DAO, something similar to FDA), hospital, customer, and the doctor.

We'll now discuss how drug-related information moves in the blockchain to all the peers, starting from the manufacturer and ending at the customer. Suppose the manufacturer creates a drug, and he enters the details of the drugs in the blockchain.

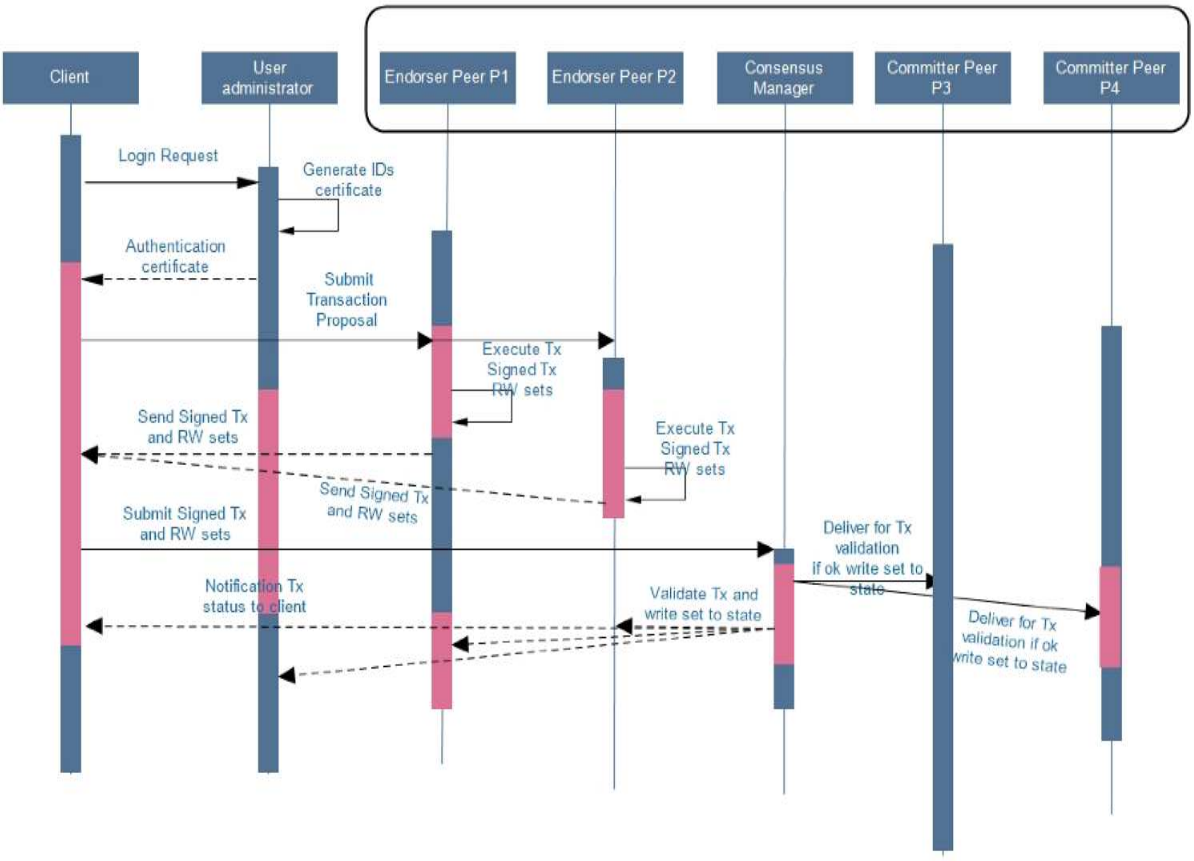
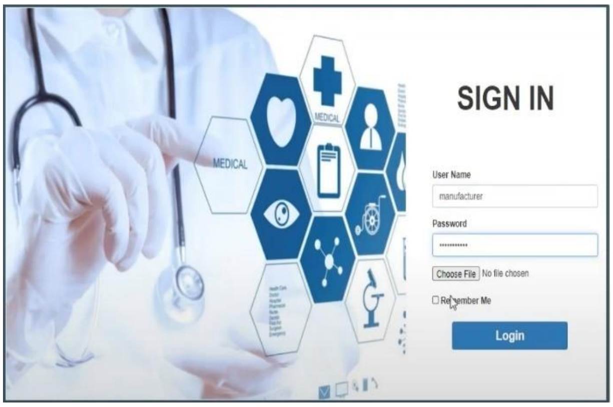

Client
User administrator
Endorser Peer P1
Endorser Peer P2
Consensus Manager
Committer Peer P3
Committer Peer P4
Login Request
Generate IDs certificate
Authentication certificate
Submit Transaction Proposal
Execute Tx Signed Tx RW sets
Send Signed Tx and RW sets
Execute Tx Signed Tx RW sets
Send Signed Tx and RW sets
Submit Signed Tx and RW sets
Deliver for Tx validation if ok write set to state
Notification Tx status to client
Validate Tx and write set to state
Deliver for Tx validation if ok write set to state

Figure 2: Transaction Execution Procedure

These drugs need to be approved at the government end and the DAO. Say, once the manufacturer enters the drugs, the drug details are added in the blockchain, and it moves to the government, the details are then distributed to all the other peers as well. The government entity will check whether the manufacturer is proper or not, and then approves the drug. The DAO will check whether the quality of the drug is okay and approve the drug. Once both the government and the DAO have approved the drug, it is then moved to the hospital or the pharmacy, where it is clear and safe to the customers and patients and then a customer can purchase it. Once the customer retains a particular drug by inputting what condition he is suffering from, this particular information of the customer and the drug that he has purchased, i.e., the prescription, is then sent to the doctors for approval.

The doctors or the medical professionals, verify that the drug, which is bought by the customer, is prescribed properly, and then approve the purchase. So, once it is approved, the customer is able to purchase that particular drug. This is a flow that starts from the manufacturer who creates the drug and then ends with the customer who purchases the drug. Hence, the drug-related information moves through the blockchain among all the peers. Now that we have shown the blockchain network itself, and the basic flow diagram in Figure 4, we will now see how the entire application works, through the self-designed user interface. Every peer has to log in through a web page, so we will now log in as the manufacturer. Figure 3 is the sample login page of the manufacturer.

There is a choose file option in which we need to select the manufacturer-specific ID certificate to authorize him and encrypt transactions in the blockchain network.



SIGN IN

User Name: manufacturer

Password: ••••••••••

Choose File No file chosen

□ Remember Me

Login

Figure 3: Manufacturer Login Interface

These certificates are created when the blockchain network itself is started and act as a sort of fingerprint or an ID card to authenticate that particular identity into the blockchain network.These are peers specific certificates and every peer has a unique ID. So, now we are in the drug manufacturer interface. This is where the entire flow starts, as shown in the architecture diagram in Figure 4. In this manufacturer interface, we create the drug details as shown in Figure 5.

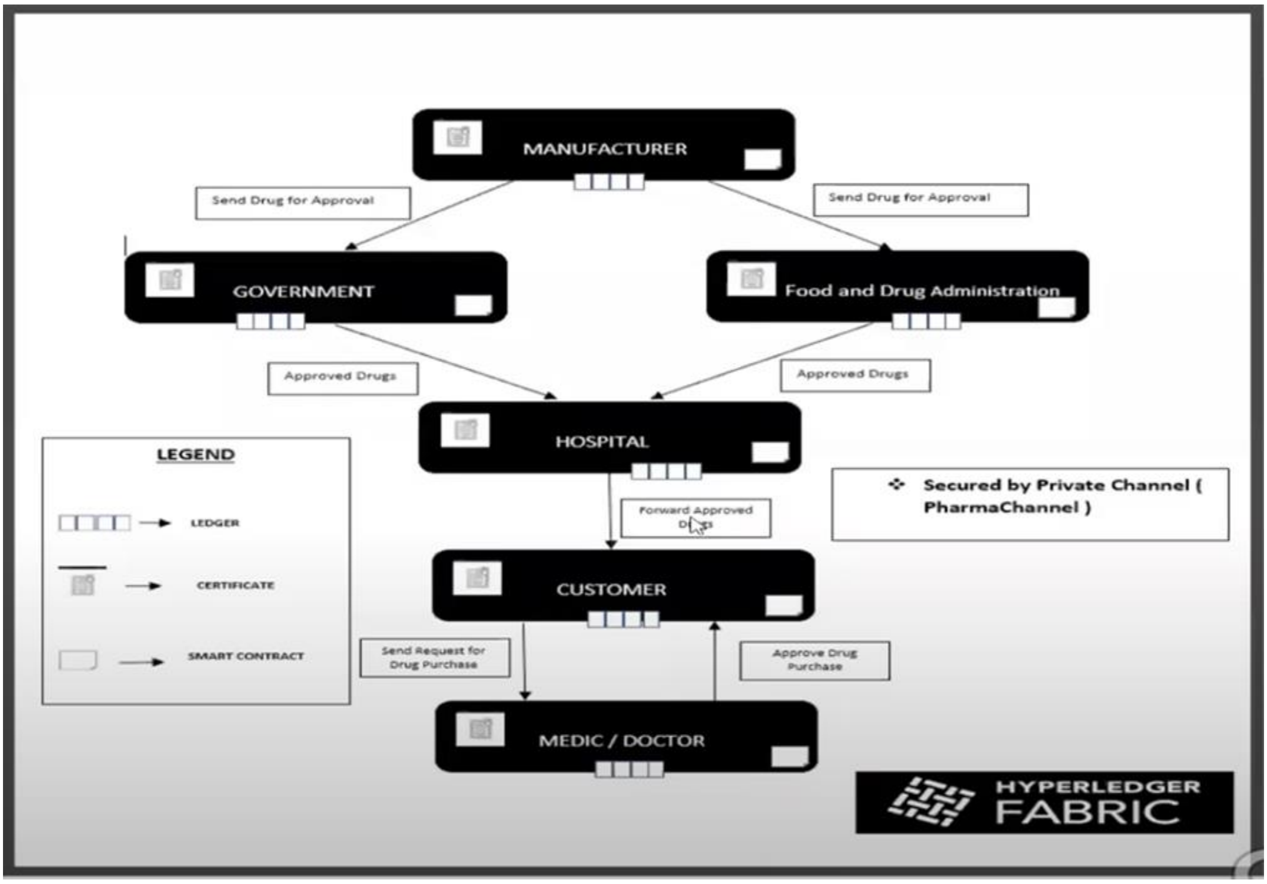
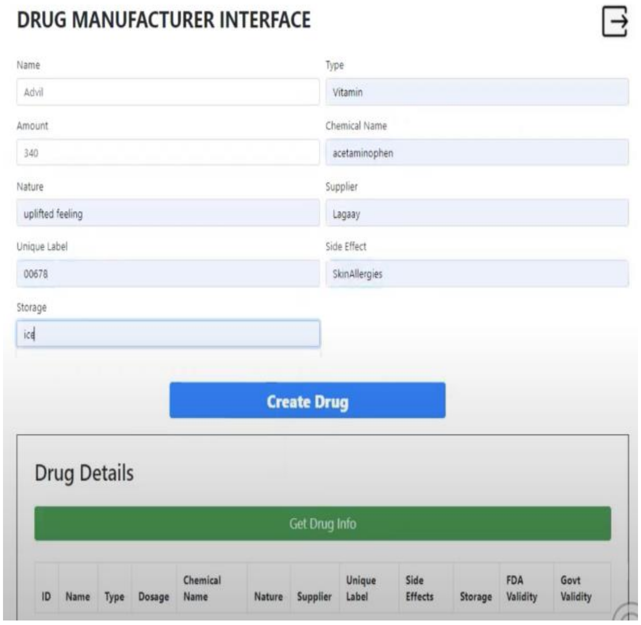


Figure 4: Architectural View of Proposed Model

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication
© Copyright: All rights reserved.



DRUG MANUFACTURER INTERFACE

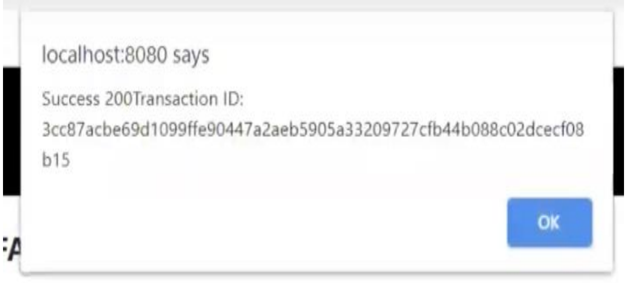
Name: Advil	Type: Vitamin
Amount: 340	Chemical Name: acetaminophen
Nature: uplifted feeling	Supplier: Lagaay
Unique Label: 00678	Side Effect: SkinAllergies
Storage: ice	

Drug Details

ID	Name	Type	Dosage	Chemical Name	Nature	Supplier	Unique Label	Side Effects	Storage	FDA Validity	Govt Validity

Figure 5: Drug Manufacturing Interface

We have given the details of a drug to create it in Figure 5. And once we click the button *Create drug*, then we'll have a successful transaction, and the transaction ID will be shown as a notification prompt as shown in Figure 6.



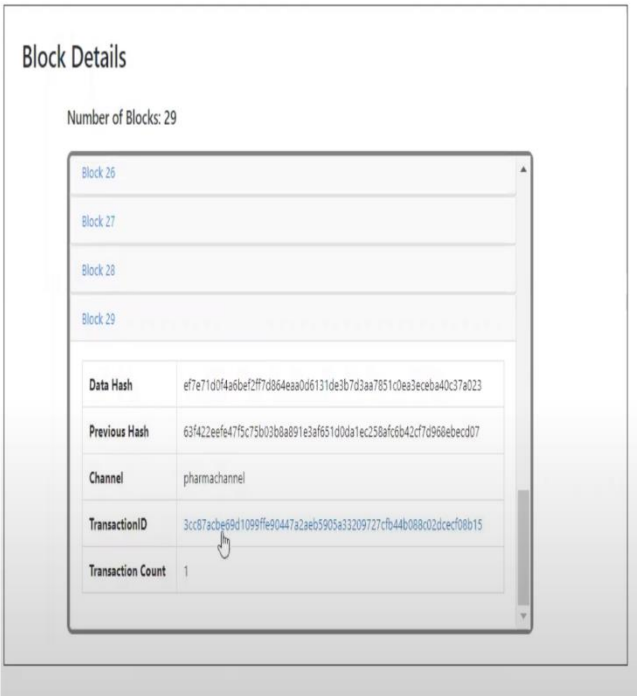
localhost:8080 says

Success 200Transaction ID: 3cc87acbe69d1099ffe90447a2aeb5905a33209727cfb44b088c02dcecf08b15

Figure 6: Notification Prompt

It goes like 3cc87. So in every peer interface, we also intend to show the blockchain itself and the number of blocks for transparency, as shown in Figure 7. Whereas, in Ethereum, there are tools like ganache to show the entire state of the blockchain, including different log details and the different transaction details. However, in this demo application, we intend to show it by ourselves by leveraging the Hyperledger fabric.

For instance, we have created the ADVIL drug that is added into the blockchain with block number 29. If we click on a particular block number as shown in Figure 8, we will get information like Transaction ID, Block Hash, Channel name, drug name, Time Stamp, etc. This is the powerful feature of blockchain and the Hyperledger fabric. Also, as this data is immutable, no one can change it. Each successful entry will be recorded as a transaction or a block in the blockchain network. In our application, whenever we create a drug, it is considered as a transaction and this particular transaction is added as a block into the blockchain, which is then distributed among all the peers in the network.



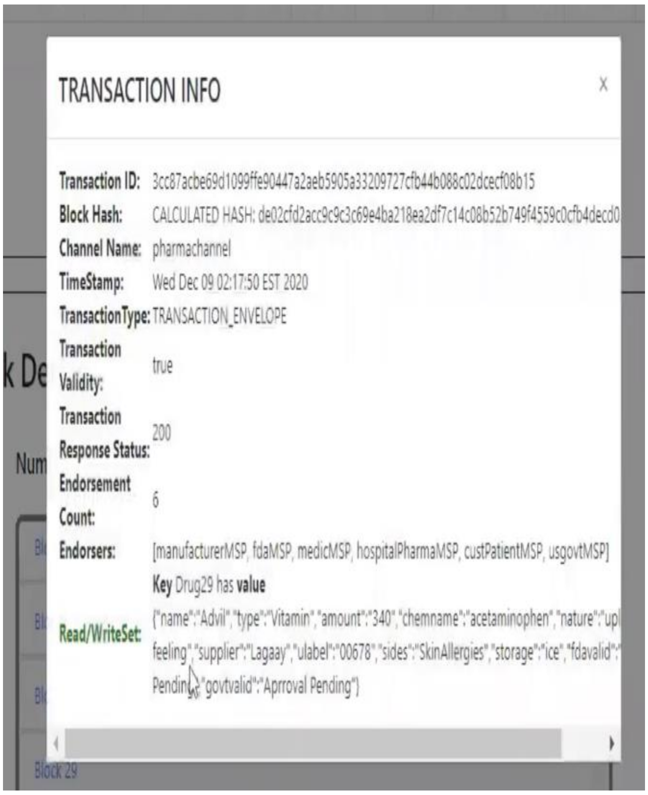
Block Details

Number of Blocks: 29

Data Hash	ef7e71d0f4a6bef2ff7d864eaa0d6131de3b7d3aa7851c0ea3eceba40c37a023
Previous Hash	63f422eefe47f5c75b03b8a891e3af651d0da1ec258afc6b42cf7d968ebecd07
Channel	pharmachannel
TransactionID	3cc87acbe69d1099ffe90447a2aeb5905a33209727cfb44b088c02dcecf08b15
Transaction Count	1

Figure 7: Block Details

Similarly, we have a button named Get Drug Info which is a query to the blockchain to view the block details. This button queries the smart contract to get all the drug details.



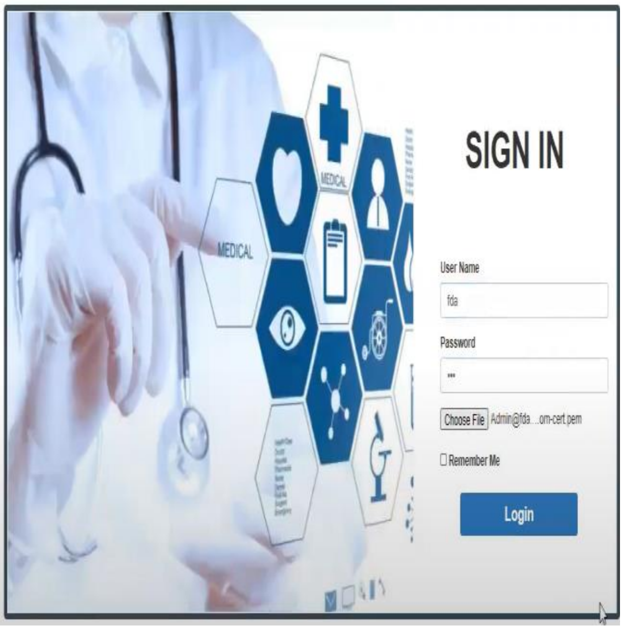
TRANSACTION INFO

Transaction ID:	3cc87acbe69d1099ffe90447a2aeb5905a33209727cfb44b088c02dcecf08b15
Block Hash:	CALCULATED HASH: de02cfd2acc9c9c3c69e4ba218ea2df7c14c08b52b749f4559c0cfb4decd0
Channel Name:	pharmachannel
TimeStamp:	Wed Dec 09 02:17:50 EST 2020
TransactionType:	TRANSACTION_ENVELOPE
Transaction Validity:	true
Transaction Response Status:	200
Endorsement Count:	6
Endorsers:	[manufacturerMSP, fdaMSP, medicMSP, hospitalPharmaMSP, custPatientMSP, usgovtMSP]
Read/WriteSet:	**Key** Drug29 has **value** {"name":"Advil","type":"Vitamin","amount":"340","chemname":"acetaminophen","nature":"upl feeling","supplier":"Lagaay","ulabel":"00678","sides":"SkinAllergies","storage":"ice","fdavalid":" Pending","govtvalid":"Aprroval Pending"}

Figure 8: Transaction Information

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication

Once we create the drug, the FDA and the government entity need to approve it for retail in pharmacies. Figure 9 shows the login interface of the DAO.



SIGN IN

User Name
fda

Password
***

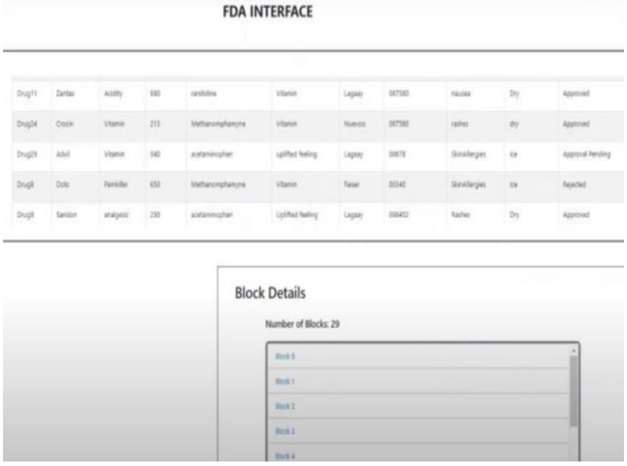
Choose File Admin@fda....om-cert.pem

□ Remember Me

Login

Figure 9: DAO Login Interface

We get the drug info from the blockchain. We created a few drugs, and the number of blocks till now is 29. The ADVIL medicine has pending approval as shown in Figure 10.



FDA INTERFACE

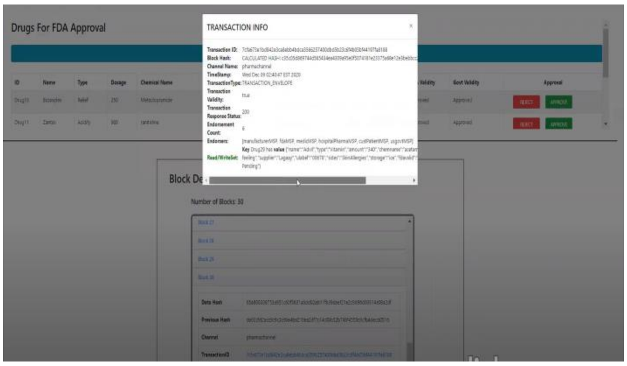
Drug11	Zantax	Acidity	980	ranitidine	Vitamin	Lagaay	067580	nausea	Dry	Approved
Drug24	Crocin	Vitamin	213	Methanomphamyne	Vitamin	Nuevoo	067580	rashes	dry	Approved
Drug29	Advil	Vitamin	340	acetaminophen	uplifted feeling	Lagaay	00678	SkinAllergies	ice	Approval Pending
Drug8	Dolo	Painkiller	650	Methanomphamyne	Vitamin	Faiser	00340	SkinAllergies	ice	Rejected
Drug9	Saridon	analgesic	250	acataminophen	Uplifted feeling	Lagaay	098452	Rashes	Dry	Approved

Block Details

Number of Blocks: 29

Figure 10: Approval Pending from DAO

So,once we approve, say ADVIL medicine, then after a successful operation, the number of blocks is now 30 with the 30th block content containing the key-value pair with the approved message, as shown in Figure 11. That is how the DAO will approve drugs.



Drugs For FDA Approval

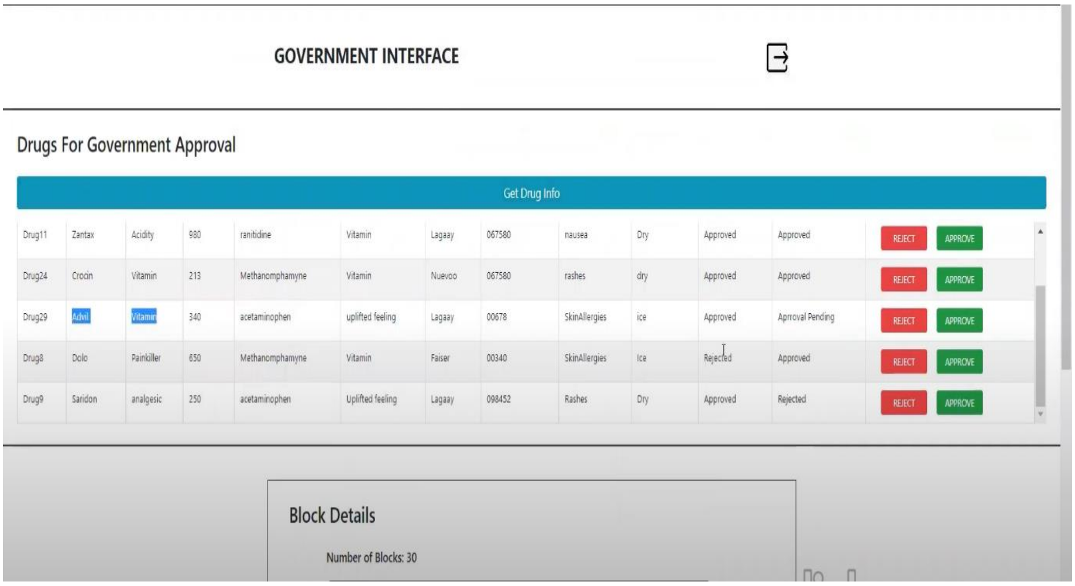
TRANSACTION INFO

Transaction ID: [illegible]
Block Hash: CALCULATED HASH: [illegible]
Channel Name: pharmachannel
TimeStamp: Wed Dec 09 02:40:47 EST 2020
TransactionType: TRANSACTION_ENVELOPE
Transaction Validity: true
Transaction Response Status: 200
Endorsement Count: 6
Endorsers: [manufacturerMSP, fdaMSP, medicMSP, hospitalPharmaMSP, custPatientMSP, usgovMSP]
Read/WriteSet: Key Drug29 has value {"name":"Advil","type":"Vitamin","amount":"340","chemname":"acetam [illegible] feeling","supplier":"Lagaay","ulabel":"00678","sider":"SkinAllergies","storage":"ice","fdavalid":" [illegible] Pending"}

Block De

Number of Blocks: 30

Figure 11: Drug Approval by DAO



GOVERNMENT INTERFACE

Drugs For Government Approval

Get Drug Info

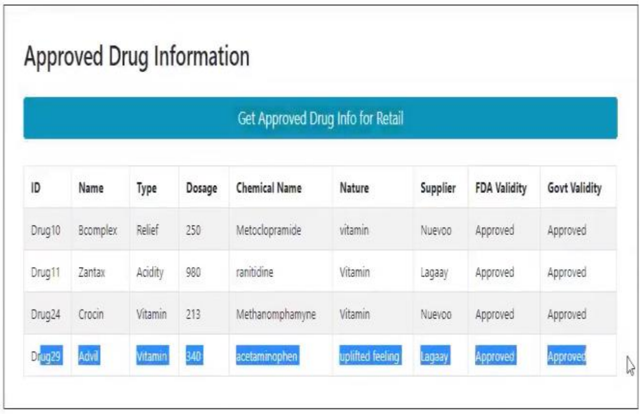
Drug11	Zantax	Acidity	980	ranitidine	Vitamin	Lagaay	067580	nausea	Dry	Approved	Approved	REJECT APPROVE
Drug24	Crocin	Vitamin	213	Methanomphamyne	Vitamin	Nuevoo	067580	rashes	dry	Approved	Approved	REJECT APPROVE
Drug29	Advil	Vitamin	340	acetaminophen	uplifted feeling	Lagaay	00678	SkinAllergies	ice	Approved	Aprroval Pending	REJECT APPROVE
Drug8	Dolo	Painkiller	650	Methanomphamyne	Vitamin	Faiser	00340	SkinAllergies	Ice	Rejected	Approved	REJECT APPROVE
Drug9	Saridon	analgesic	250	acetaminophen	Uplifted feeling	Lagaay	098452	Rashes	Dry	Approved	Rejected	REJECT APPROVE

Block Details

Number of Blocks: 30

Figure 12: Approval Pending from Government

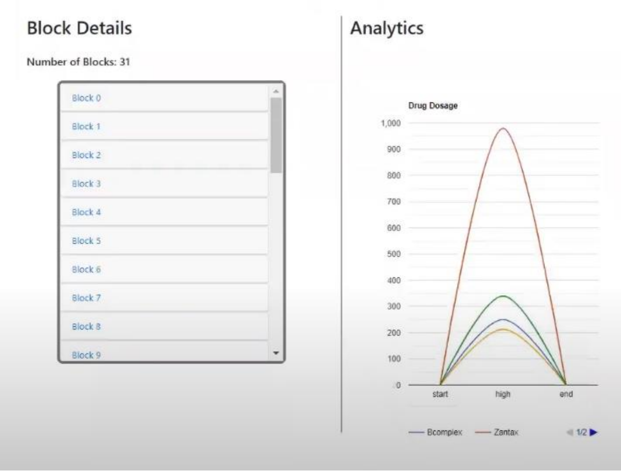
Now moving to the government interface. After login, we will see the drugs approved by DAO. Now, the government needs to approve it. After the government approves it, a successful notification will come up and this transaction is then recorded into the blockchain network. Now, as shown in Figure 12, the drug ADVIL has been approved by the DAO and has approval pending from the Government. So, once the government approves ADVIL, then a successful transaction notification will pop up with the transaction ID, and this successful transaction is added as a block in the blockchain network. Hence, now the total blocks become 31.So, now that the DAO and the government have approved the drug, it is now cleared for retail in pharmacies. So, if we log in to the pharmacy or the hospital interface, this will show all the drugs that are genuine and have been approved by FDA and the government, as summarized in Figure 13.



Approved Drug Information

ID	Name	Type	Dosage	Chemical Name	Nature	Supplier	FDA Validity	Govt Validity
Drug10	Bcomplex	Relief	250	Metoclopramide	vitamin	Nuevoo	Approved	Approved
Drug11	Zantax	Acidity	980	ranitidine	Vitamin	Lagaay	Approved	Approved
Drug24	Crocin	Vitamin	213	Methanomphamyne	Vitamin	Nuevoo	Approved	Approved
Drug29	Advil	Vitamin	340	acetaminophen	uplifted feeling	Lagaay	Approved	Approved

Figure 13: Pharmacies / Hospital Interface



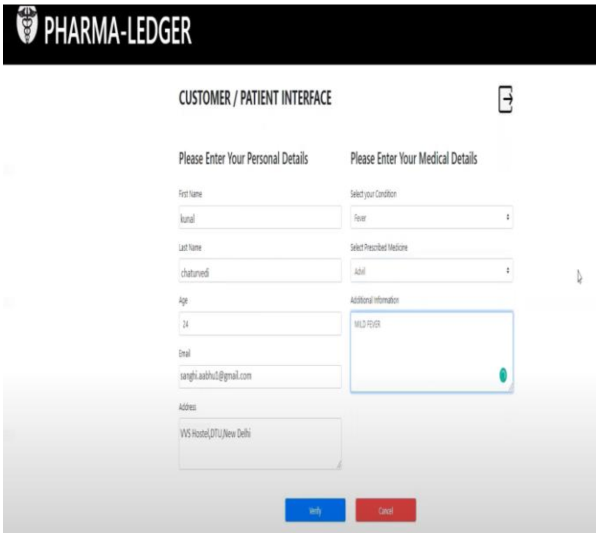
Block Details

Number of Blocks: 31

Figure 14: Analytical Section

This analytic section is just provided to show how data comes in near real time through the blockchain and it can be used for different charting and statistical purposes also. So, this is just to show that data appears very fast and can be used by many analytics companies for reporting statistics data for those purposes

Now that drugs are approved, they are available to be purchased at pharmacies. So, now, we log into the customer interface. In this interface, the patient, or the customer can purchase approved drugs that are at the pharmacy, as shown in Figure 15. So, he will enter his details like name, age, email, etc. Then select a condition like fever, influenza, etc., and then select the medicines that are available for purchase (the approved drugs) and ready for sale. Once we submit these customer details, it is sent to the doctor or the medical professional who will approach the prescription, and this will be added as a transaction/block in the network. Hence, now the total blocks become 32.



PHARMA-LEDGER

CUSTOMER / PATIENT INTERFACE

Please Enter Your Personal Details

First Name: kunal

Last Name: chaturvedi

Age: 24

Email: sanghi.aabhu1@gmail.com

Address: VVS Hostel,DTU,New Delhi

Please Enter Your Medical Details

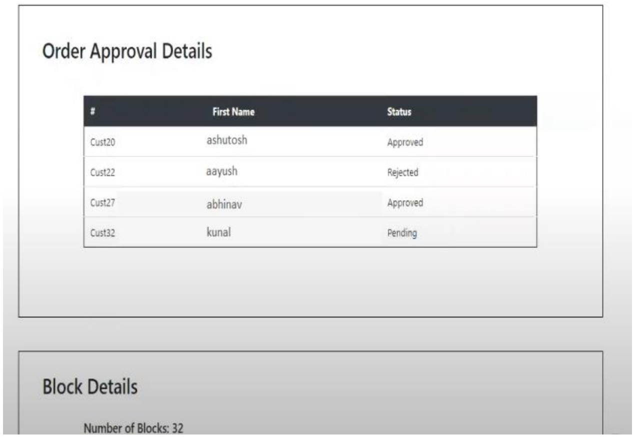
Select your Condition: Fever

Select Prescribed Medicine: Advil

Additional Information: MILD FEVER

Figure 15: Patient Interface

Once we submit these customer details, it is sent to the doctor or the medical professional who will approach the prescription, and this will be added as a transaction/block in the network. Hence, now the total blocks become 32. We can see that the customer details have been passed on to the doctor and has pending approval, shown in Figure 16.



Order Approval Details

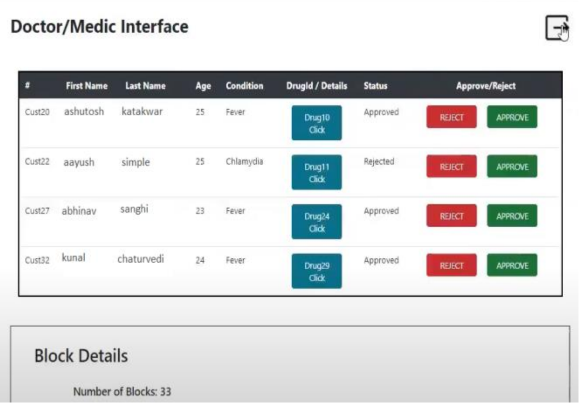
#	First Name	Status
Cust20	ashutosh	Approved
Cust22	aayush	Rejected
Cust27	abhinav	Approved
Cust32	kunal	Pending

Block Details

Number of Blocks: 32

Figure 16: Pending Approval from Doctor

So, now we will finally log in with the doctor's interface, by giving the required certificate. Here, we can see the customer's prescription & the drugs that he requires. So if the doctor feels that the prescription is okay, then he can approve it. The successful approval counts as a transaction and will be recorded as a block in the blockchain network. Now, the block count becomes 33, as shown in Figure 17



Doctor/Medic Interface

#	First Name	Last Name	Age	Condition	DrugId / Details	Status	Approve/Reject
Cust20	ashutosh	katakwar	25	Fever	Drug10 Click	Approved	REJECT APPROVE
Cust22	aayush	simple	25	Chlamydia	Drug11 Click	Rejected	REJECT APPROVE
Cust27	abhinav	sanghi	23	Fever	Drug24 Click	Approved	REJECT APPROVE
Cust32	kunal	chaturvedi	24	Fever	Drug29 Click	Approved	REJECT APPROVE

Block Details

Number of Blocks: 33

Figure 17: Doctor's Approval for Medicine

Therefore, after that, if the customer logs into the customer's interface, he can see that the required drug is now approved for purchase. Hence, this forms the entire flow of the proposed model. We started from the manufacturer and ended with the customer. With the help of Hyperledger Fabric, we can remove the counterfeiting drugs from the market because this type of network is totally transparent and tamper-free.

IV. CONCLUSION

Serious health issues, including deaths, may occur if the users consume counterfeit drugs. Several counterfeit drugs have been detected in the market of lower / middle income countries. Hence, detecting such fake drugs in the market is a big challenge. Keeping their threats in mind, in this paper, we have proposed a blockchain based model to detect such fake drugs. The proposed model also aims to track the movement of drugs from the industry to the patient. We have used the Hyperledger fabric to implement the entire model. In the proposed model, the manufacturer first has to upload the details of a drug which is sent further to the government for approval. Once it is approved, the pharmacies can request the approved drugs within the blockchain network. Further, if a patient needs to get some medicine/drugs, then a corresponding request is made into the blockchain network. Then, a medical officer/doctor approves or rejects his request. Because the entire model is implemented in a blockchain network, it can prevent counterfeiting of drugs and we can easily track the movement of drugs from the manufacturer up to the patient.

REFERENCES

1. The impact of counterfeit drugs in south and south-east Asia, Available:https://www.europeanpharmaceuticalreview.com/article/92194/the-impact-of-counterfeit-drugs-in-south-and-south-east-asia/
2. Hasan et al. , "A Blockchain-Based Approach for the Creation of Digital Twins", IEEE Access, vol. 8, pp. 34113-34126, 2020.
3. C. Lin, D. He, N. Kumar, X. Huang, P. Vijayakumar and K. R. Choo, "HomeChain: A Blockchain-Based Secure Mutual Authentication System for Smart Homes," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 818-829, 2020.
4. C. Zhang *et al*., "BSFP: Blockchain-Enabled Smart Parking With Fairness, Reliability and Privacy Protection," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 6578-6591, 2020.
5. X. Liu, S. X. Sun and G. Huang, "Decentralized Services Computing Paradigm for Blockchain-Based Data Governance: Programmability, Interoperability, and Intelligence," *IEEE Transactions on Services Computing*, vol. 13, no. 2, pp. 343-355, 2020.
6. S. Seven, G. Yao, A. Soran, A. Onen, and S.M. Muyeen "Peer-to-Peer Energy Trading in Virtual Power Plant Based on Blockchain Smart Contracts" IEEE Access, vol. 8, pp. 175713-175726, 2020.
7. V. Jaiman, and V. Urovi, "A Consent Model for Blockchain-Based Health Data Sharing Platforms" , IEEE Access, vol. 8, pp. 143734-143745, 2020.
8. H. Guo, W. Li, M. Nejad and C. C. Shen, "Proof-of-Event Recording System for Autonomous Vehicles:A Blockchain-Based Solution", IEEE Access, vol. 8, pp. 182776-182786, 2020.
9. X. Yang, X. Yi, S. Nepal, A. Kelarev and F. Han, "Blockchain voting: Publicly verifiable online voting protocol without trusted tallying authorities", Future Generation Computer System, vol. 112, pp. 859-874, 2020.
10. Y. Yuan and F. Wang, "Towards blockchain-based intelligent transportation systems," *in IEEE 19th International Conference on Intelligent Transportation Systems*, pp. 2663-2668, Brazil, 2016
11. E. Mengelkamp, B. Notheisen, C. Beer, D. Dauer and C. Weinhardt, "A blockchain-based smart grid: towards sustainable local energy markets", Computer Science-Research and Development, vol. 33, pp. 207-214, 2018.
12. M. Turkanovic, M. Holbl, K. Kosic, M. Hericko, and A. Kamisalic, "EduCTX: A Blockchain-Based Higher Education Credit Platform", IEEE Access, vol. 6, pp. 5112-5127, 2018.
13. J. Kishigami, S. Fujimura, H. Watanabe, A. Nakadaira and A. Akutsu, "The Blockchain-Based Digital Content Distribution System," *IEEE Fifth International Conference on Big Data and Cloud Computing*, pp. 187-190, China, 2015.
14. M. Mettler, "Blockchain technology in healthcare: The revolution starts here," *IEEE 18th International Conference on e-Health Networking, Applications and Services, pp. 1-3,* Germany, 2016.
15. T. McGhin, K. R. Choo, C. Z. Liu, and D. He, "Blockchain in healthcare applications: Research challenges and opportunities", Journal of Network and Computer Applications, vol. 135, pp. 62-75, 2019.
16. C. C. Agbo , Q. H. Mahmoud and J. M. Eklund, "Blockchain Technology in Healthcare: A Systematic Review", Healthcare, vol. 7, no. 2, article no. 56, 2019.
17. A. A. Siyal et al., "Applications of Blockchain Technology in Medicine and Healthcare: Challenges and Future Perspectives", Cryptography, vol. 3, no.1, article no. 3, 2019.
18. C. Esposito, A. Santis, G. Tortora, H. Chang, and K. R. Choo, "Blockchain: A Panacea for Healthcare Cloud-Based Data Security and Privacy?", IEEE Cloud Computing, vol. 5, 2018.
19. W. J. Gordon, and C. Catalini, "Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability", Computational and Structural Biotechnology Journal, vol. 16, pp. 224-230, 2018.
20. T. Kumar, V. Ramani, I. Ahmad, A. Braeken, E. Harjula and M. Ylianttila, "Blockchain Utilization in Healthcare: Key Requirements and Challenges," *in IEEE 20th International Conference on e-Health Networking, Applications and Services, pp. 1-7,* Czech Republic, 2018.
21. L. Bell, W. J. Buchanan, J. Cameron, and O. Lo, "Applications of Blockchain Within Healthcare", *Blockchain in Healthcare Today*, vol. *1*., 2018.
22. A. Hasselgren, K. Kralevska, D. Gligoroski, S. A. Pedersen, and A. Faxvaag, "Blockchain in healthcare and health sciences—A scoping review", International Journal of Medical Informatics, vol. 134, article no. 104040, 2020.
23. X. Liang, J. Zhao, S. Shetty, J. Liu and D. Li, "Integrating blockchain for data sharing and collaboration in mobile healthcare applications," *IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications, pp. 1-5,* Canada, 2017.
24. A. Dwivedi, G. Srivastava, S. Dhar, and R. Singh, "A Decentralized Privacy-Preserving Healthcare Blockchain for IoT", Sensors, vol. 19, no. 2, article no. 326, 2019.
25. A. Ekblaw, A. Azaria, J. Halamka and A. Lippman, "A Case Study for Blockchain in Healthcare : " MedRec " prototype for electronic health records and medical research data", 2016.
26. A. Omar, M. S. Rahman, A. Basu, and S. Kiyomoto, "MediBchain: A Blockchain Based Privacy Preserving Platform for Healthcare Data", International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage, pp. 534-543, 2017.
27. G. Leeming, J. Cunningham, and J. Ainsworth, "A Ledger of Me: Personalizing Healthcare Using Blockchain Technology" , Frontiers in Medicine, vol. 6, 2019.

28. P. Bhattacharya, S. Tanwar, U. Bodke, S. Tyagi and N. Kumar, "BinDaaS: Blockchain-Based Deep-Learning as-a-Service in Healthcare 4.0 Applications," *IEEE Transactions on Network Science and Engineering*, 2019.
29. S. Jiang, J. Cao, H. Wu, Y. Yang, M. Ma and J. He, "BlocHIE: A Blockchain-Based Platform for Healthcare Information
30. Exchange," *IEEE International Conference on Smart Computing*, pp. 49-56, Italy, 2018.
31. A. Theodouli, S. Arakliotis, K. Moschou, K. Votis and D. Tzovaras, "On the Design of a Blockchain-Based System to Facilitate Healthcare Data Sharing," *17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications, pp. 1374-1379,* USA, 2018.
32. L. Ismail, H. Materwala and S. Zeadally, "Lightweight Blockchain for Healthcare," *IEEE Access*, vol. 7, pp. 149935-149951, 2019.
33. K. N. Griggs et al. "Healthcare Blockchain System Using Smart Contracts for Secure Automated Remote Patient Monitoring", Journal of Medical Systems, vol. 42, article no.130, 2018.
34. R. Kumar and R. Tripathi, "Traceability of counterfeit medicine supply chain through Blockchain," *11th International Conference on Communication Systems & Networks*, pp. 568-570, India, 2019.
35. I. Haq, and M. Olivier, "Blockchain Technology in Pharmaceutical Industry to Prevent Counterfeit Drugs", International Journal of Computer Applications, vol. 180, pp. 8-12, 2018.
36. I. Radonovic and R. Likic, "Opportunities for Use of Blockchain Technology in Medicine", *Applied Health Economics and Health Policy,* vol. 16, pp. 583–590, 2018.
37. M. Debe, K. Salah, R. Jayaraman and J. Arshad, "Blockchain-Based Verifiable Tracking of Resellable Returned Drugs," *IEEE Access*, vol. 8, pp. 205848-205862, 2020.
38. Z. Shae and J. J. P. Tsai, "On the Design of a Blockchain Platform for Clinical Trial and Precision Medicine," *IEEE 37th International Conference on Distributed Computing Systems*, pp. 1972-1980, USA, 2017.
39. N. Saxena, I. Thomas, P. Gope, P. Burnap and N. Kumar, "PharmaCrypt: Blockchain for Critical Pharmaceutical Industry to Counterfeit Drugs," *Computer*, vol. 53, no. 7, pp. 29-44, 2020.

AUTHORS PROFILE

Abhinav Sanghi, is currently pursuing Bachelor of Technology in Discipline of Mathematics and Computing, Delhi Technological University, India. His main areas of interest are Android development, Augmented Reality(AR), Machine learning, Blockchain etc. He has developed several android applications. He has done internship at Samsung Research and Development, Noida in the field of AR, android and developed an android app for Samsung future phones. Apart from that, he has qualified National Talent Search Examination in 2013.

Aayush, is currently pursuing Bachelor of Technology in the Discipline of Mathematics and Computing, Delhi Technological University, India. Currently he is working as a Software Developer Intern at OYO. He has a very good exposure and experience in the field of Data Analytics and he currently works as a server side engineer. His main areas of interest are Blockchain, Sequential Models, Natural Language Processing, API Development. Recently, he did his research internship at Delhi Technological University, in the field of video tampering detection using Blockchain. He also has a good understanding of event streaming platforms, integrations and has very thorough knowledge of Software Development Life Cycle and itsphases(SDLC).

Ashutosh Katakwar, is currently pursuing Bachelor of Technology in Mathematics and Computing from Delhi Technological University. He has strong interest in mathematics and programming, and as part of his academics and curriculum, have developed proven analytical, problem-solving and computing skills through his projects and internship. He is proficient in python, Java, SQL, VBA, and have good experience on projects related to Machine Learning, Data based models and Financial models.

Anshul Arora, is currently working as Assistant Professor in Discipline of Mathematics and Computing, Delhi Technological University Delhi, India. He has pursued Masters and Ph.D. from Department of Computer Science and Engineering, Indian Institute of Technology Roorkee, India. His areas of research include Mobile Security, Mobile Malware Detection, Network Traffic Analysis, and Blockchain.

Aditya Kaushik, is currently working as Associate Professor in Discipline of Mathematics and Computing, Delhi Technological University, Delhi, India. He has pursued M.Sc. NET, Ph.D. P.Doc. from renowned colleges. His areas of research are Differential equations, numerical analysis, on-linear diffusion equations, fluid dynamics, applied mathematics, finite element methods, numerical mathematics and error estimation etc.

JMEPEG
https://doi.org/10.1007/s11665-023-08043-9

©ASM International
1059-9495/$19.00

TECHNICAL ARTICLE

Development of ZrB_2-SiC Plasma-Sprayed Ceramic Coating for Thermo-chemical Protection in Hypersonic Vehicles

Kalpana Gupta, Qasim Murtaza, and N. Yuvraj

Submitted: 21 December 2022 / Revised: 14 February 2023 / Accepted: 18 February 2023

The aim of this research is to explore the potential of ZrB_2-SiC-based ceramic coatings as a protective layer for Inconel 718 substrates. The study focuses on the use of shrouded plasma spraying technology to deposit the coating and investigates the effect of exposure to molten sulfate-vanadate salts (45% Na_2SO_4 and 55% V_2O_5) on its properties. The tribological behavior and corrosion resistance of the coated samples were evaluated, and the results revealed that the ZrB_2-SiC coating showed resistance to hot corrosion. Upon exposure to the molten salt mixture, the ZrB_2-SiC coating was found to interact with V_2O_5 and Na_2SO_4, forming binary oxide phases, such as $NaVO_3$ and SiO_2. The formation of the ZrO_2 monoclinic phase was observed due to the leaching of B_2O_3 from the ZrB_2-SiC coating. However, the physical properties of ZrB_2-SiC were found to prevent the penetration of the molten salt, thus reducing its corrosive effect. Overall, the results indicate that ZrB_2-SiC-based ceramic coatings have the potential to offer effective protection to Inconel 718 substrates against hot corrosion in harsh environments.

Keywords hypersonic vehicles, molten salt, shrouded plasma spraying, silicon carbide, tribology, wear mechanism, zirconium diboride

1. Introduction

Hypersonic vehicle components must perform under high temperatures, such as scramjet nozzles, wing leading edges, and nose tips. A protective surface coating is recommended to safeguard these parts against high-temperature oxidation and a corrosive environment (Ref 1). The remarkable ability of ultra-high temperature ceramics (UHTCs) to withstand harsh environments has led to widespread use. Among various UHTC materials, ZrB_2 is the most often studied UHTC material because of its superior thermal conductivity (58.2 W/mK), low coefficient of thermal expansion (5.82×10^{-6} °C^{-1}), melting temperature (3500 K), higher electrical conductivity, low-density (6.09 g/cm^3), strong thermal shock resistance, and high hardness (Ref 2, 3).

Because of the properties mentioned above, ZrB_2-based protective layers [like a thermal protection system (TPS)] have received much attention in recent years. Thermal barrier coating solutions based on ZrB_2 are often used. In recent years, ZrB_2-based coatings have attracted much attention as it protects against oxidation at high temperatures by acting as ablation-resistant coatings. Above 800 °C, it has been observed that ZrB_2 is rapidly oxidized, producing B_2O_3 and ZrO_2. This B_2O_3 forms a protective liquid layer (above 450 °C) on the surface, which restricts oxygen diffusion. However, at above 1200 °C B_2O_3, get starts evaporating severely and no longer protects the matrix. It has been investigated that the amalgamation of ZrB_2 with SiC increases the oxidation resistance at 1000-1800 °C by producing less volatile boro-silicate glass with reduced oxygen permeability (Ref 3). As a result, they've been employed in high-temperature situations to endure oxidation, corrosion, and wear. ZrB_2-SiC ceramics gradually oxidize at 800-1200 °C. The oxidation of SiC generates a SiO_2 film on the coating surface as the temperature rises, which can further avoid oxygen entry into the matrix (Ref 4, 5). Simultaneously, $ZrSiO_4$ may be produced by reacting ZrO_2 and SiO_2, which improves the coating oxidation resistance. Incorporating a SiC ceramic into ZrB_2-based ceramics can considerably improve their oxidation behavior within the moderately elevated temperature, with optimal SiC content ranging between 15-20 vol.%. It is found that 20 wt% compositions of SiC is best suited for hypersonic vehicles in test conducted by US Air Force (Ref 6).

In the past, several studies investigated the oxidation resistance of ZrB_2-based ceramics at higher temperatures. The characteristic scale formed during oxidation of a ZrB_2-SiC ceramic has been proven to consist of four different layers: an unaffected layer, a SiC-depleted layer, a SiO_2 ironic smooth film, and a SiO_2-ZrO_2-layer (Ref 10, 11). Researchers evaluated the properties of reinforcement, sliding speed and load on the tribology of ZrB_2based composites. Shirshendu et al. studied the scratch resistance of ZrB_2-TiB_2 composites counter to diamond and discovered that ZrB_2-TiB_2 had superior wear resistance because of greater fracture toughness and hardness (Ref 7). Jitendra et al. examined tribological behavior and found a decrease in COF as well as an increased tendency for the formation of tribochemical films (Ref 8). Despite this, few studies in the literature concentrate on the oxidation and wear behavior of atmospheric plasma spray ZrB_2-based coatings. As

Kalpana Gupta, Qasim Murtaza, and **N. Yuvraj**, Mechanical Engineering, Delhi Technological University, Delhi, India. Contact e-mail: kalpanaguptapandey@gmail.com.

Published online: 16 March 2023

a result, the engineering and scientific research community is working to devise ways for impurity infiltration resistance to enhance the service life of UHTC components in presence of high temperatures or different types of fuel. Still, it is needed to research the sliding wear and hot corrosion characteristics of the ZrB_2-SiC coating contrary to molten corrosive salt, notably Na_2SO_4 and V_2O_5, at extreme temperatures of 900 °C (Ref 9). As a result, it is clear that the obstacles to preventing molten salt flow in ZrB_2-SiC still exist in creating efficient hypersonic vehicles or UHTC components.

This study intends to establish ZrB_2-SiC coatings over Inconel-718 material using a shrouded plasma spraying technique. Inert gas (shroud gas like nitrogen and argon) develops an envelope around the plasma flame and minimizes the chance of oxidation. Shrouding gas helps to shield the inflight particle so that it will not make direct contact with the atmosphere, preventing it from oxidizing the coating (Ref 9). This is the first study on the influence of ZrB_2-SiC coating on hot corrosion behavior and the total mitigation of molten corrosive salt penetration in ZrB_2-SiC coating. The coating will be subjected to Na_2SO_4 + V_2O_5 for 30 h at a high temperature (1600 °C). The ZrB_2-SiC coating is thought to show an essential part in molten salt penetration mitigation by operating as a physical barrier. Furthermore, ZrB_2-SiC can interact thermally and chemically with the molten salts, altogether preventing molten salt penetration and strengthening the root. This research aims to demonstrate the oxidation behavior of the shrouded plasma-sprayed ZrB_2-SiC-based coatings and elucidate the density impact on ZrB_2-SiC oxidation resistance). The role of ZrB_2-SiC in the mitigation of in dry unlubricated states of sliding against counter body WC ball and corrosive molten salt (55 wt.% V_2O_5+ 45 wt.% Na_2SO_4) intrusion will be explored.

2. Materials and Methods

The SiC (700-800 nm) and ZrB_2 (10 μm) were supplied by Trixotech Advanced Materials Pvt. Ltd. (India). The powder compositions of ZrB_2 80 vol.%-SiC 20 vol.% were obtained by mixing ZrB_2 and SiC. To agglomerate the mixture, a spray-drying method was used, resulting spherical grain powder with the mean particle size distribution in between 40-60 μm, suitable in thermal spray deposition. A Field Emission Scanning Electron Microscope (FESEM) made by Zeiss, Sigma H.D., was used for the study of powder, coating and fracture surface morphology, also EDS analyzer was coupled. In contrast, x-ray diffractometer (XRD) (model: TTRAX III, made by Rigaku, Japan) was used to examine phase composition using Cu-K_α radiation (energy is 8.04 keV and λ = 0.15 nm) having scan rate of 2°/min (step size = 0.02 degree) in 2θ range between 20° and 80°.

The mixed powders were used to deposit coating on Inconel-718 substrates using the shrouded APS technique. The primary plasma and powder carrier gas are argon, whereas hydrogen is employed as the secondary gas. For the controlled atmosphere, an inert atmospheric shroud is placed at the forward-facing of the 9 MB plasma gun as depicted in Fig. 1. Several experiments were carried out, and the process parameters for coating samples are determined and represented in Table 1.

Helium gas pycnometer (Model: 1200e, Country: USA) was used to determine the coating true densities with 0.689 bar pressure as the exit gas. The theoretical density of ZrB_2 and SiC are 6.085 g/cm^3 and 3.217 g/cm^3, respectively (Ref 10). The theoretical density of the composites was determined using the rule of mixtures. The tribological test wear conducted in ball on disk tribometer using WC ball, for 3600 s, speed 250 rpm, normal load 5 N and wear track diameter of 6 mm. The classical Archard's equation is utilized for determination of specific wear rate.

A mixture of Vanadium Oxide (V_2O_5) and Sodium Sulfate salt (Na_2SO_4) with 55 wt.% and 45 wt.% respectively was prepared as a corrosive agent is shown in schematic Fig. 2. The specification of each salt is listed in Table 2. The powders of Na_2SO_4 and V_2O_5 were blended in a planetary ball mill for a period of 4 h using Zirconia balls as a mixing medium (ZrO_2 ball to powder ratio 2:1).

The ZrB_2-SiC coating was preheated at 250 °C using high temperature chamber furnace for 1 h before applying the Na_2SO_4-V_2O_5 mixture on the surface of coating, to achieve better adhesion with salts on the coating surface. A quantity of 5 gm/cm^2 Na_2SO_4-V_2O_5 salt was evenly distributed on the coating's surface using a glass fibre brush. Then, furnace was used to dry these salt-containing coatings for 1 h at 250 °C, to remove moisture from the deposited salt. The sample was placed inside a furnace set to 1600 °C in an air-filled furnace for 30 h, and the coatings were removed for characterization after cooling to ambient temperature.

3. Results and Discussion

3.1 Microstructural and Tribological Investigation of Powder and Coatings

The high and low magnification FESEM images of the ZrB_2-SiC powder, which is utilized in the production of plasma-sprayed coating, are shown in Fig. 3(a) and (b). All the spherical powder agglomerates had an average size of 40-70 μm. Figure 3(a) depicts the SiC particles in the ZrB_2 matrix using yellow line arrows. The surface of the spherical agglomerates exhibited a coarse structure because of the asymmetrical morphology and sharp corners of the base material that made up the entire agglomeration.

To examine the elemental compositions in the powder, an elemental mapping analysis was conducted. The high-resolution FESEM images in Fig. 4(a) give us an in-depth look at the ZrB_2-SiC powder's microstructure. This analysis went a step further by utilizing elemental mapping techniques (presented in Fig. 4b-e), which allowed us to study the distribution of the key elements Zr, B, Si, and C within the sample. The findings of this analysis are extremely important, as they provide evidence that the elemental ratio of the ZrB_2-SiC powder is consistent with the desired composition. This can be seen clearly in Fig. 4(f), where the elemental ratio is plotted and compared to the desired specifications. This information can also be used to optimize processing conditions and further improve the properties of the material.

The XRD spectrum of the resulting ZrB_2-SiC powder and its coating is revealed in Fig. 5. Coating depicted significant right shifting of the peak as compared to the powder due to the lattice contraction in crystal during solidification of molten particles on the substrates. According to the JCPDS, the major peaks in ZrB_2-SiC coating are ZrB_2 (α), SiC (γ), and a trace of ZrO_2 (β):

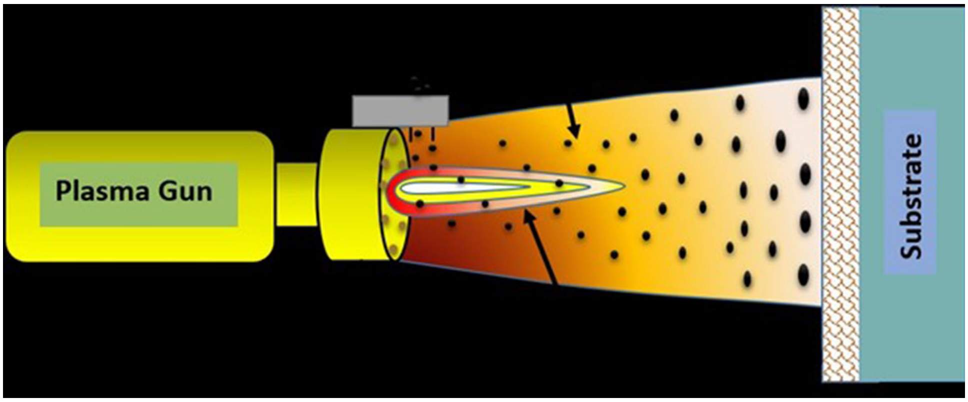


Fig. 1 Schematic of the shrouded plasma spraying setup

Table 1 Optimized process parameters used to fabricate plasma-sprayed ZrB_2-SiC coatings

Plasma	Current	Voltage	Standoff	Primary	Feed	Secondary	Shroud
Spray Parameters	(A)	(V)	Distance (in mm)	Flow, Argon (slpm)	Rate (g/min)	flow, Hydrogen (slpm)	Gas Pressure (psi)
Value	506	60	110	54	12	5.1	40

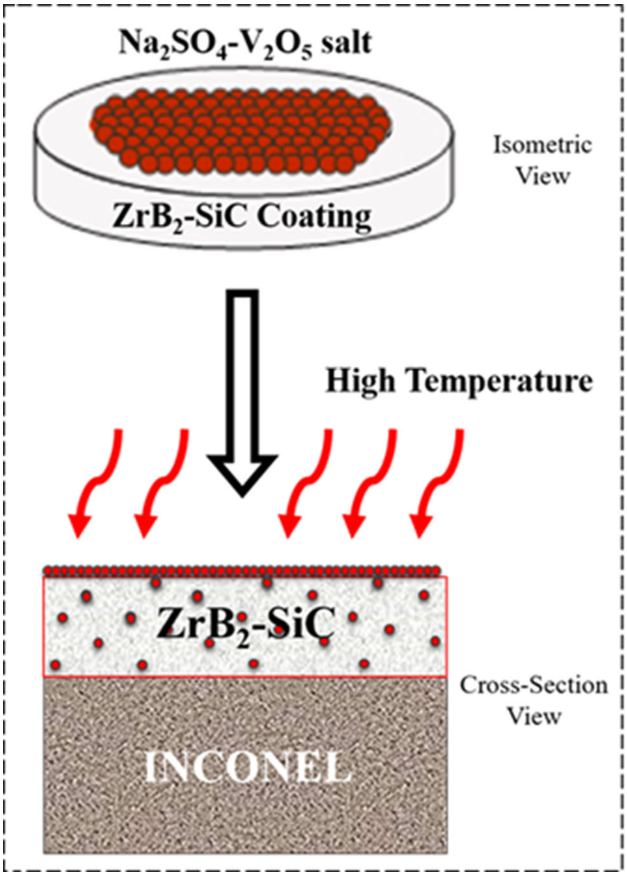


Fig. 2 Schematic of the penetration mechanism of molten Na_2SO_4-V_2O_5 salt through ZrB_2-SiC coating under high temperature environment

Table 2 Physical and thermal properties of the corrosive salts studied in this work

Corrosive salts	Na_2SO_4	V_2O_5
Manufacturer	Merck (Germany)	Merck (Germany)
Melting point, °C	888	690
Density, g/cm^3	2.70	3.36

(00-034-0423) for hexagonal ZrB_2, (00-048-0708) for hexagonal SiC, (00-036-0420) for monoclinic ZrO_2 (m-ZrO_2), and (00-024-1164) for tetragonal ZrO_2 (t-ZrO_2). t-ZrO_2 was identified in the sprayed coating because a portion of the t-ZrO_2 might not be changed to m-ZrO_2 during the quick cooling from elevated to ambient temperature. XRD was unable to identify SiO_2 due to its low concentration and amorphous nature. Furthermore, the reduced SiC intensity may be caused by disintegration of the SiC ceramic and oxidation at high temperatures. According to this, the major phase in the as-prepared coating is ZrB_2 (Reference code: 01-089-3930). In addition, comparable SiC peaks (Reference code: 00-022-1317) could be observed in the XRD pattern. In contrast, (00-036-0420) for monoclinic ZrO_2 (m-ZrO_2) and (00-024-1164) for tetragonal ZrO_2 (t-ZrO_2). However, XRD patterns indicate that the coatings retained ZrB_2 and SiC phases.

The chemical element analysis using EDS shows an amount of silicon that is consistent with the initial powders, depicted in Table 3.

The cross-sectional representation of thickness of the coating is shown in Fig. 6(a) which is found in between 300-400 μm, and are devoid of defects and demonstrated greater integrity with substrate whereas, Fig. 6(b) shows the theoretical and measured density of the desired coating. The measured

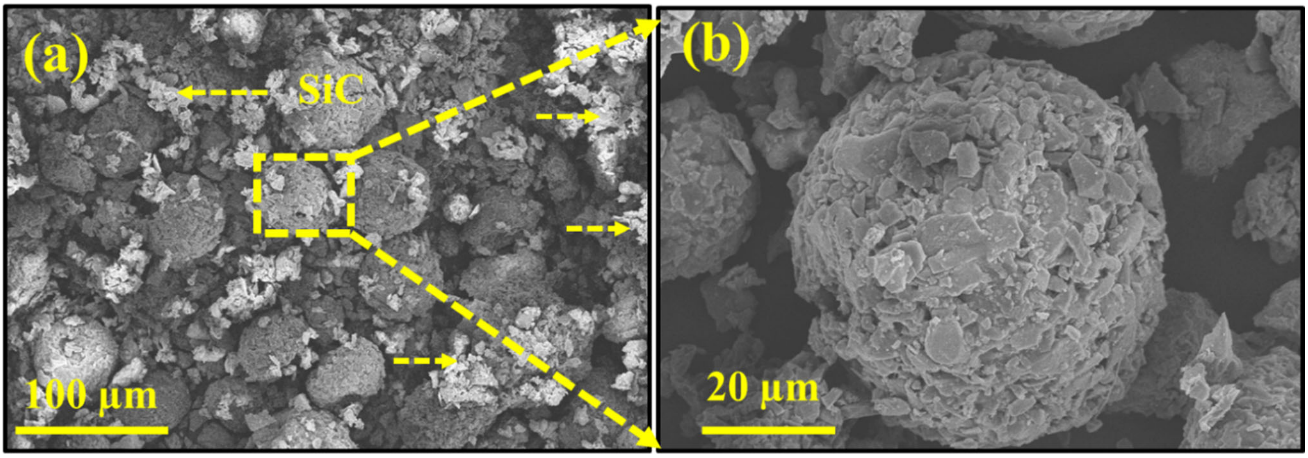


Fig. 3 (a-b) FESEM images of the ZrB_2-SiC powder at lower and higher magnification, demonstrating spherical shape

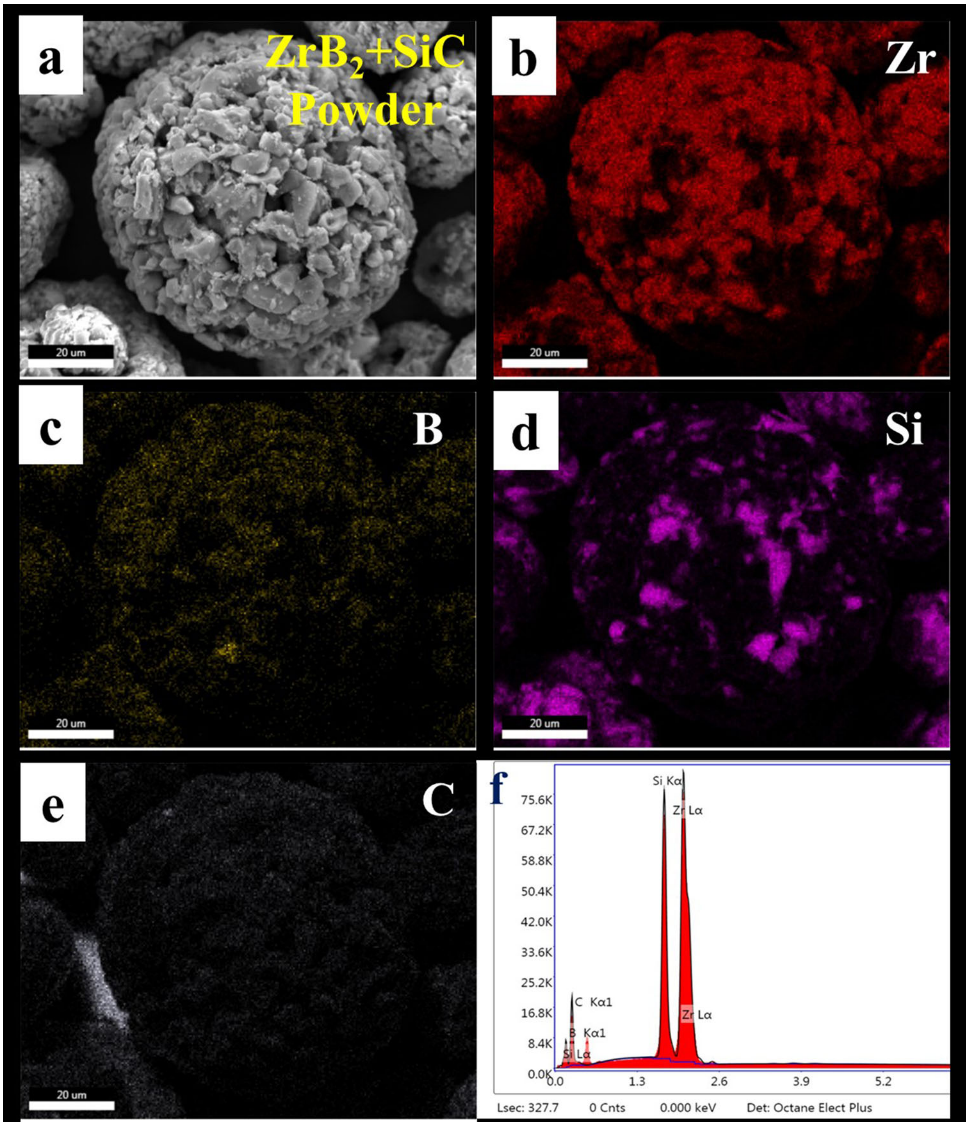


Fig. 4 (a) FESEM image of ZrB_2-SiC feedstock powder, (b-e) EDS mapping is showing the EDAX spectrum quantifying presence of Zr, B, Si, and C elements

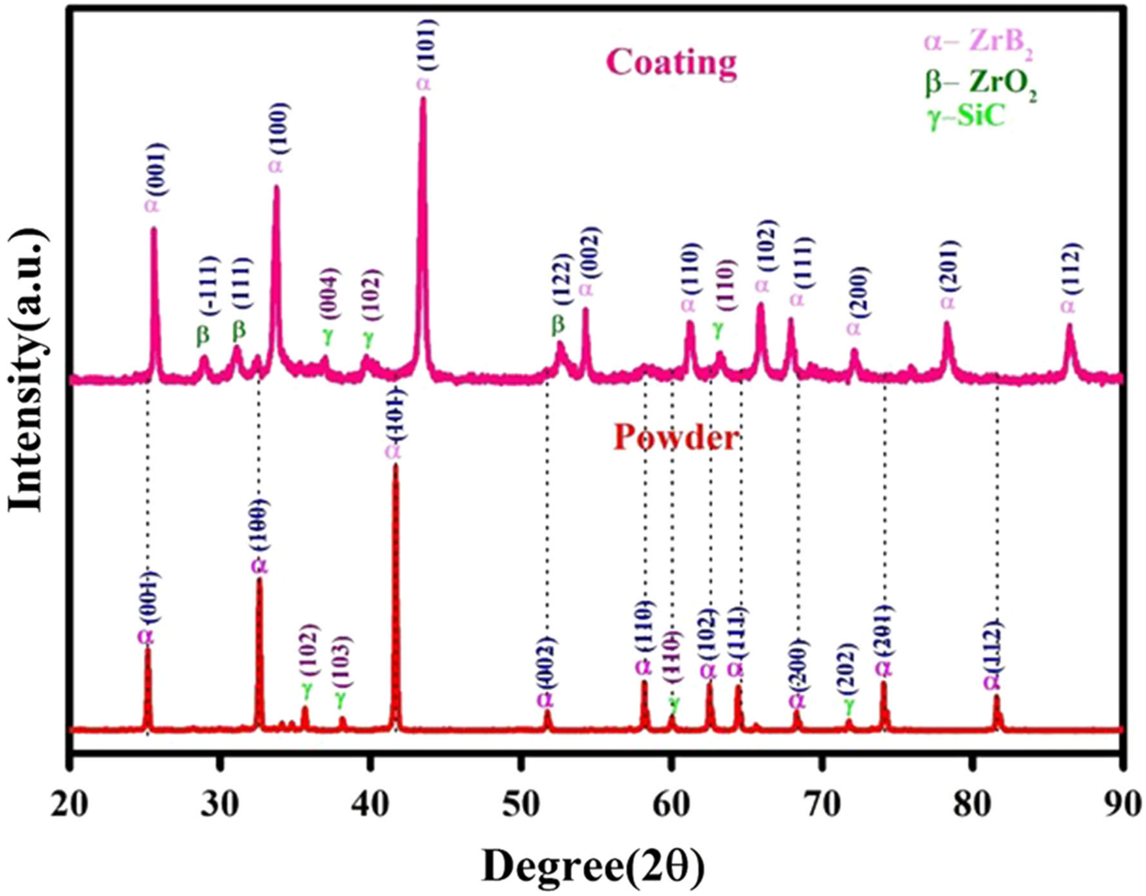


Fig. 5 X-ray diffraction (XRD) spectra of the ZrB_2-SiC powder and corresponding coatings

Table 3 Chemical elements were determined using an energy-dispersive spectrometer (EDS) analysis

S. no.	Chemical composition	Zr, wt.%	Si, wt.%
1	ZrB_2 + SiC powder	55.35	12.28
2	ZrB_2 + SiC coating	49.72	10.67

density of composite coating found to be 86.4 ± 1.3% of the theoretical density. Since relative density and porosity are inversely related, knowing the proportion of relative density allows us to determine that coating have a porosity of ~ 13%. Figure 6(c) represents the digital image of the shroud plasma-sprayed ZrB_2-SiC coating deposited over Inconel-718 substrate. In Fig. 6(d), The ZrB_2-SiC coating has a bi-modal morphology, with a moderately melted zone with a sponge-like microstructure and a completely molten zone bonded to create dense structure, as shown in high magnification FE-SEM image. The covering is devoid of significant fractures and adheres to the substrate. Micro-cracks were seen on melted zone of the coated surface, which might be attributed to shrinkage and thermal strains caused by quick solidification during plasma spraying. By intensifying corrosion process between the liquid salts ZrB_2 and SiC, the semi-molten particles and microcracks function as diffusion channels to the molten salts.

Figure 7(a) and (b) shows FESEM images of fragmented surface of as-sprayed ZrB_2-SiC coating, which exhibited a characteristic splat like microstructure mainly made of overlying lamellae delimited from splat boundaries and enclosed from a link of micro cracks (Fig. 7a). Thermal conductivity and heat transmission are commonly affected by splat boundaries. Tension relaxation during fast cooling causes horizontal microcracks to emerge at splat edges. Splat boundaries form as a result of the weak connection between deposited splats caused by the collision of fast solidification and their molten droplets.

Figure 8 displays COF (Fig. 7a), wear rate and wear volume loss (Fig. 8b) of bare substrate and ZrB_2—SiC coating, respectively. This can be observed that ZrB_2-SiC coating drastically reduces COF as well as wear volume loss. The average COF decreases from 0.57 to 0.3 for bare substrate and ZrB_2-SiC coating, respectively. The wear volume loss of bare Inconel-718 and ZrB_2-SiC was calculated to be 2.63 ± 0.5 and 1.90 ± 0.8 mm^3, respectively. As demonstrated in Fig. 7(b), a significant reduction in wear rate was seen in ZrB_2-SiC (2.02 ± 0.03 10^{-3}/N-m) coating compared to Inconel-718 (2.8 ± 0.04 10^{-3}/N-m) bare substrate. Generally, hardness and toughness are essential parameter influencing the underlying tribological mechanism of most materials. These findings are consistent with the Archard's equation for abrasive wear, specifies that the wear rate will be lower for harder material. Hence, the formation of wear debris will be low for coating of increased hardness (ZrB_2-SiC) which lead to better wear resistance and lowest COF than bare substrate. The incorpo-

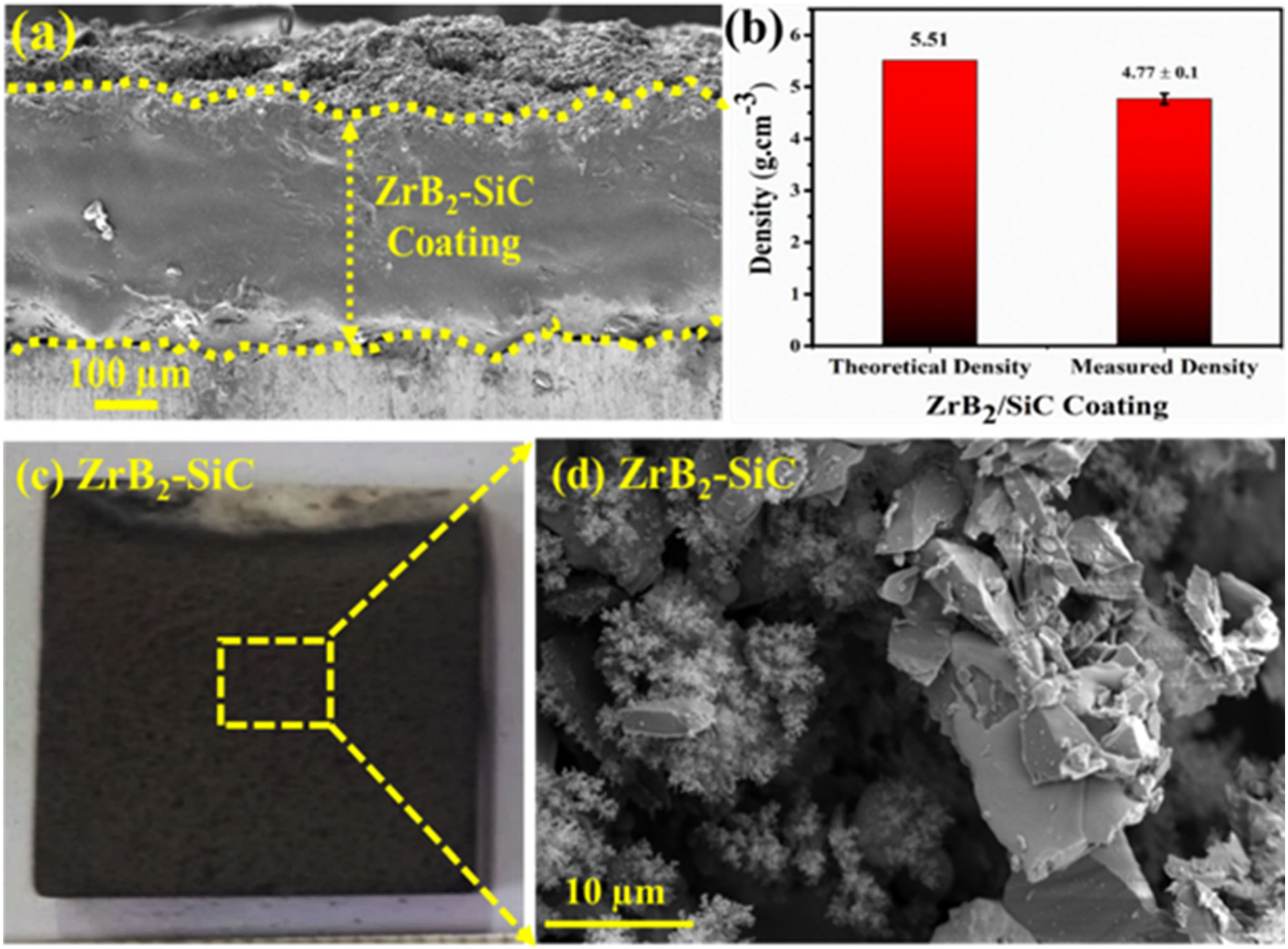


Fig. 6 (a) Cross-sectional FESEM image of shrouded plasma-sprayed ZrB_2/SiC coating, (b) shows comparative plot of theoretical and measured relative density of the ZrB_2-SiC coating, (c) digital image of plasma-sprayed coating on Inconel-718 substrate, and (d) high magnification FESEM image of ZrB_2-SiC coating

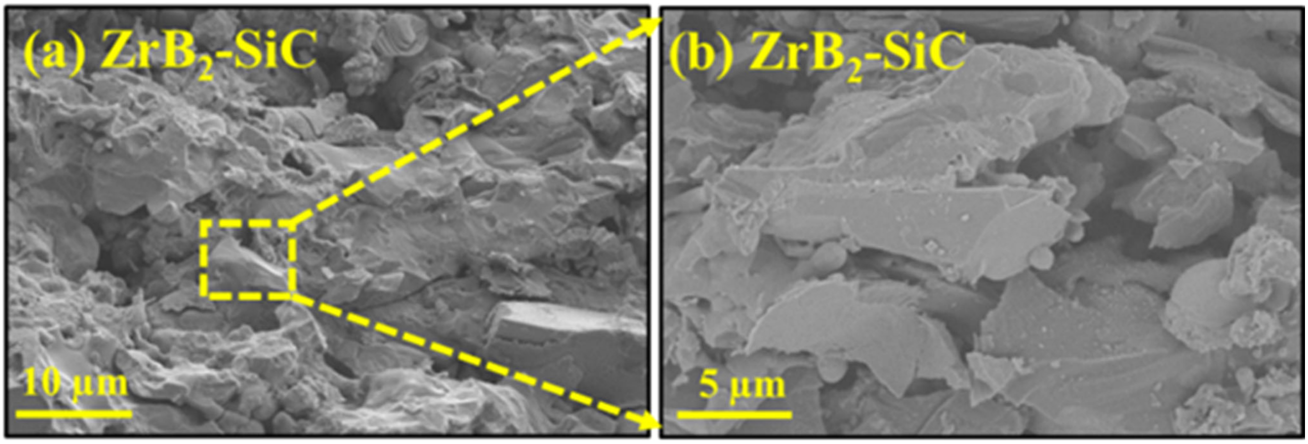


Fig. 7 (a-b) High-magnification FESEM images of fragmented cross section of shrouded plasma-sprayed ZrB_2-SiC coatings

ration of SiC in ZrB_2 also enhances the tribological properties of the composite. Apart from that, studies also shown that the ceramics consisting SiC reacts with the moisture in the air and silica/hydrated silica reaction film is produced. These formed tribo films partially covered the wear track which further prevents wearing of the surface.

3.2 Characterization of Sample After Hot Corrosion

Figure 9(a) illustrates a digital image of the ZrB_2-SiC coating prior to hot corrosion testing. Figure 9(b) and (c) illustrates the morphology of the plasma spraying coatings after 30 h of exposure to corrosive salts ($V_2O_5 + Na_2SO_4$) at 1600 °C. After hot corrosion with $V_2O_5 + Na_2SO_4$ at

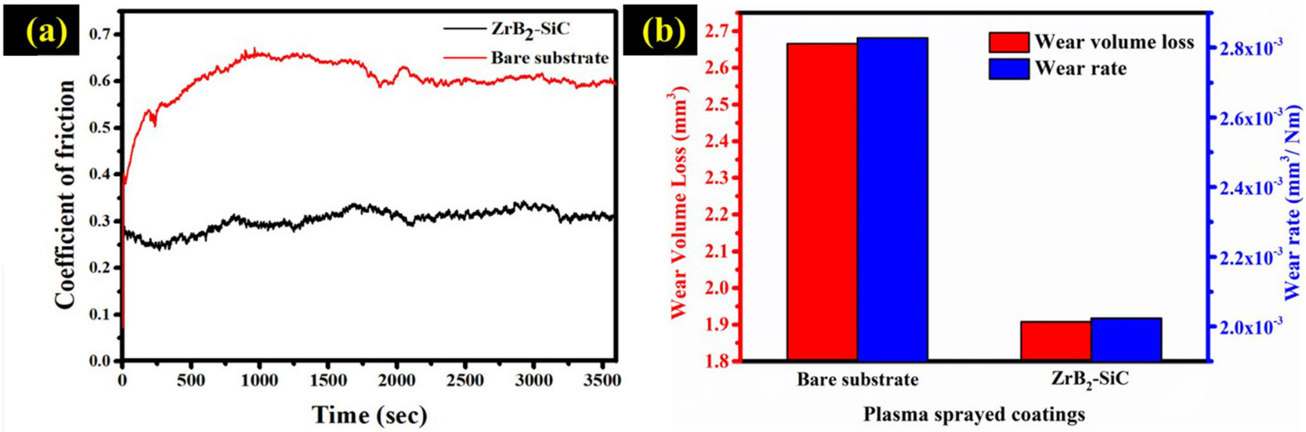


Fig. 8 (a) COF of bare substrate and ZrB_2-SiC coating (b) Graph depicting wear volume loss along with wear rate

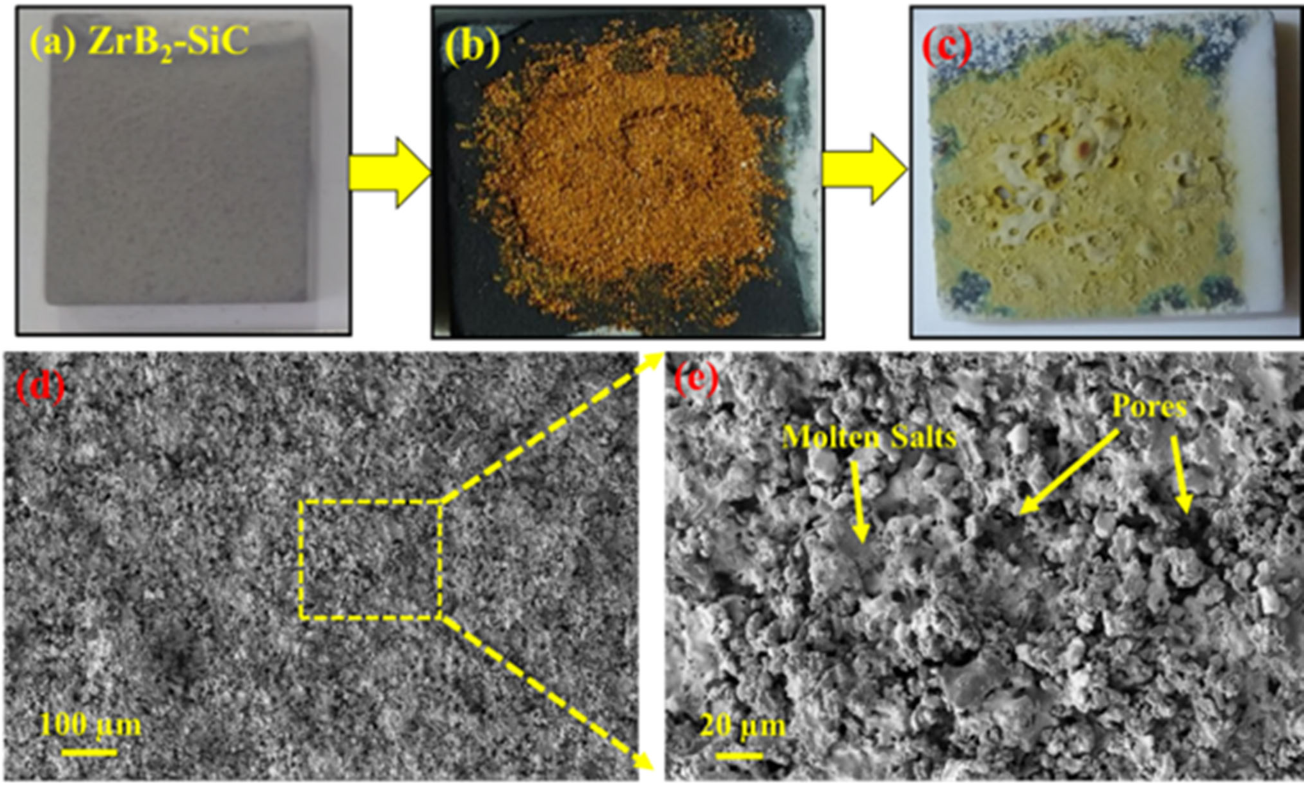


Fig. 9 (a) Digital picture of the ZrB_2-SiC coating before hot corrosion test, (b) digital image shows V_2O_5 + Na_2SO_4 salt onto the ZrB_2-SiC coating, and (c) digital image shows the post-hot corrosion on coating, (d) FESEM images shows the surface of ZrB_2-SiC coating after hot corrosion test, and (e) high-magnification image of marked area (Fig. 10d), arrows indicating pores and molten salts of $NaSO_4$ and V_2O_5 deposited and corroded over the surface

1600 °C, ZrB_2-SiC is converted to monoclinic ZrO_2. SiO_2 is generated from SiC, and a minor amount of SiC reflection is noticed. The coated surface is covered by corrosion products, as exposed in Fig. 9(c). Figure 9(d) and (e) depicts the topography of a plasma-sprayed ZrB_2-SiC coating surface after 30 h of exposure to corrosive salts (Na_2SO_4-V_2O_5) at 1600 °C. It was observed that several cylindrical-shaped and crystal dendritic were formed over the surface of the ZrB_2-SiC coating. The surface morphology of ZrB_2-SiC coatings also showed a porous surface with minor cracks in addition to the formation of cylindrical-shaped and crystal dendritic.

Figure 10(a) represents the top surface of the coating prior to hot corrosion, where three distinct regions are identified, and Fig. 10(b) demonstrates the EDAX spectrum with verified elements, namely Zr, B, Si, and C. Figure 10(c) depicts the FESEM of the top surface of the coating after hot corrosion. In this case, three distinct areas are taken to validate the corrosion area with the corrosive materials. Figure 10(d) depicts the EDAX scanning of the coating's surface after hot corrosion at three distinct stages, confirming the elements Zr, B, Si, C, Na, S, O, and V after reacting with the corrosive salts.

From the experiment, it is observed that at higher temperature, the salts of Na_2SO_4-V_2O_5 were melted and reacted with

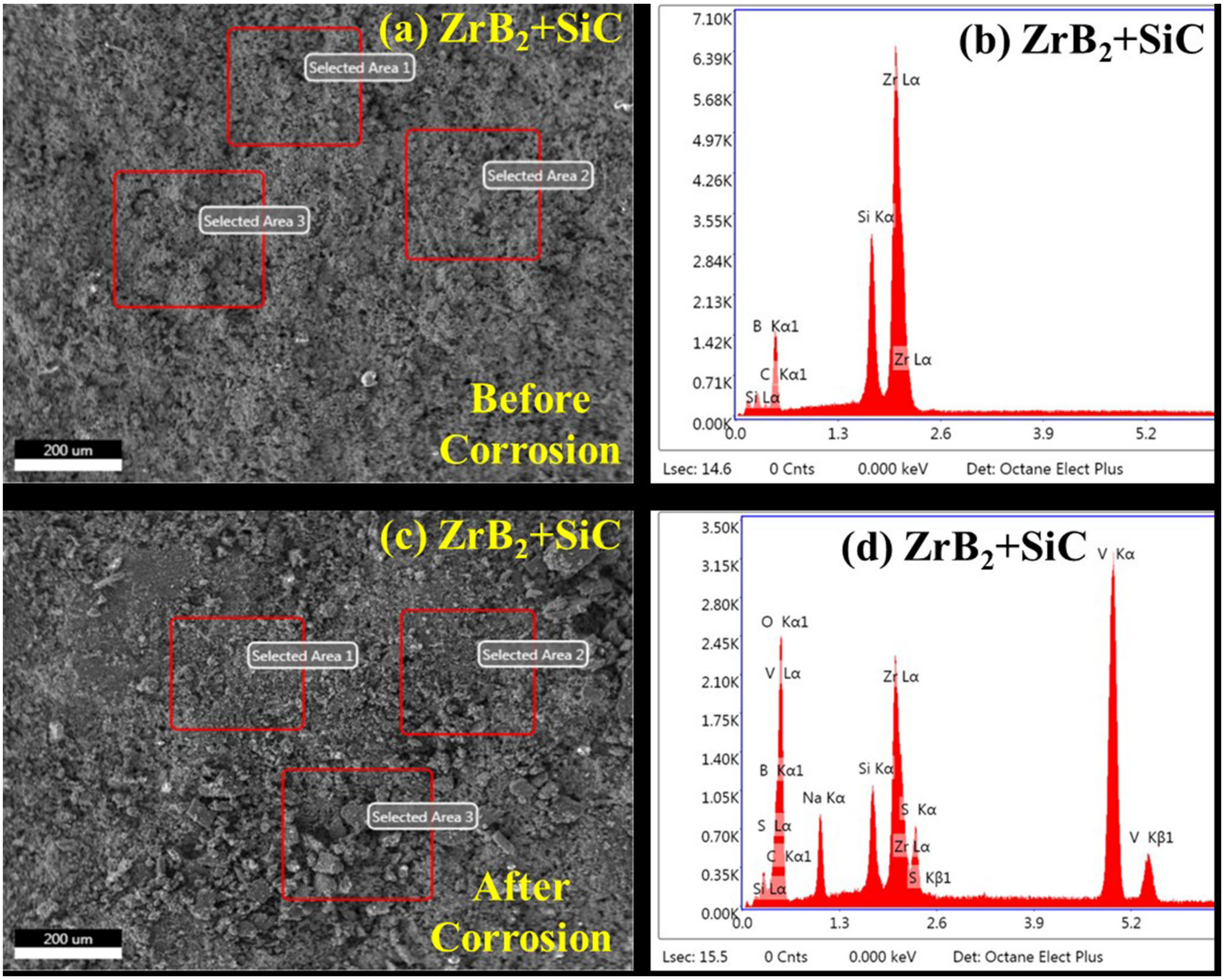


Fig. 10 (a-c) FESEM images of the ZrB_2-SiC coating depicts, and (b-d) shows the EDAX spectrum before and after corrosion, respectively

ZrB_2/SiC which leads to oxide's formation such as ZrO_2 and SiO_2. These oxides act as protective layer and stop further reaction of salts with coating.

To better recognize, V_2O_5 may be melted first in the mixed sulfate–vanadate salt because of the relatively low melting point (690 °C). The reaction involving in formation of $NaVO_3$ and Na_3VO_4 is summarized in Eq. 3, 4 and 5.

$$V_2O_5 + Na_2SO_4 = SO_3 + NaVO_3 \qquad \text{(Eq 1)}$$

$$V_2O_5 + Na_2O = 2NaVO_3 \qquad \text{(Eq 2)}$$

$$V_2O_5 + 3Na_2O = 2Na_3VO_4 \qquad \text{(Eq 3)}$$

When the ZrB_2-SiC coating was in direct contact to V_2O_5-Na_2SO_4 molten corrosive salts, a molten mixture of Na_2O-$NaVO_3$-Na_2SO_4-V_2O_5 -Na_3VO_4 likely to cover the coating's surface because these reactions might not be completed entirely and their reversibility should be addressed. This combination might have entered the coating through the built-up pores, resulting in micro-cracks over the coated surface from the plasma spraying process.

Earlier studies of Hakon Flood and Hermann Lux demonstrated that when acid and base react with inorganic salts can be shown as acidic by the accepter of oxide ion (O.$^{2-}$) and basic by the donor of oxide ion; this base-acid concept could be interpreted as a kinetic revival of the oxygen theory for bases and acids: (Ref 11)

$$O^{2-} + \text{acid} = \text{Base} \qquad \text{(Eq 4)}$$

Or else,

$$SO_3(\text{acid}) + Na_2O\ (\text{base}) = Na_3SO_4(\text{salt}) \qquad \text{(Eq 5)}$$

The molten salts of concern in the current investigation were largely sodium meta-vanadate and sodium ortho-vanadate, where sodium meta-vanadate, for illustration [chemical reaction (7 and 8)], may stand as per

$$Na_2O(\text{base}) + V_2O_5(\text{acid}) = 2NaVO_3(\text{salt}) \qquad \text{(Eq 6)}$$

As a result, the thermodynamically acceptable processes that occur during plasma spraying coating are as follows:

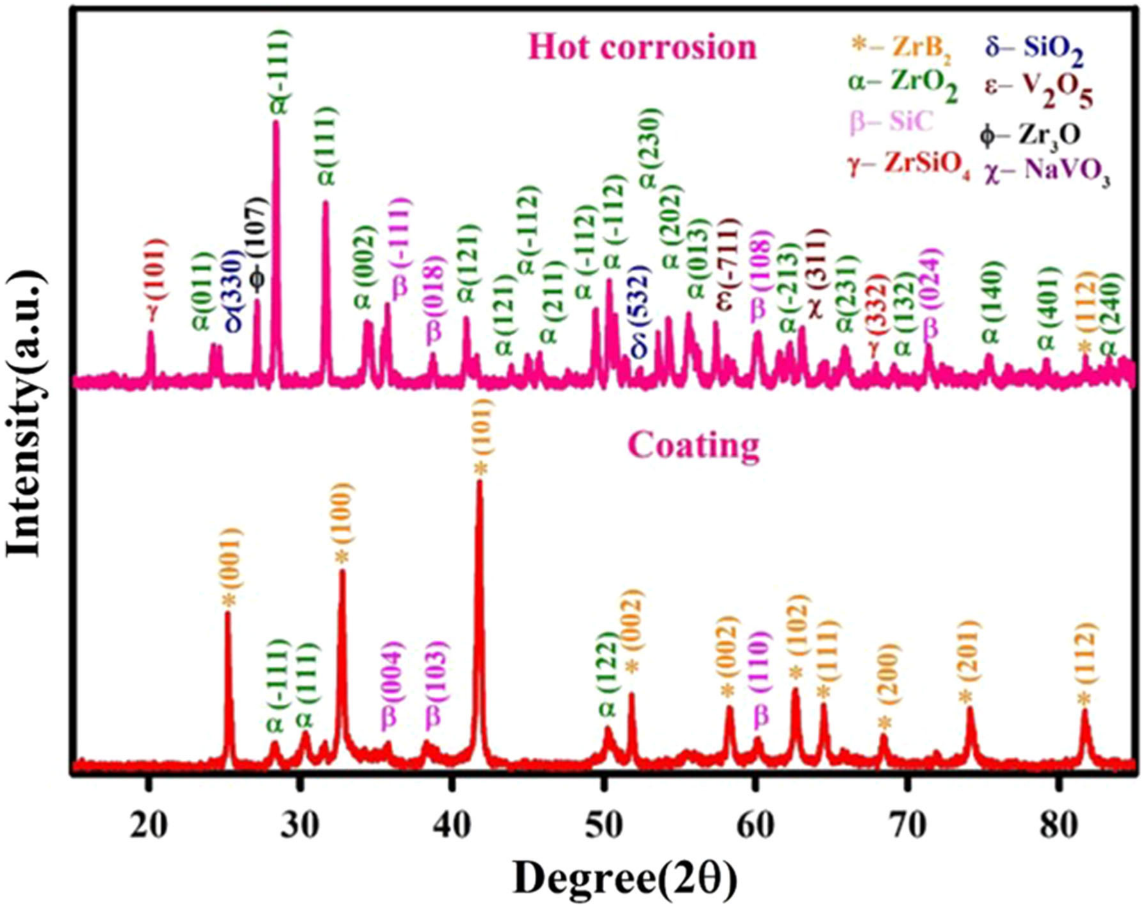


Fig. 11 XRD spectra of the ZrB_2-SiC plasma-sprayed coating before and after molten salt test

$$ZrB_2 + SiC + NaVO_3 \rightarrow ZrO_2 + Na_2CO_3 + SiO_2 + V_2O_5 + B_2O_3 \qquad \text{(Eq 7)}$$

$$4B_2O_3 + 4ZrO_2 + 10NaVO_3 + 5\ SiC = 4\ ZrB_2 + 2NaO + 3CO_2 + 5SiO_2 + 5V_2O_5 \qquad \text{(Eq 8)}$$

As previously stated, the interaction of the melted corrosive salts with ZrB_2-SiC led to the loss of ZrO_2 and SiO_2, leading to the formation of a ZrO_2 crystalline phase which caused the tetragonal ZrO_2 to become unstable and transition for the aforementioned stable monoclinic (m) ZrO_2 crystalline structure. The diffusionless, martensitic transition of tetragonal—t′ to monoclinic—m ZrO_2 occurs via a volume expansion of roughly 2-6%, resulting in spallation and cracking of the ZrO_2-SiC coating. When the volume expansion pressures are relieved, they may cause fractures in the coating and, as a result, coating failure.

$$(ZrO_2)_{tetragonal} = (ZrO_2)_{monoclinic} \qquad \text{(Eq 9)}$$

Thus, the interactions of these ceramics with vanadium salts are primarily evaluated using Lux Flood type of acid base reaction which are explicable on basis of the oxides' comparative acid—basic nature. The least affected stabilizing oxides are those that must be acidic to interact with V_2O_5 and $NaVO_3$.

XRD pattern of the SiC-ZrB_2 coating before and after the molten salt (hot corrosion) test is shown in Fig. 11. After the hot corrosion experiment, zirconia was detected in both monoclinic and tetragonal phases, despite the fact that the as-sprayed coating only contained the tetragonal phase. In addition to these phases, a few SiO_2 and V_2O_5 peaks were also seen in corroded ZrB_2-SiC coating. This finding validated the development of SiO_2 crystals on top of the corroded ZrB_2-SiC coating. However, the corroded ZrB_2-SiC coating's XRD spectra revealed sizable ZrB_2 peaking. Moreover, the SiC starting and ZrB_2 phases, binary phases like SiO_2, SiC, and ZrO_2 reaction products, can be found in the ZrB_2-SiC plasma-sprayed coating. The produced B_2O_3 is expected to evaporate at temperatures exceeding 1400 °C.

The creation of the disilicide phase, caused by the weakening of Zr_2Si bonds, results in the production of the less thermodynamically stable Zr_2Si phase. Kumar et al. revealed that at temperatures above 1400 °C, SiO_2 interacts with ZrO_2 to generate $ZrSiO_4$, and ZrO_2 also reacts with TiO_2 to form titanium-rich zirconium titanate $Zr_5Ti_7O_2$ phase (Ref 12). But, in this study, ZrO_2 and SiO_2 didn't get react with other oxide elements. As a consequence, oxygen ions react with ZrB_2-SiC components upon exposure to produce its corresponding oxides. The diffraction patterns of V_2O_5 and Na_2SO_5 are seen in salt exposed surface of ceramic due to the remains of salts on the coating surface. The XRD pattern shows that when the ceramic is exposed to molten salt, the diffracted peaks related to

$ZrSiO_4$ binary compounds become more evident, this might be because the molten salt creates an oxidizing atmosphere.

Previous research on electrochemical behavior of ZrB_2 has revealed that metal behaves passively in acidic chloride or neutral solutions. During chemical deterioration, presence of the SiC phase generates a thin protective layer of Si-O, demonstrating passive metal behavior (Ref 13, 14).

4. Conclusions

In this study, a coating was produced with the aim of providing resistance against corrosive environments at 1600 °C for 30 h and having a relatively low coefficient of friction (COF) on its surface. The coating was made using plasma spray technology and consisted of a mixture of ZrB_2 and SiC with a thickness of 300-400 μm, on Inconel-718 substrate. The COF values for the coated substrates ranged from 0.57 to 0.3. The wear rate of the ZrB_2-SiC coating was found to be $2.02 \pm 0.03 \times 10^{-3}$ (mm^3/Nm), which was lower than that of the bare Inconel-718 substrate ($2.8 \pm 0.04 \times 10^{-3}$ (mm^3/Nm)). The coated surfaces were then subjected to hot corrosion tests using molten sulfate/vanadate salt mixtures, and it was found that the ZrO_2 component of the coating formed a protective layer that prevented further penetration of the salts. The study suggests that nanostructured ZrB_2-SiC coatings can be a potential candidate for further development as hypersonic features with improved resistance to hot corrosion via sulfate/vanadate melts.

References

1. V.T. Le, N. San Ha and N.S. Goo, Advanced sandwich structures for thermal protection systems in hypersonic vehicles: a review, *Compos. B Eng.*, 2021, **1**(226), p 109301
2. S. Mungiguerraa, G.D. Di Martinoa, A. Cecerea, R. Savinoa, L. Zoli, L. Silvestroni and D. Sciti, Ultra-high-temperature testing of sintered ZrB_2-based ceramic composites in atmospheric re-entry environment, *Int. J. Heat Mass Transf.*, 2020, **156**, p 119910
3. M. Chen, H. Li, X. Yao, G. Kou, Y. Jia and C. Zhang, High temperature oxidation resistance of La_2O_3-modified ZrB_2-SiC coating for SiC-coated carbon/carbon composites, *J. Alloys Compd.*, 2018, **15**(765), p 37-45
4. P. Sarin, P.E. Driemeyer, R.P. Haggerty, D.K. Kim, J.L. Bell, Z.D. Apostolov and W.M. Kriven, In situ studies of oxidation of ZrB_2 and ZrB_2-SiC composites at high temperatures, *J. Eur. Ceram. Soc.*, 2010, **30**(11), p 2375–2386
5. R.V. Krishnarao, M.Z. Alam and D.K. Das, In-situ formation of SiC, ZrB_2-SiC and ZrB_2-SiC-B4C-YAG coatings for high temperature oxidation protection of C/C composites, *Corros. Sci.*, 2018, **15**(141), p 72–80
6. E. Clougherty, R.J. Hill, W.H. Rhodes and E.T. Peters, Research and development of refractory oxidation-resistant diborides, Part II, vol. II: Processing and Characterization. Tech. Rept. No. AFML-TR-68-190. 1970 Jan
7. S. Chakraborty, D. Debnath, A.R. Mallick, R.K. Gupta, A. Ranjan, P.K. Das and D. Ghosh, Microscopic, mechanical and thermal properties of spark plasma sintered ZrB_2 based composite containing polycarbosilane derived SiC, *Int. J. Refract. Metal Hard Mater.*, 2015, **1**(52), p 176–182
8. Y. Gupta and B.V. Kumar, ZrB_2-SiC composites for sliding wear contacts: influence of SiC content and counterbody, *Ceram. Int.*, 2022, **48**(10), p 14560-14567
9. B. Mukherjee, A. Islam, K.K. Pandey, O.A. Rahman, R. Kumar and A.K. Keshri, Impermeable CeO_2 overlay for the protection of plasma sprayed YSZ thermal barrier coating from molten sulfate-vanadate salts, *Surf. Coat. Technol.*, 2019, **25**(358), p 235-246
10. Y. Yang, Y.H. Qian, J.J. Xu and M.S. Li, Effects of $TaSi_2$ addition on room temperature mechanical properties of ZrB_2-20SiC composites, *Ceram. Int.*, 2018, **44**(14), p 16150-16156
11. Scholz F, Kahlert H. Chemical Equilibria in Analytical Chemistry: The Theory of Acid–Base, Complex, Precipitation and Redox Equilibria. Springer; 2019 Aug 1
12. P.R. Kumar, M.A. Hasan, A. Dey and B. Basu, Development of ZrB_2-based single layer absorber coating and molten salt corrosion of bulk ZrB_2-SiC ceramic for concentrated solar power application, *J. Phys. Chem. C*, 2021, **125**(24), p 13581-13589
13. C. Monticelli, A. Bellosi and M. Dal Colle, Electrochemical behavior of ZrB_2 in aqueous solutions, *J. Electrochem. Soc.*, 2004, **151**(6), p B331
14. C. Monticelli, F. Zucchi, A. Pagnoni and M. Dal Colle, Corrosion of a zirconium diboride/silicon carbide composite in aqueous solutions, *Electrochim. Acta*, 2005, **50**(16-17), p 3461-3469

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

Received: 4 April 2022 | Revised: 19 September 2022 | Accepted: 28 November 2022

DOI: 10.4218/etrij.2022-0113

ORIGINAL ARTICLE

ETRI Journal WILEY

Diversity characteristics of four-element ring slot-based MIMO antenna for sub-6-GHz applications

Vipul Kaushal | **Amit Birwal** | **Kamlesh Patel**

Department of Electronic Science, University of Delhi South Campus, New Delhi, India

Correspondence
Kamlesh Patel, Department of Electronic Science, University of Delhi South Campus, New Delhi. India.
Email: kpatel@south.du.ac.in

Funding information
This work was supported by the Faculty Research Programme (FRP) of the Institute of Eminence (IoE) scheme of the University of Delhi (Letter ref. no. IoE/FRP/PCMS/2020/27 dated 31.12.2020).

Abstract
This paper proposes four-ring slot resonator-based MIMO antennas of $75 \times 150\ \text{mm}^2$ without and with CSRR structures in the sub-6-GHz range. These orthogonal-fed antennas have shown diverse characteristics with dual polarization. L-shaped parasitic structures have increased the isolation (i.e., >40 dB) in the single-element antenna over the band of 3.4 GHz–3.8 GHz. A set of three CSRR structures in the MIMO antenna reduced the coupling between antenna ports placed in an inline arrangement and enhanced the isolation from 12 dB to 20 dB and the diversity characteristics. The S-parameters of both MIMO antennas are measured and used to evaluate MIMO parameters like ECC, TARC, MEG, and channel capacity loss. The simulation results show the variations in the gain and directivity on exciting linear and dual polarizations. The diversity performance of the reported MIMO antennas is suitable for 5G applications.

KEYWORDS
channel capacity loss, ECC, MEG, MIMO antenna, mutual coupling, S-parameters, sub-6 GHz, TARC

1 | INTRODUCTION

Shannon's formula computes an instantaneous capacity for a single-input–single-output (SISO) channel in the transmitter and receiver units that uses the number of antennas at each end of the link [1]. The ergodic SISO capacity (in dB-b/s/Hz) as a function of distance is less than the multiple-input multiple-output (MIMO) capacities for line-of-sight geometry. MIMO wireless technology can increase a channel's capacity by maintaining equal antenna spacing. It is possible to increase the channel's throughput linearly with every pair of transmitting and receiving antennas added to the system. This feature makes MIMO technology one of the essential wireless technologies to be employed recently. Additionally, MIMO communication channels offer an exciting solution to multipath interference by utilizing multiple signal paths [2]. As one of the applications, the MIMO antenna is widely used for 5th Generation (5G) communication systems. The advantages of MIMO in 5G technology are higher data rates, unified connectivity, low latency, and more connected nodes [3]. Several MIMO antenna designs have been reported for 5G applications recently. A dual-band MIMO antenna is reported to cover bands of 3.3 GHz–3.6 GHz and 4.8 GHz–5.0 GHz with an isolation of approximately 12 dB [4]. Using a self-isolated antenna element, a broadband MIMO antenna system reported a 14-dB isolation in the 5G NR (3.3 GHz–4.2 GHz) frequency range [5]. However, it is a relatively complex structure. Furthermore, such separation is insufficient to prevent any intervention.

The main characteristics of the MIMO antenna for 5G are low profile, high isolation between antenna ports, envelope correlation coefficient (ECC) of <0.5 [6], minimum total average reflection coefficient (TARC) (<−20 dB), and mean effective gain (MEG) between −3 dB and −12 dB [6]. The future aspect of MIMO design is to make it massive by adding numbers to transmitting and receiving antennas, hence increasing the system's overall channel capacity. Polarization diversity can meet the need for more antennas, resulting to space diversity, by adopting dual polarization. Hence, the overall size of a dual-polarized MIMO antenna can be cut in half when compared with identical single-polarized antennas [7, 8]. Reference [9] compares the features of various isolation improvement techniques used to reduce mutual coupling between feed ports in a MIMO antenna. One of the isolation techniques reported in previous studies [9] is mushroom-type electromagnetic band gap (EBG) structures that increase the isolation by approximately 24.6 dB. A row of mushroom-type EBG structures can be used for further improvement in the isolation. Although it provides sound isolation, its design is complex. Some structures like slitted patterns in the ground plane, rectangular defected ground structure, and slots provide 20-dB, 32-dB, and 45-dB improvement in the isolation, respectively. These designs are simpler; however, the techniques utilizing these structures are reported to degrade the radiation pattern of the antenna. A U-shaped microstrip structure increases isolation by 31 dB, and it has the advantage of a more straightforward design without any degradation of the radiation pattern. In a two-element MIMO antenna array consisting of two rectangle patch antennas with 50-Ohm coaxial feeding, five 3-D meta-material cells composed of an upper M-shaped patch and two lower U-shaped patches were implemented and >18-dB mutual coupling reduction has been obtained without any deviation in the operating bandwidth and radiation characteristics [10]. A recent study introduces a metasurface-based decoupling method between two coupled MIMO antennas. A metasurface superstrate comprises pairs of nonuniform cut wires with different lengths [11]. The measured results confirmed that the isolation between two dual-band antennas is improved to >25 dB at both 2.5 GHz−2.7 GHz and 3.4 GHz–3.6-GHz bands. In another approach [12], a ceramic superstrate-based decoupling method is proposed to reduce the mutual coupling between two closely packed dipole antennas and cross-polarization suppression. This ceramic superstrate is taken on a dielectric slab with ε_r of 20.5 and thickness of 2 mm and kept suspended over the antennas in the H-plane. This increased the isolation from 10 dB to >25 dB in the operation band of 3.3 GHz–3.7 GHz. The minimization of ECC values between antenna elements can improve the radiation pattern, decrease the interference between channels, and enhance the MIMO antenna performance characteristics. Hence, by adding an L-shaped structure on the top layer, the isolation improvement is achieved in this work up to 44 dB, which is simple to design without affecting the antenna's radiation pattern. Additionally, aperture-coupled feeding is preferred as it is offers advantages like the feed-line radiation isolated from slot radiation and higher bandwidth [13]. Hence, placing feedlines and radiators on different planes provides these advantages. Moreover, the main advantages of using a ring slot as a radiator are a low profile, fabrication simplicity, and bidirectional radiation, which lead to the preferred employment of a ring slot radiator antenna in a 5G smartphone PCB [14]. A new reduced-coupling slot MIMO antenna design is proposed for 5G MIMO mobile terminals composed of dual-polarized square-ring slot antennas placed at four corners [15]. This antenna reported >75% radiation efficiency (RE), 60% total efficiency (TE), ECC ≤ 0.005, and TARC better than −25 dB between the ports in the frequency range of 3.4 GHz–3.8 GHz. In previous studies [16], an eight-antenna MIMO array is proposed to operate at the LTE band 42 and LTE band 46 for the 5G mobile terminals, which is composed of eight slot antenna elements based on stepped impedance resonators and activated by shorted microstrip feed lines printed on the top side of the substrate. This MIMO antenna of size of $70 \times 140\ \text{mm}^2$ reported an ECC of ≤0.08 and isolation of ≥11.2 dB at 3.4 GHz–3.6 GHz and 5.15 GHz–5.925 GHz bands. Enhancement of isolation between ports, ease of design, and robustness make slot radiator antennas suitable for antenna design. Such a feature provides one additional advantage of higher channel capacity, so proclaiming the channel capacity loss (CCL) makes the MIMO antenna design more selective for a high data rate transmission. Simultaneously, the evaluation of MIMO parameters and CCL from the measured results confirms the performance of the proposed MIMO antenna in the real-time environment.

In this paper, a four-element dual-polarized MIMO circular ring slot antenna with a pair of L-shaped parasitic structures is proposed that offers a low profile of $75 \times 150\ \text{mm}^2$, high isolation between the ports (>20 dB), an RE of up to 80%, a TE of up to 76%, and various diversity performance like lower ECC (<0.005), less TARC (up to −12 dB) between ports, MEG between −3 dB and −12 dB for all the antenna ports, and CCL < 0.7 b/s/Hz in the sub-6-GHz band, that is, 3.4 GHz–3.8 GHz. The MIMO parameters and CCL are also calculated from the measured S-parameters to validate the performance. Furthermore, the MIMO design is improved by affixing a set of three complementary splitring resonators (CSRRs) to the ground layer in order to reduce mutual coupling between ports in an inline arrangement without compromising its radiation characteristics. However, when the number of ports is increased from four to eight with CSRR structures, the ECC and TARC are found to be inefficient, whereas the minimum CCL is reduced to 0.4 b/s/Hz.

2 | SINGLE-ELEMENT ANTENNA (SEA) CONFIGURATION

Figure 1 illustrates an SEA consisting of a circularring slot radiator on the bottom layer and orthogonally positioned two microstrip feed lines on the top layer on a FR4 substrate ($\varepsilon_r = 4.5$, height = 1.5 mm). Instead of a ring, a pair of L-shaped parasitic devices are introduced to improve isolation [15]. The antenna structure is dual-polarized and has a 35 × 35 mm^2 low profile. Dual polarization has the advantage of boosting channel capacity, which is necessary for 5G applications, using fewer antennas and enhancing envelope correlation between nearby ports. Table 1 shows the dimensions of SEA.

The S-parameters of this SEA with optimized physical dimensions (given in Table 1) are shown in Figure 2. The return loss (S_{nn}) of SEA is found to be −31.73 dB at 3.6 GHz and its impedance bandwidth is 425 MHz (for −10 dB).

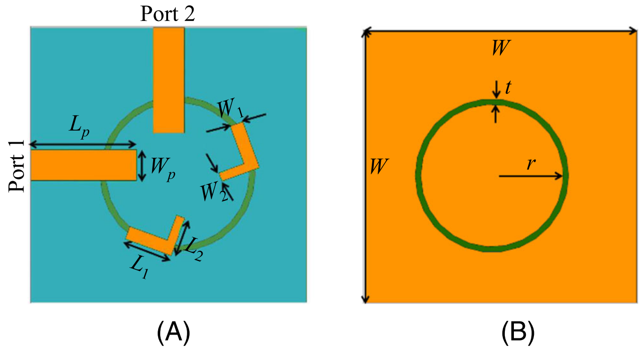


FIGURE 1 Single-element antenna (SEA) geometry: (A) Top view. (B) Bottom view.

TABLE 1 Design parameter values for SEA.

Parameter	Values (mm)	Parameter	Values (mm)
W	35.0	r	8.10
L_p	11.9	t	0.65
W_p	3.50	W_1	1.50
L_1	5.50	W_2	1.00
L_2	4.50	W_s	1.60

The isolation (S_{mn}) between two ports of SEA has reached a minimum of 44.53 dB in the sub-6 GHz range (3.4 GHz–3.8 GHz) by introducing an L-shaped parasitic structure on the top layer. On exciting Ports 1 and 2 individually, the 3D radiation patterns of SEA at 3.6 GHz are illustrated in Figure 3. The directivity is found to be 3.77 dBi at 3.6 GHz, whereas the beamwidths are obtained as ∼121° and ∼88° on exciting Port 1 and Port 2, respectively.

3 | MIMO ANTENNA CONFIGURATION

3.1 | Proposed MIMO based on SEA

As illustrated in Figure 4, the improved antenna element was used to create a MIMO antenna in a 2 × 2 array on

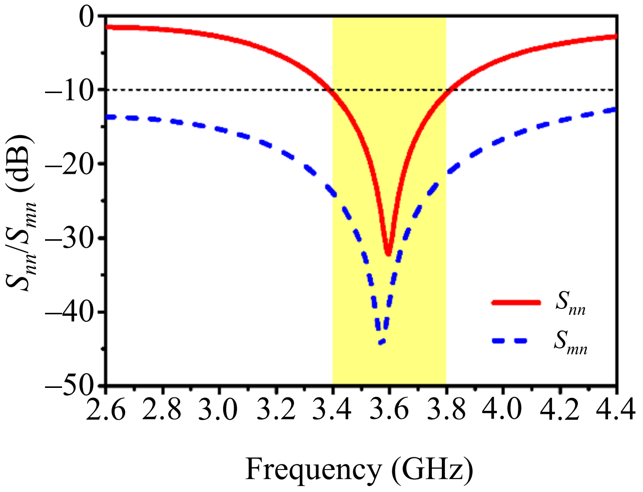


FIGURE 2 S-parameters response of SEA, where m and $n = 1$ or 2; $m \neq n$.

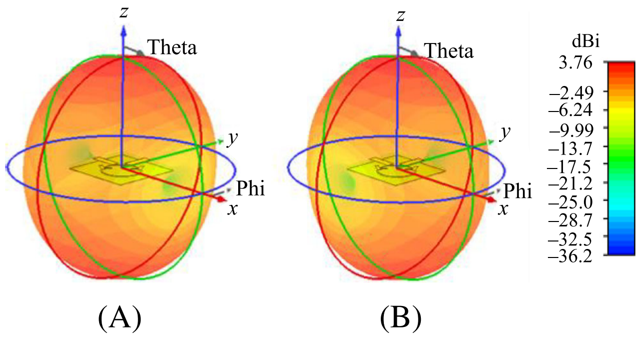


FIGURE 3 Three-dimensional radiation pattern of single-element antenna (SEA) when fed at (A) Port 1 and (B) Port 2.

the same FR4 substrate, with an overall geometry 75 × 150 mm^2. Furthermore, the size of element antennas placed at all four corners of the MIMO structure has been reduced to 29 × 29 mm^2,that is, by 31.3% as the parameters (length of feed line, thickness, and inner radius of circular slot on ground plane) for SEA are reduced a little as in Table 2. These new dimensions improved matching, and a MIMO antenna was fabricated accordingly. In the designed MIMO antenna, the orthogonally placed ports in each SEA support different transmission channels with independent or very less fading of the signals, that is, signals with horizontal and vertical polarization transmitting from or receiving at these ports experience low correlation, which makes the scope of polarization diversity.

Thus, polarization diversity offers the advantage of acquiring less space on the smartphone PCB for efficient diversity characteristics. However, it is limited to establishing only two separate transmission channels. To overcome this limitation, one feasible approach is to install these SEAs in the four corners of the smartphone PCB to establish several channels. Consequently, the four sets of either horizontal or vertical ports can receive several versions of the same signal by utilizing separate transmission pathways with less fading correlation between them; this configuration accredits another sort of diversity known as space diversity. Figure 5 depicts a visual representation of polarization and space variety.

Figure 6 (A,B) shows the completed MIMO antenna (B). The S-parameter measurement (S_{nn}) is performed by connecting one MIMO antenna port to the vector network analyzer (VNA) R&S ZVH8 and leaving the

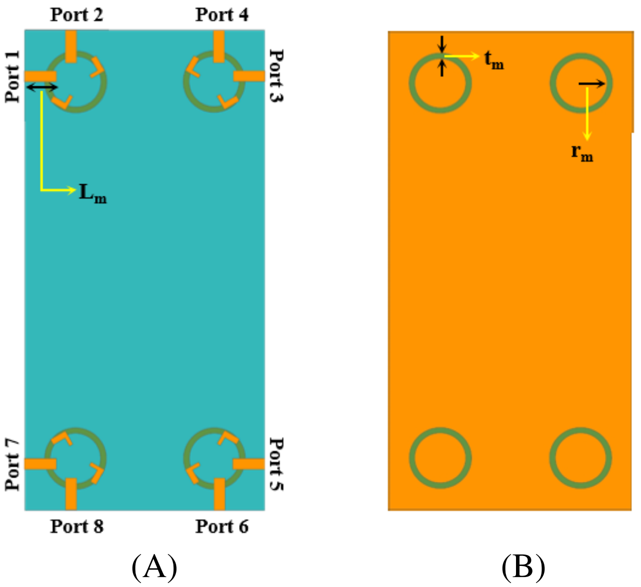


FIGURE 4 (A) Top layer and (B) bottom layer of proposed 5G MIMO antenna.

TABLE 2 Modified design parameters in proposed MIMO antenna.

Parameter	Values (mm)
L_m	10.00
r_m	8.00
t_m	1.85

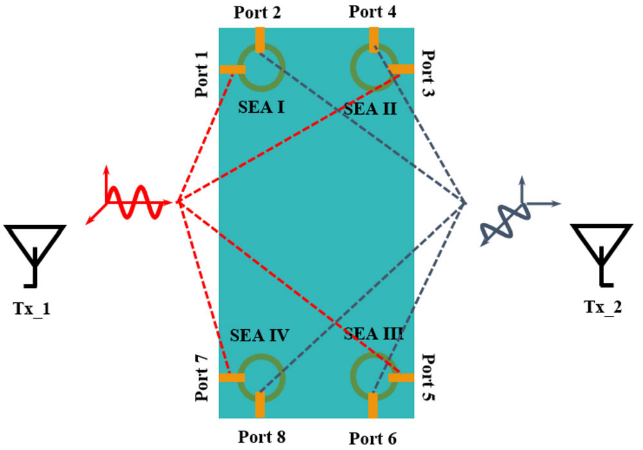


FIGURE 5 Polarization and space diversity scenario in the proposed MIMO structure.

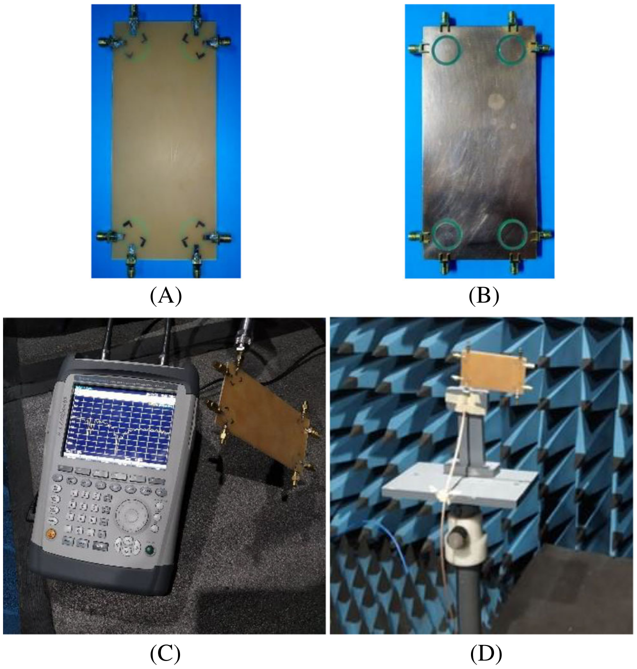


FIGURE 6 Fabricated MIMO. (A) Top layer. (B) Bottom layer. (C) S-parameter measurement setup. (D) Radiation pattern measurement setup in the anechoic chamber.

remaining ports terminated with 50-ohm loads, as shown in Figure 6C. The isolation (S_{mn}) between two antenna ports is measured using the same VNA, whereas all other ports are terminated. The MIMO antenna's radiation pattern is measured in an anechoic chamber using signal generator RIGOL DSG3060 and spectrum analyzer R&S FSL6 as shown in Figure 6D. Figures 7 and 8 show the simulated and measured results of return loss and isolation characteristics of the proposed 5G MIMO antenna. This antenna has a return loss of better than −30 dB at each port, whereas the measured S_{nn} values are from −18 dB to −38 dB in the desired band for all ports as shown in Figure 7B. The isolation (S_{mn}) from Port 1 to other ports is found below −20 dB except with Port 3 due to their inline alignment in the 3.6-GHz band. Similar isolation is observed between Ports 5 and 7, that is, −12.9 dB at 3.6 GHz. The measured Smn values agree with the simulated one by comparing Figure 8 (A,B), which ensures the use of this designed MIMO antenna in the sub-6-GHz range, that is, the n78 band. Directivity values exiting various ports range from 4.69 dBi to 6.11 dBi due to dual polarization horizontal polarization (on stimulating Ports 1, 3, 5, and 7) and vertical polarization (on exciting Ports 2, 4, 6, and 8). Table 3 shows the details of the simulated gain, directivity, and beamwidth values for these excitations.

Figure 9 shows the simulated and measured radiation patterns of the designed MIMO antenna at 3.6 GHz when exciting horizontal and vertical ports. It could be seen that while exciting the horizontal port, the MIMO antenna gives a dumbbell-shaped radiation pattern having a gain of 3.8 dB in the XZ plane while covering both sides of the smartphone PCB and the YZ plane. On exiting the vertical port, the MIMO antenna in the YZ plane radiates a better gain of 5.2 dB on the top side compared

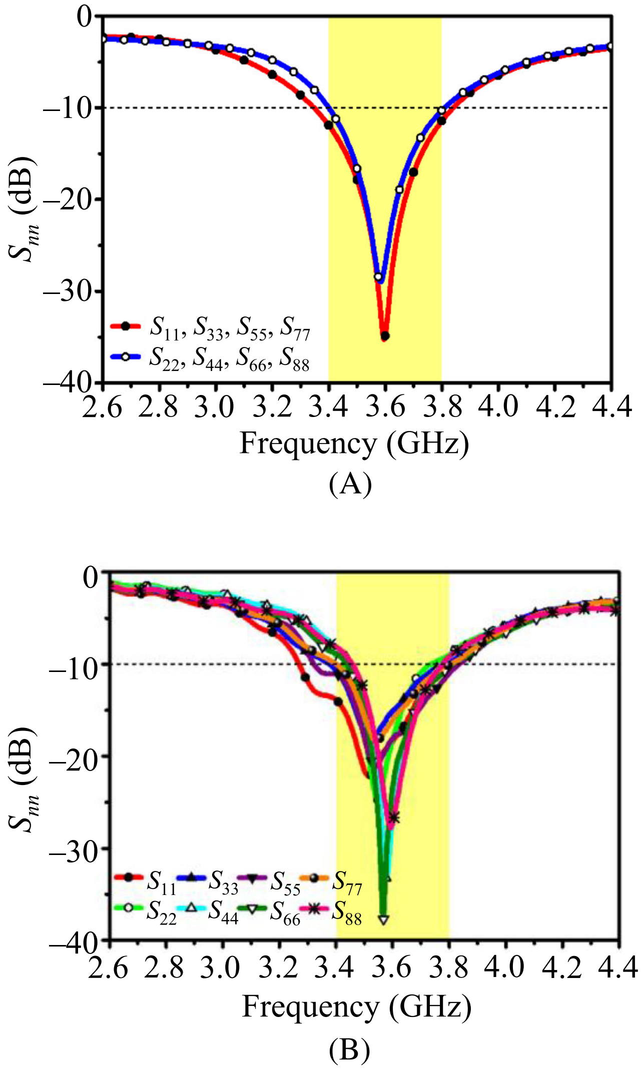


FIGURE 7 S-parameter response of the proposed MIMO antenna, where $n = 1$–8. (A) Simulated S_{nn}. (B) Measured S_{nn}.

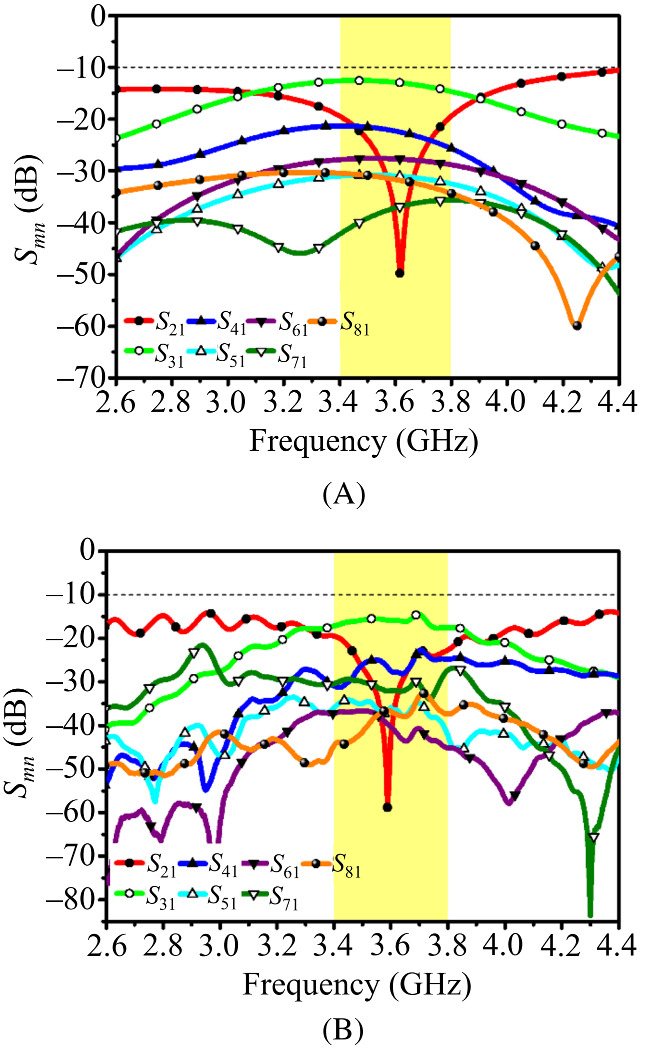


FIGURE 8 S-parameter response of proposed MIMO antenna; where $m = 2$–8 and $n = 1$. (A) Simulated S_{mn}. (B) Measured S_{mn}.

TABLE 3 Simulated gain, directivity, and beamwidth of MIMO antenna at 3.6 GHz on various polarizations.

Excited ports	Gain (dB)	Directivity (dBi)	3 dB beamwidth
1, 3, 5, 7 (horizontal)	3.82	4.69	58.4°
1, 3, 2, 4 (dual)	3.50	4.18	40.2°
2, 4, 6, 8 (vertical)	5.21	6.11	39.3°

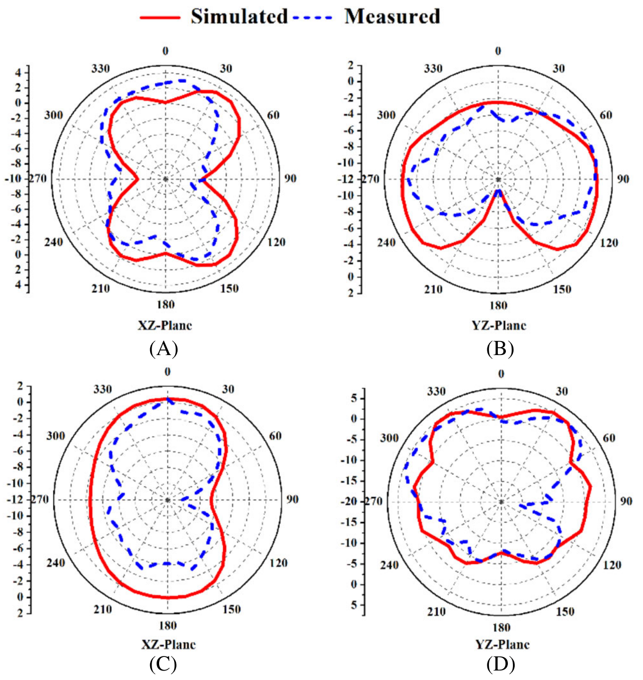


FIGURE 9 The radiation pattern of the designed MIMO antenna at 3.6 GHz (A) horizontal port (1/3/5/7) in the XZ plane, (B) horizontal port (1/3/5/7) in the YZ plane, (C) vertical port (2/4/6/8) in the XZ plane, and (D) vertical port (2/4/6/8) in the YZ plane.

to a gain of −3 dB on the other side. However, the XZ plane pattern covers the left and right sides of the antenna PCB with the same gain values. A possible higher mutual coupling between horizontal ports results in lesser gain than vertical ports.

For smartphone applications, the proximity of a human head or hand can reduce the proposed antenna's overall radiation characteristics and gain, and TE will vary as reported for similar MIMO antennas [15, 16]. In the presence of user head/hand, the gain of such a MIMO antenna varies from 1.3 dB to 4 dB. However, with double-hand use, the gain is >1.9 dB [15], as antenna pairs placed far from the user body's vicinity will exhibit good radiation characteristics due to back radiation. Additionally, depending on the position, the TE of a single antenna element or antenna pair may be reduced by 13%–18% [16]. In such a scenario, the designed 8 × 8 MIMO antenna array behaves as a 6 × 6 or 4 × 4 antenna array with satisfactory performance in human proximity at the 5G band.

3.2 | Isolation improvement using CSRR

Figure 7 shows the poor isolation between port combinations 1/3 and 5/7, which is approximately 12 dB–15 dB in the range of 3.4-GHz–3.8-GHz band for this MIMO antenna design (Figure 4). These isolation values lower the antenna efficiency and the ECC [17]. As these parameters have a massive impact on a MIMO system's channel capacity and diversity gain performance [18], the isolation between port combinations 1/3 and 5/7 requires to be higher, at least >15 dB, without impacting the isolation between other ports. This performance of the MIMO antenna system is more important, regardless of the ports underutilized.

To improve isolation, a CSRR is proposed, as the CSRR-loaded transmission lines offer rejection bands depending on the design of CSRR and are modeled by a series inductor connecting the line capacitance to the corresponding LC resonator of CSRRs [19]. As the CSRRs are etched in the ground plane and are mainly excited by the electric field induced by the feed line on the top layer, the radiation pattern has minimal effect. The dimensions of CSRR are obtained by keeping 3.6 GHz as a rejection frequency and are given in Table 4.

In Figure 10, using one and a set of two CSRR structures between Port 1 and Port 3, the isolation between these ports is found to be approximately −13 and −12 dB, respectively. The isolation between the horizontal ports (i.e., 1 and 3) is improved to less than −20 dB after the addition of the third CSRR, which is sound isolation between two neighboring ports. To improve isolation between port combinations 1/3 and 5/7, a set of three CSRR structures is created on the bottom layer of the MIMO antenna, as shown in Figure 11, with dimensions up to the diameter of circular slots.

The S-parameters of the MIMO antenna (Figure 4) and CSRR-loaded MIMO antenna (Figure 11) are compared in Figures 12 and 13. In the simulation results of Figure 12, the S_{nn} values are improved in all ports with a slight shift in the center frequency for CSRR-loaded

TABLE 4 Dimensions of complementary split-ring resonator.

Parameter	Values (mm)
$R_{\max}$	4.3
g	0.5
W_r	1.0

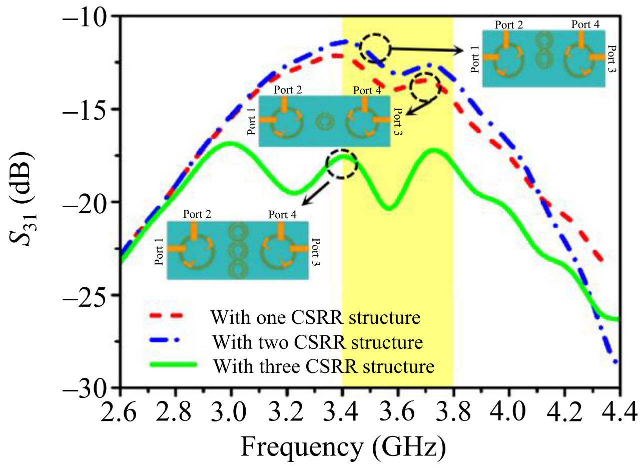


FIGURE 10 Evolution of CSRR structure set between Ports 1 and 3.

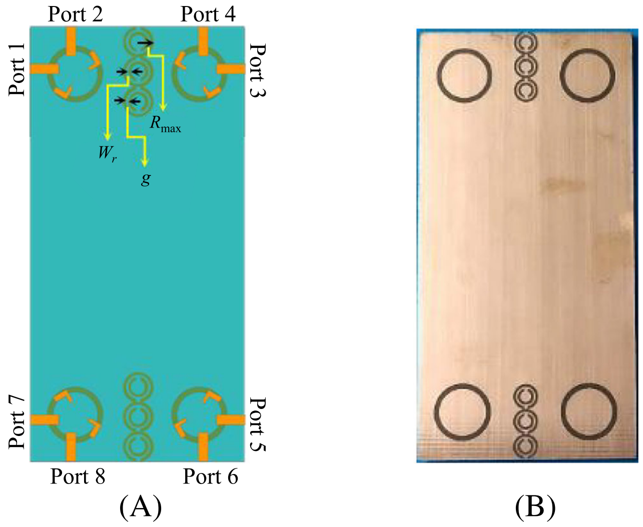


FIGURE 11 (A) Designed and (B) fabricated CSRR-loaded MIMO antenna (bottom layer).

MIMO antenna and are found to be better than 30 dB at 3.575 GHz, as indicated by the responses of S_{11}, S_{22}, S_{33}, and S_{44} for both antennas. Additionally, the measured S_{nn} data of both antennas comply with their respective simulation data.

Figure 13 shows the improvement in simulated S_{31}, that is, −20.69 dB at 3.6 GHz, with a slight degradation in the simulated S_{21} parameter for CSRR-loaded MIMO antenna. The isolation for parameters S_{31} and S_{75} improves by 60.35%, whereas the measurement results show approximately 1 dB better S_{31} and S_{75} than simulated data. Similar performances are observed for other transmission parameters.

For the proposed 5G MIMO antenna and CSRR-loaded MIMO antenna, the simulated radiation and TE are shown in Figure 14. RE is >80% in the sub-6 range (3.4 GHz–3.8 GHz) and approximately 81% at 3.6 GHz.

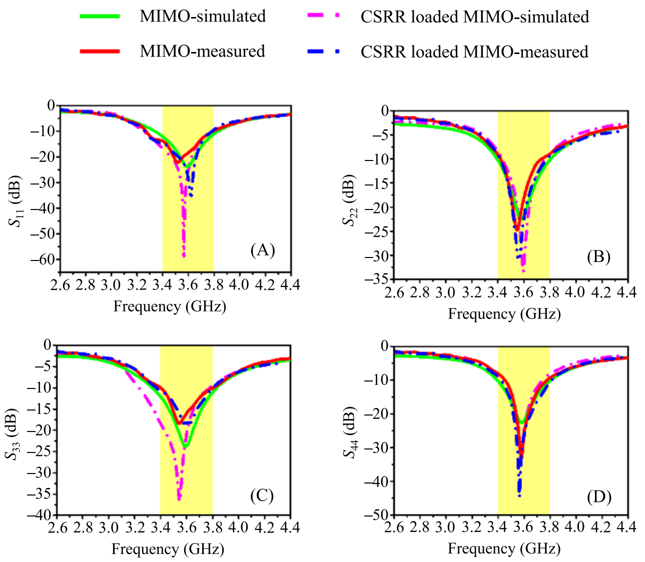


FIGURE 12 Simulated and measured S_{nn} comparison between MIMO with and CSRR-loaded MIMO.

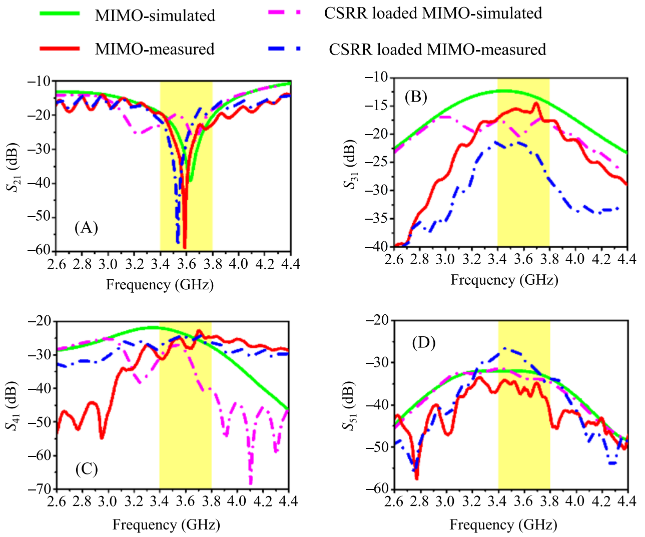


FIGURE 13 Simulated and measured S_{mn} comparison between MIMO with and CSRR-loaded MIMO.

Moreover, >71% of TE is achieved in the same frequency range and approximately 76% at 3.6 GHz. These results confirm that these antennas are highly efficient in the operating frequency band, so they are employable in 5G applications.

Figure 15 shows the surface current distribution for the proposed MIMO antenna and CSRR-loaded MIMO antenna when only Port 1 is excited. It is observed that at the frequency of 3.6 GHz, the maximum current is focused on the ring slot in the bottom layer of the proposed MIMO antenna (Figure 4). Furthermore, a critical part of the current is being coupled to the other slot radiator placed at the ground plane; hence, it shows poor isolation. The introduction of three CSRR structures breaks this coupling path in the MIMO antenna (Figure 11). The isolation is successfully increased to approximately 21 dB between Ports 1/3 and 5/7 due to the current flow into the rings of CSRR.

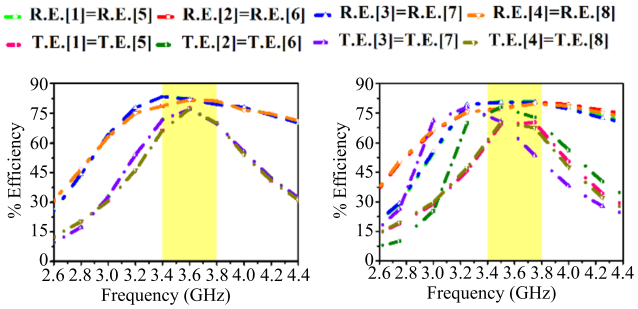


FIGURE 14 Radiation efficiency (RE) and total efficiency (TE) of (A) proposed MIMO antenna and (B) CSRR-loaded MIMO, where 1–8 are antenna port no.

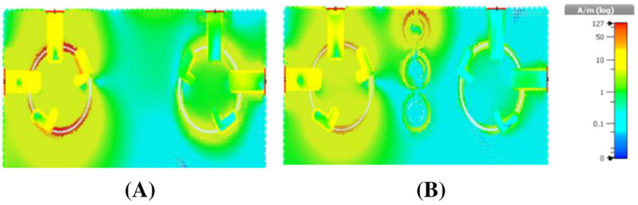


FIGURE 15 The surface current density of (A) proposed MIMO structure and (B) CSRR-loaded MIMO structure when only Port 1 is excited.

To further validate the performance of these MIMO antennas (Figures 4 and 11) for isolation with the physical dimensions in the same band, a comparison is made with the recent MIMO antenna design and is shown in Table 5.

4 | DIVERSITY CHARACTERISTICS OF THE PROPOSED MIMO ANTENNA

The diversity characteristics of a MIMO antenna system are evaluated relating to key performance parameters like ECC, TARC, MEG, and CCL. For the proposed MIMO antenna, the ECC between two single antenna elements, that is, Ant i and Ant j, can be calculated from their far-field radiation patterns using expression (1) [16]:

$$\mathrm{ECC}\,(\rho ij) = \frac{\left|\int\int_{4\pi}\overline{F}_i(\theta,\phi).\overline{F}_j^*(\theta,\phi)\mathrm{d}\Omega\right|^2}{\int\int_{4\pi}\left|\overline{F}_i(\theta,\phi)\right|^2\mathrm{d}\Omega\int\int_{4\pi}\left|\overline{F}_j(\theta,\phi)\right|^2\mathrm{d}\Omega}, \qquad (1)$$

where the complex vectors, $\overline{F}_i(\theta,\phi)$ and $\overline{F}_j(\theta,\phi)$, denote the far-field radiation patterns of Ant i, and Ant j, respectively, and * denotes the Hermitian transpose of a matrix. Furthermore, using the simulated/measured S-

TABLE 5 Isolation comparison with previously published MIMO designs.

Structure used	Size (mm^2)	Freq. (GHz)	No. of ports	Isolation (dB)	Directivity (dBi)	Efficiency
Four bent lines and floor protruding branches [4]	74 × 130	3.3–3.6 & 4.8–5.0	4	>15	NA	<85% & <60%
Self-isolated antenna element with a T-shaped [5]	75 × 150	3.3–4.2	8	>14	NA	<70%
Rectangular ring slot radiator [15]	75 × 150	3.3–3.8	8	23	6.5	75%
Diamond-ring slot antenna [20]	75 × 150	3.3–3.9	8	17	5.6	80%
Dual-polarized antenna with an AMC reflector [21]	79.6 × 79.6	3.1–3.8 & 4.4–5.0	4	20	9.1	90%
Circular ring slot radiator antenna*	75 × 150	3.4–3.8	8	12	6.12	82.6%
Circular ring slot radiator antenna with CSRR*	75 × 150	3.4–3.8	8	20	5.71	75%

*This work.

parameters of the MIMO antenna, the ECC is also obtained by following mathematical expression [2]:

$$\mathrm{ECC}\,(\rho ij) = \frac{|S_{ii}{}^{*}\,S_{ij} + S_{ji}{}^{*}S_{jj}|^{2}}{\left(1-|S_{ii}|^{2}-|S_{ij}|^{2}\right)\left(1-|S_{ji}|^{2}-|S_{jj}|^{2}\right)}, \qquad (2)$$

where $i = 1$ and j (= 1–8) represents the antenna port numbers (Figure 4).

The ECC values of pairs of antenna ports for both MIMO antennas (without and with CSRR) are shown in Figure 16, where Figure 16 (A,B) represents the ECC of both antennas obtained from the simulated far-field results, whereas Figure 16 (C,D) shows the ECC of both MIMO antennas obtained from the measured S-parameters using the expression (2).

The calculated ECC is found to be a very low value of <0.04 in the operating frequency band (from 3.6 GHz to 3.8 GHz) for both MIMO antennas (Figures 4 and 11), which is an acceptable standard for a desirable MIMO system [6]. The CSRR-loaded MIMO antenna has shown ECC < 0.04 for the range 3.3 GHz–4.5 GHz, so it is useful for broadband applications.

TARC is defined as the ratio of root square total reflected power (r_k) to root square total incident power (i_k) by the antenna. For N-port MIMO antenna [2],

$$\mathrm{TARC} = \sqrt{\frac{\sum_{k=1}^{N}|r_k|^2}{\sum_{k=1}^{N}|i_k|^2}}, \qquad (3)$$

where the port number $k = 1$ to N (here, the number of antennas, $N = 8$). Moreover, in terms of S-parameters,

$$\mathrm{TARC} = \sqrt{\frac{\left(S_{11}+S_{12}e^{j\theta}\right)^2+\left(S_{21}+S_{22}e^{j\theta}\right)^2}{2}}. \qquad (4)$$

TARC shown in Figure 17 (A) and (C) illustrates the same simulated and measured response as a single-port excitation response but less bandwidth when the coupling is higher between the ports (i.e., Ports 1 and 3).

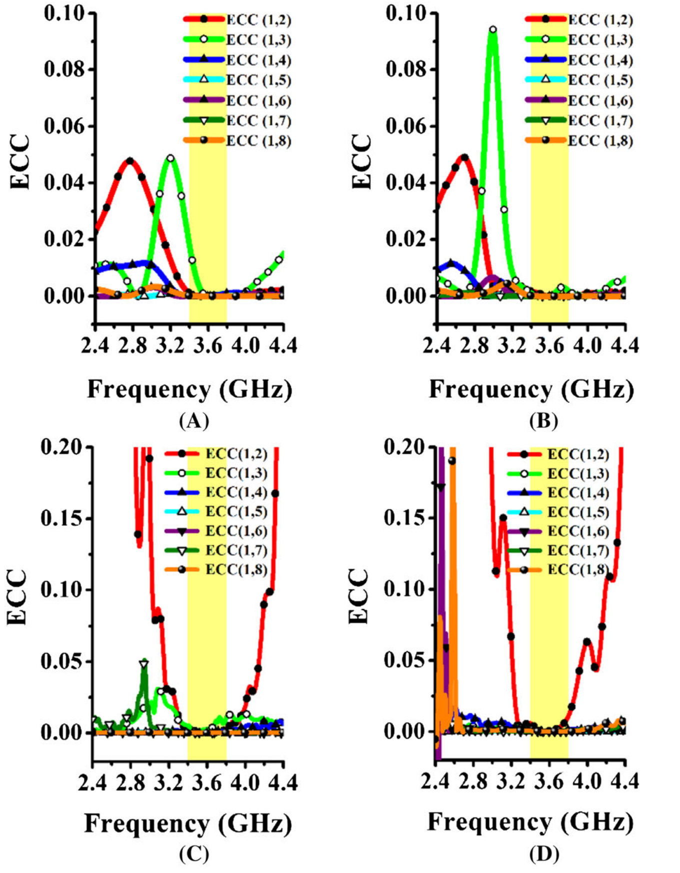


FIGURE 16 Calculated ECC values from simulated data (A) MIMO antenna (Figure 4) and (B) CSRR-loaded MIMO antenna (Figure 11) and measured data (C) MIMO antenna and (D) CSRR-loaded MIMO antenna.

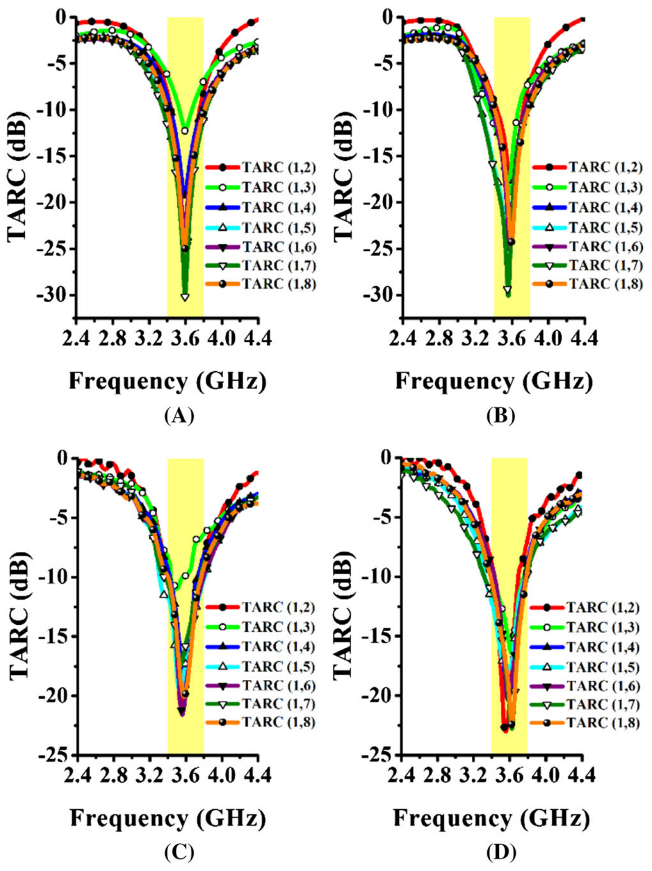


FIGURE 17 Calculated TARC v/s frequency from simulated data (A) MIMO antenna (Figure 4) and (B) CSRR-loaded MIMO antenna (Figure 11) and from measured data (C) MIMO antenna (D) CSRR-loaded MIMO antenna.

However, bandwidth increases in CSRR-loaded MIMO as the coupling is reduced between these ports, as shown in Figures 17 (B,D).

For evaluation of the diversity performance of a MIMO system, the MEG is defined as the ratio of the average received power by a MIMO antenna and the power received by an isotropic antenna [22]. For the N-port antenna system, MEG can be evaluated by Khan et al. [22],

$$\mathrm{MEG}_i = 0.5\left(1 - \sum_{j=1}^{N} |S_{ij}|^2\right) \tag{5}$$

where N is the number of antennas, i, and j (= 1 to 8) represents the antenna port numbers. The required value of MEG to exhibit good diversity characteristics should be −12 dB < MEG < −3 dB. The simulated MEG of the proposed eight-port MIMO antenna is illustrated in Figure 18 for all eight ports. For all ports of the MIMO antenna (Figure 4), the MEG is found between −3 dB and −8 dB in 3.4 GHz and 3.8 GHz with slight variation for each port (as shown in inset), whereas for CSRR-loaded MIMO antenna (Figure 11), it is −3 dB to −12 dB. The same values (approximately −7 dB to −5 dB) are found for Ports 1, 3, 5, and 7, and for Ports 2, 4, 6, and 8, it is in the range of −12 dB to −6 dB.

The final diversity parameter, CCL describes the loss in information transmission rate due to the effect of correlation. CCL for N-port MIMO can be calculated using the following equations [23]:

$$\mathrm{CCL} = -\log_2 \det(\boldsymbol{\Psi}_R) \tag{6}$$

where $\boldsymbol{\Psi}$ shows the antenna correlation matrix.

$$\boldsymbol{\Psi}_R = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{bmatrix}. \tag{7}$$

Here,

$$\sigma_{ii} = 1 - \left| \sum_{j=1}^{N} S_{ij}^{*} S_{ji} \right|, \tag{8}$$

$$\sigma_{ij} = -\left| \sum_{k=1}^{N} S_{ik}^{*} S_{kj} \right|. \tag{9}$$

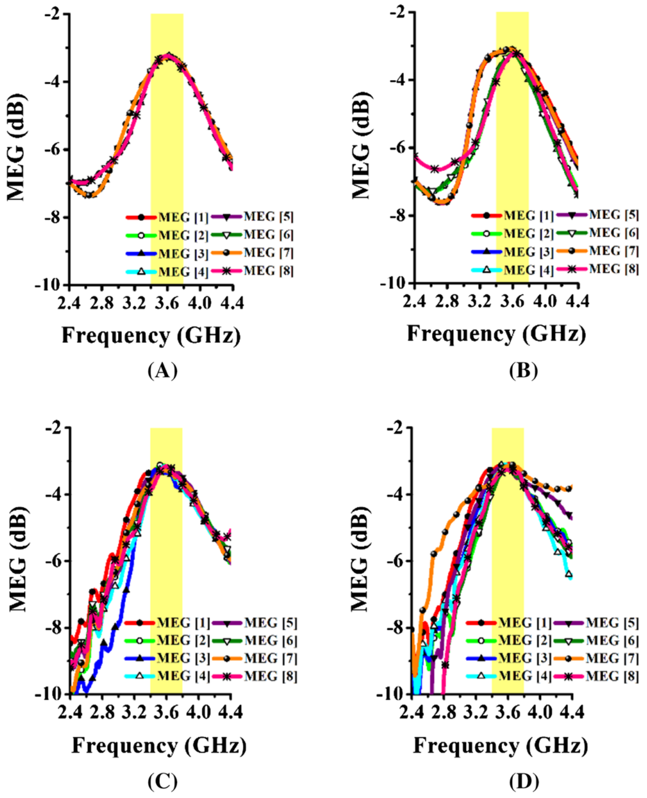


FIGURE 18 Calculated MEG versus frequency from simulated data (A) MIMO antenna (Figure 4) and (B) CSRR-loaded MIMO antenna (Figure 11) and from measured data (C) MIMO antenna (D) CSRR-loaded MIMO antenna.

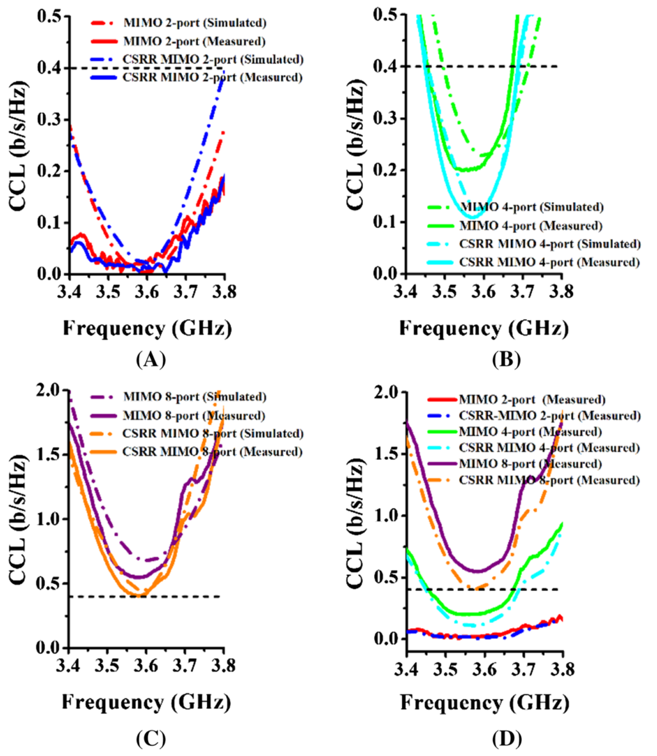


FIGURE 19 Channel capacity loss (CCL) versus frequency for (A) two-port, (B) four-port, and (C) eight-port MIMO antenna. (D) Comparison between calculated CCLs.

The calculated CCL for the proposed MIMO and CSRR-loaded MIMO is shown in Figure 19 by considering various sets of port arrangements. When only two ports of MIMO and CSRR-loaded MIMO are considered as in Figure 19 (A), the CCL values from simulated and measured S-parameters are found to be <0.4 b/s/Hz in the 3.4-GHz–3.8-GHz band for both antennas. By considering four ports for both MIMOs, as in Figure 19 (B), CCL is calculated to be <0.4 b/s/Hz for the range of 3.45-GHz–3.65-GHz band from measured S-parameters. In this band, the calculated values of CCL are 0.22 and 0.12 b/s/Hz at 3.6 GHz, respectively, for the proposed MIMO and CSRR-loaded MIMO antenna.

Furthermore, on increasing the number of ports to 8, CCL values are increased, as shown in Figure 19 (C). From simulation data, CCL is found to be 0.67 b/s/Hz for MIMO, and 0.45 b/s/Hz for CSRR-loaded MIMO, whereas it is 0.56 b/s/Hz for MIMO and 0.42 b/s/Hz at 3.6 GHz for CSRR-loaded MIMO from the measurement data. In Figure 19 (D), CCL from measured data is compared for various sets of ports, indicating that with the increasing number of antennas in the MIMO system, the CCL increases from a more significant correlation between multiple ports.

This result also confirms that lower CCL is obtained by inserting CSRR in the bottom layer of the MIMO, resulting in a more effective MIMO system. Similar values of CCL (approximately 0.3 b/s/Hz) in the bands 2.5 GHz–12 GHz are reported in previous works [23].

To confirm the improved diversity parameters, like ECC, TARC, MEG, and CCL values of our simple, low-profile MIMO antenna designs are compared in Table 6. On comparing with previously reported four-port MIMO antennas, the presented four-port MIMO antennas generated improved ECC (<0.003), TARC (<−17 dB), MEG (−3 dB to −12 dB), and CCL (<0.23 b/s/Hz) values in the sub-6 GHz. When the number of ports increases from 4 to 8 in the presented MIMO antenna designs, the ECC and TARC become little poor to 0.003 and better than −17 dB, respectively. However, the minimum CCL varies nonlinearly with the port numbers of MIMO antenna designs.

TABLE 6 MIMO performance comparison with previously published design.

Reference	No. of ports	Bandwidth (GHz)	ECC (max)	TARC	CCL(b/s/Hz) (min)
[15]	8	3.4–3.8	<0.002	–	–
[16]	8	3.4–3.6	<0.08	–	–
[23]	4	2.5–12.0	0.005	Better than −6 dB	0.3
[24]	4	2.0–10.6	0.005	–	0.3
[25]	4	3.1–10.6	0.1	Better than −4 dB	>1 (3–5 GHz) & ≤0.1 (>5 GHz)
[26]	4	3.1–10.6	0.0025	Better than −8 dB	0.2
MIMO with two ports*	2	3.4–3.8	0.0012	Better than −20 dB	0.01
MIMO with four ports*	4	3.4–3.8	0.018	Better than −12 dB	0.23
MIMO with eight ports*	8	3.4–3.8	0.018	Better than −12 dB	0.67
CSRR-loaded MIMO with two ports*	2	3.4–3.8	0.0018	Better than −17 dB	0.02
CSRR-loaded MIMO with four ports*	4	3.4–3.8	0.003	Better than −17 dB	0.12
CSRR-loaded MIMO with eight ports*	8	3.4–3.8	0.005	Better than −17 dB	0.4

*This work.

5 | CONCLUSION

Two dual-polarized circular ring slot MIMO antenna and CSRR-loaded MIMO antennas are presented for 5G applications in the sub-6-GHz range. The MIMO parameters like ECC, TARC, and MEG are improved by incorporating a set of three CSRRs on the bottom layer in the 3.4 GHz–3.8-GHz band. The CCLs are obtained from the measured S-parameters for both fabricated antennas are <0.12 and 0.4 b/s/Hz for four-port MIMO and eight-port MIMO antennas due to CSRR structures, respectively. The reported diversity characteristics confirm the use of these MIMO antennas in 5G applications. These antennas can be further modified for a multiband MIMO operation to achieve carrier aggregation, multiple transmissions, and reception of many signals of MIMO antenna designs.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

ORCID

Kamlesh Patel https://orcid.org/0000-0001-8645-5230

REFERENCES

1. J. R. Hampton, M. A. Cruz, N. M. Merheb, A. R. Hammons, D. E. Paunil, and F. Ouyang, *MIMO channel measurements for urban military applications*, (MILCOM 2008–2008 IEEE Military Communications Conference, San Diego, CA, USA) 2008, pp. 1–7.
2. M. S. Sharawi, *Printed MIMO antenna systems: performance metrics, implementations, and challenges*, Forum for Electromagnetic Research Methods and Application Technologies (FERMAT). **1** (2014), 1–11.
3. H.C. Huang, *Overview of antenna designs and considerations in 5G cellular phones*, (2018 International Workshop on Antenna Technology (IWAT), Nanjing, China), 2018, pp. 1–4.
4. W. Zhang, Z. Weng, and L. Wang, *Design of a dual-band MIMO antenna for 5G smartphone application*, (2018 International Workshop on Antenna Technology (IWAT), Nanjing, China), 2018, pp. 1–3.
5. A. Zhao, Z. Ren, and S. Wu, *Broadband MIMO antenna system for 5g operations in mobile phones*, Int. J. RF Microw. Comput.-Aided Eng. **29** (2019), no. 10. https://doi.org/10.1002/mmce.21857
6. S. Pahadsingh and S. Sahu, *Four-port MIMO integrated antenna system with DRA for cognitive radio platforms*, Int. J. Electron. Commun. **92** (2018), 98–110.
7. J. K. Hong, *Performance analysis of dual-polarized massive mimo system with human-care IoT devices for cellular networks*, J. Sens. (2018), **2018**(2018). https://doi.org/10.1155/2018/3604520
8. F. Jolani, Y. Yu, and Z. Chen, *A novel broadband omnidirectional dual-polarized MIMO antenna for 4G LTE applications*, (2014 IEEE International Wireless Symposium, Xi'an, China), 2014, pp. 1–4.
9. P. Nirmal, A. B. Nandgaonka, and S. L. Nalbalwa, *A MIMO antenna: Study on reducing mutual coupling and improving isolation*, (2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology, Bangalore, India), 2016, pp. 1736–1740.
10. K. Yu, Y. Li, and X. Liu, *Mutual coupling reduction of a MIMO antenna array using 3-D novel meta-material structures*, Appl. Comput. Electromagn. Soc. J. **33** (2018), no. 7, 758–763.
11. F. Liu, J. Guo, L. Zhao, G. -L. Huang, Y. Li, and Y. Yin, *Dual-band metasurface-based decoupling method for two closely packed dual-band antennas*, IEEE Trans. Antennas Propag. **68** (2020), no. 1, 552–557.
12. F. Liu, J. Guo, L. Zhao, G. -L. Huang, Y. Li, and Y. Yin, *Ceramic superstrate-based decoupling method for two closely packed antennas with cross-polarization suppression*, IEEE Trans. Antennas Propag. **69** (2021), no. 3, 1751–1756.
13. C. A. Balanis, *Microstrip antennas*, In *Antenna theory: Analysis and design*, 3rd ed., John Wiley & Sons, Inc, Hoboken, New Jersey, 2016, 811–876.
14. F. Elek, R. Abhari, and G. V. Eleftheriades, *A uni-directional ring-slot antenna achieved by using an electromagnetic band-gap surface*, IEEE Trans. Antennas Propag. **53** (2005), no. 1, 181–190.
15. N. O. Parchin, Y. I. A. Al-Yasir, A. H. Ali, I. Elfergani, J. M. Noras, J. Rodriguez, and R. A. Abd-Alhameed, *Eight-element dual-polarized MIMO slot antenna system for 5G smartphone applications*, IEEE Access **7** (2019), 15612–15622.
16. J. Li, X. Zhang, Z. Wang, X. Chen, J. Chen, Y. Li, and A. Zhang, *Dual-band eight-antenna array design for MIMO applications in 5G mobile terminals*, IEEE Access **7** (2019), 71636–71644.
17. S. H. Chae, S. Oh, and S. O. Park, *Analysis of mutual coupling, correlations, and TARC in WiBro MIMO array antenna*, IEEE Antennas Wirel. Propag. Lett. **6** (2007), 122–125.
18. C. Wang, S. Xiao, W. Wang, C. Wang, and S. Liu, *An analytical approach for antenna performance evaluation for MIMO systems*, (2015 International Symposium on Antennas and Propagation, Hobart, Australia), 2015, pp. 1–4.
19. J. D. Baena, J. Bonache, F. Martin, R. M. Sillero, F. Falcone, T. Lopetegi, M. A. G. Laso, J. Garcia-Garcia, I. Gil, M. F. Portillo, and M. Sorolla, *Equivalent-circuit models for split-ring resonators and complementary split-ring resonators coupled to planar transmission lines*, IEEE Trans. Microw. Theory Tech. **53** (2005), no. 4, 1451–1461.
20. N. Ojaroudi Parchin, H. Jahanbakhsh Basherlou, M. Alibakhshikenari, Y. Ojaroudi Parchin, Y. I. A. Al-Yasir, R. A. Abd-Alhameed, and E. Limiti, *Mobile-phone antenna array with diamond-ring slot elements for 5G massive MIMO systems*, Electronics **8** (2019), no. 5, 521.
21. Q. Liu, H. Liu, W. He, and S. He, *A low-profile dual-band dual-polarized antenna with an AMC reflector for 5G communications*, IEEE Access **8** (2020), 24072–24080.
22. A. A. Khan, M. H. Jamaluddin, S. Aqeel, J. Nasir, J. R. Kazim, and O. Owais, *Dual-band MIMO dielectric resonator antenna for WiMAX/WLAN applications*, IET Microw. Antennas Propag. **11** (2017), no. 1, 113–120.
23. K. S. Sultan and H. H. Abdullah, *Planar UWB MIMO-diversity antenna with dual notch characteristics*, Prog Electromagn Res C Pier C **93** (2019), 119–129.

24. S. Tripathi, A. Mohan, and S. Yadav, *A compact Koch fractal UWB MIMO antenna with WLAN band-rejection*, Antennas Wirel Propag Lett **14** (2015), 1565–1568.
25. M. Shehata, M. S. Said, and H. Mostafa, *Dual notched band quad-element MIMO antenna with multitone interference suppression for IR-UWB wireless applications*, IEEE Trans. Antennas Propag. **66** (2018), no. 11, 5737–5746.
26. M. Bilal, R. Saleem, H. H. Abbasi, M. F. Shafique, and A. K. Brown, *An FSS-based nonplanar quad-element UWB-MIMO antenna system*, IEEE Antennas Wirel. Propag. Lett. **16** (2017), 987–990.

AUTHOR BIOGRAPHIES

Vipul Kaushal received his BSc degree in electronics and MSc degree in electronics from the University of Delhi, Delhi, India, in 2017 and 2019, respectively. Presently, he is a research scholar at the Department of Electronic Science, University of Delhi South Campus, New Delhi, India. His research interests are MIMO antenna designs and their application in 5G communications.

Amit Birwal received his BTech degree in electronics and communication engineering from Guru Gobind Singh Indraprastha University, Delhi, India, in 2003, and his MTech degree in microwave electronics and PhD degree from the University of Delhi, Delhi, India, in 2006 and 2021, respectively. From 2006 to 2013, he worked at various levels in the development and test phases of telecom devices at Ericsson India. Since 2014, he holds the position of assistant professor at the Department of Electronic Science, University of Delhi South Campus, New Delhi, India. His main research interests are designing and optimizing planar antennas for IoT-based systems, communications systems, and so forth.

Kamlesh Patel received his MSc degree in electronics and an MTech degree in microwave electronics from Rani Durgavati Vishwavidyalaya, Jabalpur, India, and the University of Delhi, India, in 1999 and 2003, respectively. He has a PhD from the Delhi Technological University, Delhi, India. From 2004 to 2013, he worked as a scientist at CSIR-National Physical Laboratory, India. Since 2013, he has been working with the Department of Electronic Science, University of Delhi South Campus, New Delhi, India, where he is now an associate professor. His research interests include microwave components, material characterization, and planar antennas for mobile communications.

How to cite this article: V. Kaushal, A. Birwal, and K. Patel, *Diversity characteristics of four-element ring slot-based MIMO antenna for sub-6-GHz applications*, ETRI Journal (2023), 1–13. https://doi.org/10.4218/etrij.2022-0113

Doped graphene characterized via Raman spectroscopy and magneto-transport measurements

Marie-Luise Braatz[1,2], Nils-Eike Weber[3,+], Barthi Singh[4,5], Klaus Müllen[4], Xinliang Feng[4*], Mathias Kläui[1#] and Martin Gradhand[1,6]

[1]Institute of Physics, Johannes Gutenberg University Mainz, Staudingerweg 7, 55128 Mainz, Germany

[2]Graduate School of Excellence Materials Science in Mainz, Staudingerweg 9, 55128 Mainz, Germany

[3]Carbon Materials Innovation Center (CMIC), BASF SE, 67056 Ludwigshafen, Germany

[4]Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany

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

[6]H. H. Wills Physics Laboratory, University of Bristol, Tyndall Ave, BS8-1TL, United Kingdom

# Corresponding author: klaeui@uni-mainz.de

+ Present address: Scienta Omicron GmbH, Limburger Strasse 75, 65232 Taunusstein, Germany

* Present address: Max Planck Institute of Microstructure Physics, Weinberg 2, 06120 Halle, Germany

Abstract

Functionalizing graphene beyond its intrinsic properties has been a key concept since the first successful realization of this archetype monolayer system. While various concepts, such as doping, co-doping and layered device design, have been proposed, the often complex structural and electronic changes are often jeopardizing simple functionalization attempts. Here, we present a thorough analysis of the structural and electronic properties of co-doped graphene via Raman spectroscopy as well as magneto-transport and Hall measurements. The results highlight the challenges in understanding its microscopic properties beyond the simple preparation of such devices. It is discussed how co-doping with N and B dopants leads to effective charge neutral defects acting as short-range scatterers, while charged defects introduce more long-range scattering centers. Such distinct behavior may obscure or alter desired structural as well as electronic properties not anticipated initially. Exploring further the preparation of effective pn-junctions, we highlight step by step how the preparation process may lead to alterations in the intrinsic properties of the individual layers. Importantly, it is highlighted in all steps how the inhomogeneities across individual graphene sheets may challenge simple interpretations of individual measurements.

Introduction

Chemical doping of graphene is widely used to modify the properties of graphene [1] adapting it to a wide set of applications from biosensing to batteries or catalysis [2–4]. Among the most commonly used dopants are nitrogen and boron [5], where the induced lattice distortion is relatively small [6]. The combination of both dopants in one sample has been proposed for supercapacitors [7] or biological applications [8]. Including further dopants such as the combination of nitrogen and sulfur [9,10] or boron and beryllium [11], an even larger variety of tuning properties is possible. While several proposals have been made, not many details are established about the characterization of the electronic structure and material properties of such co-doped samples. Furthermore, combining graphene layers with different dopants and doping levels opens an even wider range of combinations and thus property tuning, which is just starting to be explored.

Here, we present investigations on selected samples with different doping levels including co-doping to establish a clearer picture of the induced changes of the underlying graphene transport properties. To investigate the effect of different types of dopants on the physical properties of graphene, co-doped samples of nitrogen and boron are compared to only nitrogen-doped graphene. By using complementary optical and electronic measurement techniques, different properties such as structural order, electronic mobility as well as the transport relaxation times of the samples are ascertained.

Furthermore, we combine two differently doped layers into a heterostructure. This allows us to explore the possible creation of junctions based on graphene sheets with different doping levels.

Methods

The nitrogen-doped (sample A and B) as well as the boron and nitrogen co-doped samples were grown by chemical vapor deposition (CVD) on copper [12]. The dopants were incorporated by adding different precursors during the growth phase: varying amounts of NH_3 for the N-doped and the in-house synthesized organic precursor (B_2N_2 Dibenzo[a,e]pentalenes ($C_{30}H_{30}B_2N_2$)) BNNB-DBP for the B,N-co-doped sample. The transfer of the graphene from copper to Si/SiO_2 (p-doped Si covered with 300 nm SiO_2) followed a standard wet transfer protocol [13]. First, the graphene was covered with polymethyl-methacrylate (PMMA), then the copper was etched by ammonium persulfate (3%). Next, the samples were placed on water for cleaning and finally transferred to the Si/SiO_2 substrate. For the double transfer, a second graphene sheet was deposited on top of the first one by using the same procedure. A sketch of this configuration can be found in the corresponding section in Fig. 5a. For both types of transfer, the graphene pieces were approximately square with an edge length between 0.5 and 1 cm. Raman spectroscopy was performed at a wavelength of 532 nm. For the subsequent electronic measurements the samples were contacted with silverpaste in a four-probe geometry. An additional contact was placed on the backside to allow for electrical gating of the samples. The measurements were performed in a helium cryostat in vacuum at temperatures down to 3K and magnetic fields up to 8T applied perpendicular to the sample surface. Furthermore, on sample A, transmission electron microscopy (TEM) and x-ray photoelectron spectroscopy (XPS) were performed. Details on the experimental methods can be found in the supplementary information.

Results and Discussion

B and N co-doped samples

Raman spectroscopy was used to characterize the structure of the samples. In all cases the characteristic graphene peaks, 2D (~2679 cm^{-1} at 532nm) and G (~1580 cm^{-1} at 532nm), can be observed. In addition, the D peak (~1350 cm^{-1} at 532nm), which is only present for imperfect graphene [14,15], can be clearly identified (see inset Fig. 1(a)). The symmetric, single-peak, shape of the 2D-peak and the intensity ratio of the 2D- and G-peak, equal or larger than 2 (Fig. 1a), identify our samples as monolayer graphene with a reasonably low level of defect concentrations [16,17]. To confirm the monolayer nature of the samples, TEM was performed on sample A. The image in Fig. 2 shows an area of polycrystalline graphene with the grain boundary indicated by the red arrows. As has been shown before [18] the ratio of the intensities of 2D and G peak I(2D)/I(G) is generally taken to be a good measure of the defect concentration. For relatively pure graphene I(G) changes weakly while the intensity of the 2D peak becomes gradually lower with higher defect concentration. While this ratio is on average lower for the co-doped sample, the spread suggests

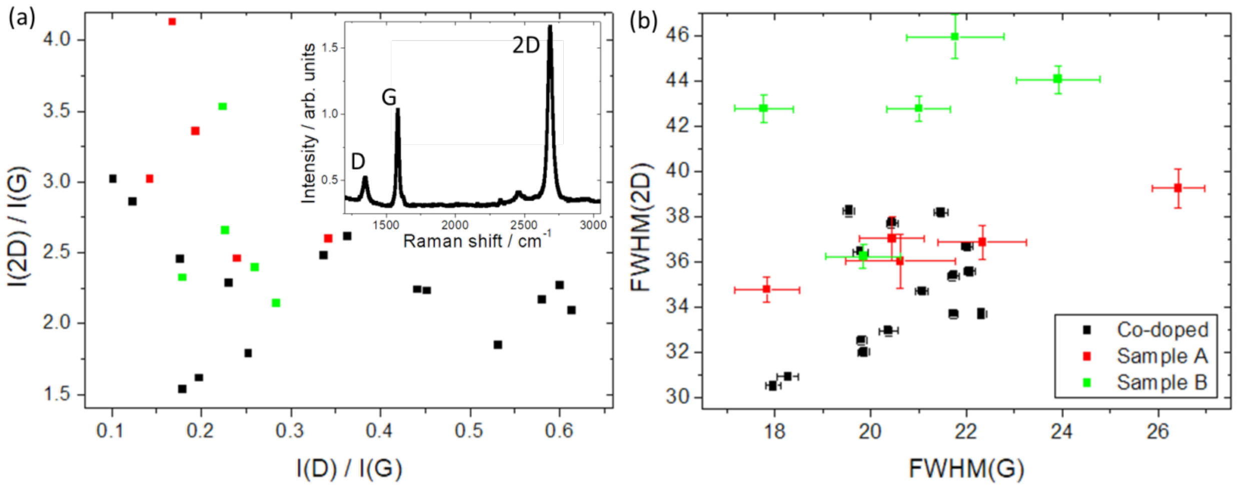


Figure 1: The Raman spectroscopy for all graphene samples with the N doped samples A (red) and B (green) and the Co-doped sample (black). a) Ratio of 2D to G peak intensity vs. the ratio of the D to the G peak with an inset of the Raman spectrum for the co-doped sample. b) The FWHM of the 2D peak vs. the FWHM of the G peak for all three samples.

significant inhomogeneities across the sample. However, the considerably larger values for the I(D)/I(G) ratio in the co-doped graphene points to a larger defect concentration at least on average for this system. Nevertheless, the scenario for the I(D)/I(G) is slightly more complicated as discussed previously [19]. Due to competing mechanisms this peak intensity will increase proportionally with the defect concentration but will decrease again beyond a certain threshold [20]. This behavior makes it difficult to identify the defect concentration solely by the I(D) intensity.

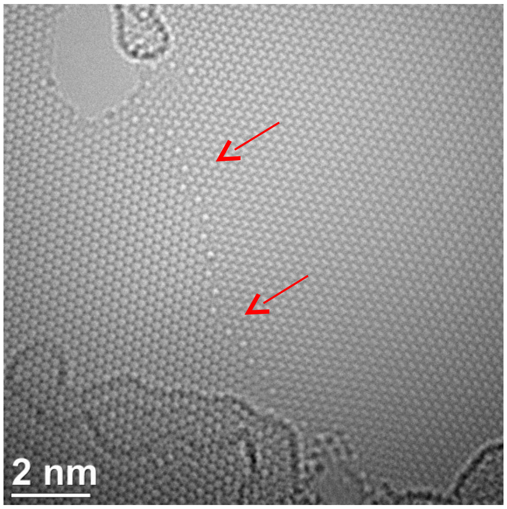


Figure 2: TEM image of sample A showing a grain boundary in monolayer graphene

In Fig. 1b we summarize the results for the full width at half maximum (FWHM) of the 2D as well as the G resonances. It is well established that the FWHM(G) will increase with increasing defect level, but the precise quantitative change will depend on the nature of the defect [21]. Any drastic increase of the FWHM of the 2D-peak would be associated with n-layer graphene but the maximal values observed here are substantially below any indication of multilayer graphene[17] and clearly identifies the systems as monolayer graphene. As for the intensities, the data is not conclusive and underlines the significant inhomogeneities across even individual samples. For any practical application this becomes a challenge as the properties might change rapidly across individual samples over reasonably short scales (probed by the spot of the laser in the Raman spectroscopy).

Finally, it is important to note the strong dependence of the intensities as well as the FWHM on the charge carrier concentration. Generally, the G resonances stiffen away from the Dirac point while the intensities of the D peak increase at the Dirac point [19]. This makes it difficult to draw direct conclusions from the Raman data as the different defect levels will change the scattering, the electron mobility, and the effective mass at the same time as the doping level and the defect level are affected simultaneously. This demonstrates that relying on only one probe to characterize individual samples will prove futile in many systems and that one needs to carry out measurements with significant statistics for robust information on doped graphene samples.

To compare the structural analysis by Raman spectroscopy to the electronic properties and further elucidate the types of scattering present in the samples, we measured the magnetoresistance (MR) in a cryostat at 3 K in a perpendicular magnetic field of up to B=8 T. The MR is determined from the sheet resistance $R_S(B)$

$$MR(B)=\frac{R_S(B)-R_S(0)}{R_S(0)}. \qquad (0)$$

The results for sample A and B are shown in Fig. 3 together with the curve for the co-doped sample measured furthest away from the Dirac point. The other MR curves of the co-doped samples, which were shifted closer to the Dirac point by electrical gating, are shown in the supplementary information. All measurements were performed sufficiently far away from the Dirac point (between -4 and -10 x 10^{12} cm^{-1}) so that the variation of MR with charge carrier density becomes negligible. Comparing the three curves in Fig. 3, all of them show a local maximum of the resistance at zero magnetic field due to weak localization. At high fields the trends are very different (Fig. 3 inset). For the two N doped samples, A and B, the MR stays negative, going down monotonically for sample B and flattening for sample A. For the co-doped sample, we observe a sign change at around B = 4 T. As shown by McCann et al. [22], a positive high field MR can be associated to a scenario where the scattering is dominated by intervalley scattering induced by short range defects. In contrast the negative high field MR can be understood in terms of intravalley scattering dominated by long range charged impurities. Thus, the data would imply that the co-doped sample leads to effectively charge neutral defects while the N doped samples feature extended charged impurities while all the while the overall disorder remains comparable as indicated by the Raman measurements.

This is further supported as the co-doped sample shows the highest MR indicating low doping [23,24] while the negative MR values for samples A and B indicate higher doping. The nitrogen-doping as measured on sample A by XPS is 1.0% ± 0.1%. This finding is also in line with the values of the charge carrier mobility, which is expected to be higher for higher quality graphene. Comparing the three samples, the highest value is found for the co-doped sample. The lowest value is found for the nitrogen-doped sample B. These values are measured in a Hall-geometry (in a cryostat at 3 K), where we used [23]

$$\frac{R_H}{R_S} = B\mu \qquad (2)$$

with R_H and R_S the Hall- and sheet resistance, the magnetic field B and the charge carrier mobility μ .

These results lead to an apparent contradiction with the co-doped sample showing a high MR and mobility, typically associated with pristine graphene [13,24], while at the same time showing the largest intensities for the defect induced Raman D peak. This can be resolved by distinguishing defects, or localized neutral impurities, from dopants, which act as charged long range scattering regions. In case of purely N doping the number of defects and dopants is equivalent, while having both nitrogen and boron atoms these dopants partially compensate. They all contribute as short-range localized defects enhancing the D peak intensity in the Raman measurement. However, the resulting small effective doping and thus an effectively small number of charged long range dopants will enhance the MR.

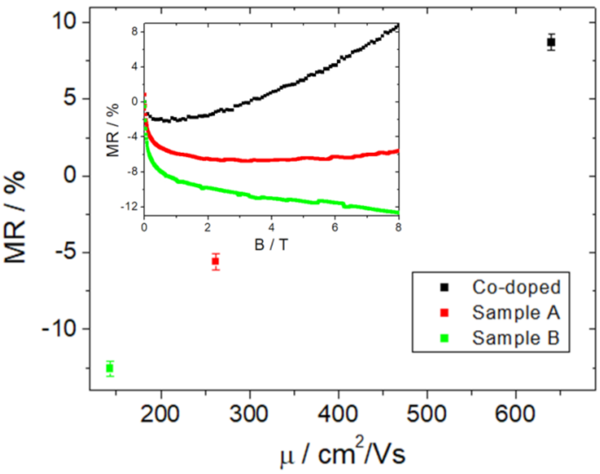


Figure 3: Comparison of the magnetoresistance at 8 T and the charge carrier mobility extracted from Hall-measurements. Inset: MR from 0 to 8 T for all three samples.

This qualitative discussion is further supported by a more quantitative analysis in terms of the electron scattering times. For the co-doped sample, MR measurements at different charge carrier densities were performed by applying gate voltages (Fig. SI1). They are all used within the quantitative analysis in this section. To extract the various scattering times, we fit the magnetoresistance for fields up to 0.3 T with a model that considers the phase coherence time τ_{ϕ} , the intervalley scattering time τ_i and the intravalley scattering $\tau_{¿}$ [22]:

$$\Delta R(B) = \frac{-e^2 \rho_s^2}{\pi h}\left[F(dB\tau_{\phi}) - F\left(\frac{dB}{\tau_{\phi}^{-1}+2\tau_i^{-1}}\right) - 2F\left(\frac{dB}{\tau_{\phi}^{-1}+\tau_i^{-1}+\tau_{¿}^{-1}}\right)\right] \quad (3)$$

with $d=(4eD/\hbar)$, D the diffusion constant, $F(z)=\ln(z)+\Psi(1/2+1/z)$ and $\Psi(z)$ the digamma function. Here, we calculated the intravalley scattering according to Moser et al. [25] directly to reduce the number of free parameters. The resulting intravalley scattering times are summarized in Table 1 and support the interpretation introduced above as the scattering times for the co-doped samples are roughly a factor of two larger than for the Nitrogen-doped samples, A and B. This again indicates the rather weak intravalley scattering for the co-doped samples with predominantly short-range scatterers.

Sample	A	B	Co	Co	Co	Co
Charge carrier density [10^{12} cm^{-2}]	-9.32	-5.7	-6.5	-5.7	-5.0	-4.4
Intravalley scattering time [10^{-15}s]	9.34	3.99	18.9	18.2	16.9	15.7

Table 1: The intravalley relaxation times as calculated from the electron mobility and the charge density according to Ref. 13. Errors as determined from the fit described therein are 0.1% for the charger carrier density and 1% for the intravalleyscattering time. The Fermi velocity used is v_F=1x10^6 m/s.

From the fitting, we find the corresponding intervalley scattering and phase coherence times as shown in Fig. 4. The intervalley scattering times for the N-doped samples are comparable, sample B, or smaller, sample A than those for the co-doped sample, which does not allow for an easy conclusion. However, the phase coherence time, an indicator for the existence of effective inelastic scattering is significantly larger for the co-doped samples. This points to stronger inelastic scattering for the N-doped samples which is indicative of a stronger scattering from charged, long range, perturbations in the N-doped in comparison to the co-doped samples.

Our results show that a complex interplay of intervalley and intravalley scattering caused by predominantly short-ranged defects and long-ranged dopants complicates simple predictions for co-doped samples. The theoretical predictions of band gap engineering in slightly artificial co-doped systems where the dopants form dimers [26] or specific configurations around DV(555-777) defects[27] are probably unrealistic in real life devices where defect formation and sample preparations are much more complex. Furthermore, ensuring the homogeneity of such properties of reasonably large structures becomes an even more difficult challenge. Our works show that standard procedures lead to sizable inhomogeneities across the samples with randomly placed B and N defects partially compensating the degree of doping and in turn leading to more short-range defects in contrast to the long-ranged dopants in standard N-doped samples.

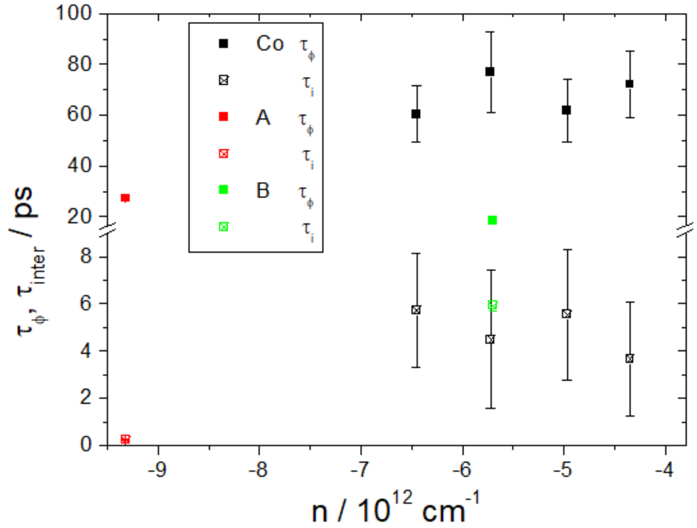


Figure 4: Phase coherence time and intervalley scattering time as extracted from the MR curve. For the co-doped samples different charge carrier densities were obtained by applying a gate voltage.

With that in mind we will explore a further method to exploit possible functionalities in differently doped graphene sheets by combining them. In the following, we discuss initial results for junctions prepared from differently doped graphene sheets brought into direct electronic contact via a region of twisted bi-layer graphene.

Lateral joining of differently doped graphene sheets

The lateral joining of areas with different doping type is well known from semiconductor physics leading to applications such as diodes and classical transistors [28,29]. To generate such a configuration using graphene, we deposited differently doped graphene sheets on top of each other. Their respective doping level with nitrogen is 0.2% for layer I and 1% for layer II and they are partly overlapping as sketched in Fig 5a. A microscope image of the different steps of the double transfer can be seen in Fig 6.

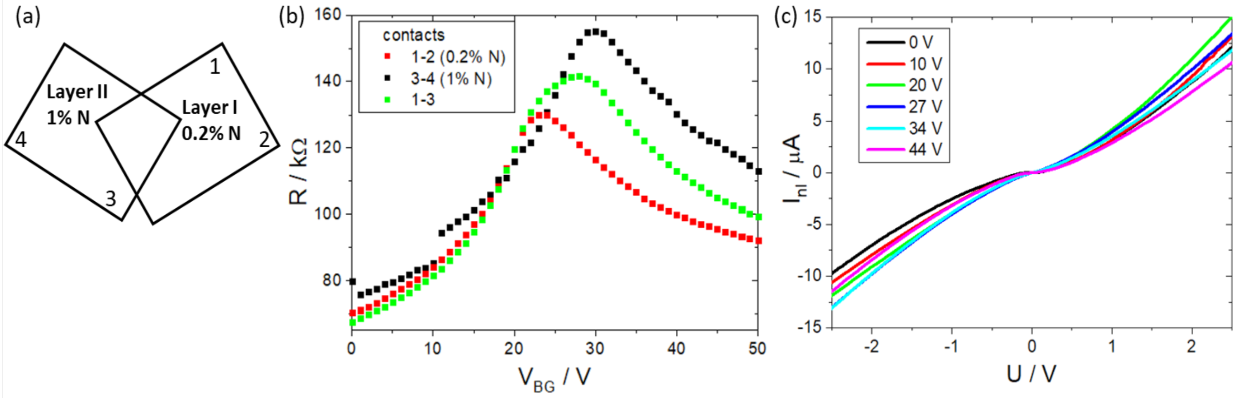


Figure 5: (a) Sketch of the sample. Layer I with 0.2% N is deposited and layer II with 1% N is deposited on top. For details see Methods. (b) Gate curves measured on different parts of the sample; contacts as depicted in (a). (c) The non-linear part of the IV curves between contacts 1-3 for different gate voltages. 27 V corresponds to an intermediary state between the charge neutrality points of the individual layers.

The sample was contacted via 4 silverpaste contacts, two on each layer. Measuring the gate dependence of the resistance (Fig. 4b), the charge neutrality point (CNP) is visible for the two layers with distinct levels of doping and changes depending on the doping from 23.5 V for 0.2% nitrogen to 30 V for 1% nitrogen. Overall, the CNP is placed at positive gate voltages due to adsorbed oxygen [30]. Measuring across the junction, the CNP is a superposition of the two individual layer peaks with a maximum in between and is slightly broadened. In the region between the two CNPs of 23.5 V and 30 V, layer I is in an n-doped state while layer II is in a p-doped state forming possibly a pn-junction. Measuring the IV-characteristics between contacts 1-3 an effective pn-junction ought to introduce stronger non-linear contributions. In Fig. 4c we present these non-linear contributions having removed the linear part as defined by the zero-field resistance depicted in Fig. 4b. In this representation, regions of horizontal IV characteristics indicate linear behavior and any deviation from this signifies non-linear contributions. From Fig. 4c it is clearly visible that this non-linear contribution becomes much more prominent in the region of gate voltages ranging from 20 V - 34 V which is precisely the region where we expect the effective pn junction to form. However, overall the effect is weak and to achieve a more pronounced pn-junction effect, it would be necessary to move the CNPs of the individual graphene layers further apart via more pronounced effects arising from doping. Presumably, this would require even more pure and clean pristine graphene flakes complicating the preparation of such junctions or nanopatterned devices where the junction forms a large part of the probed area.

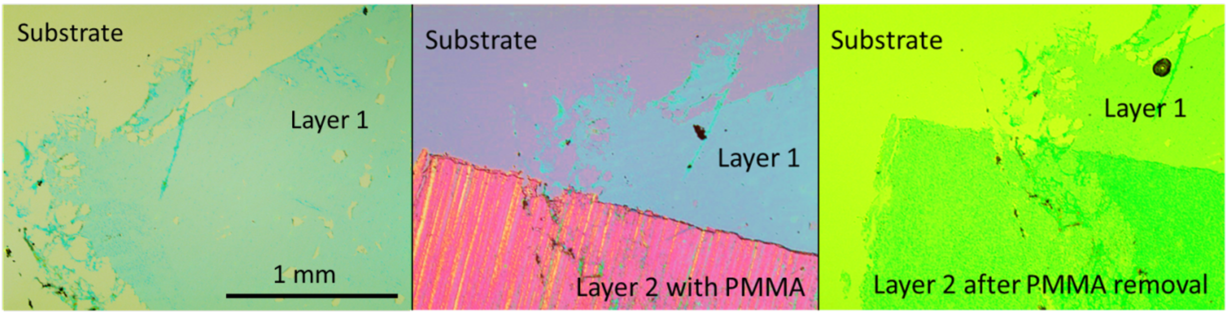


Figure 6: Microscope image of the different steps of the double transfer: First layer 1 is transferred and the PMMA removed (left image). The next layer is transferred with PMMA still present (middle image). The PMMA is removed with acetone (right image).

As the cleanliness becomes a crucial factor for these junctions, we investigated the impact of the second transfer on the properties of the first layer in more detail via Raman spectroscopy (see Fig. 6). Both, the intensity of the 2D peak as well as the intensity of the D peak is increased after the double transfer. While the increased 2D peak generally indicates more perfect graphene samples, the increased D peak is commonly associated with more defects [20]. However, as discussed above, the D peak intensity is not a monotonic function of the defect concentration, and it is hard to reach firm conclusion from the moderate change seen in Fig. 6 a. Furthermore, the sharp drop in the FWHM of the 2D peak after the transfer would point to less defects in line with the increase in the 2D intensity. This could be explained by adsorbates being removed during a second acetone bath and / or the subsequent heating. Fabricating a pn-junction via the presented transfer mechanism must take into account the impact of this procedure on the first layer. While the sample retains the basic graphene properties as shown here, the configuration of adsorbates or other defects may change considerably.

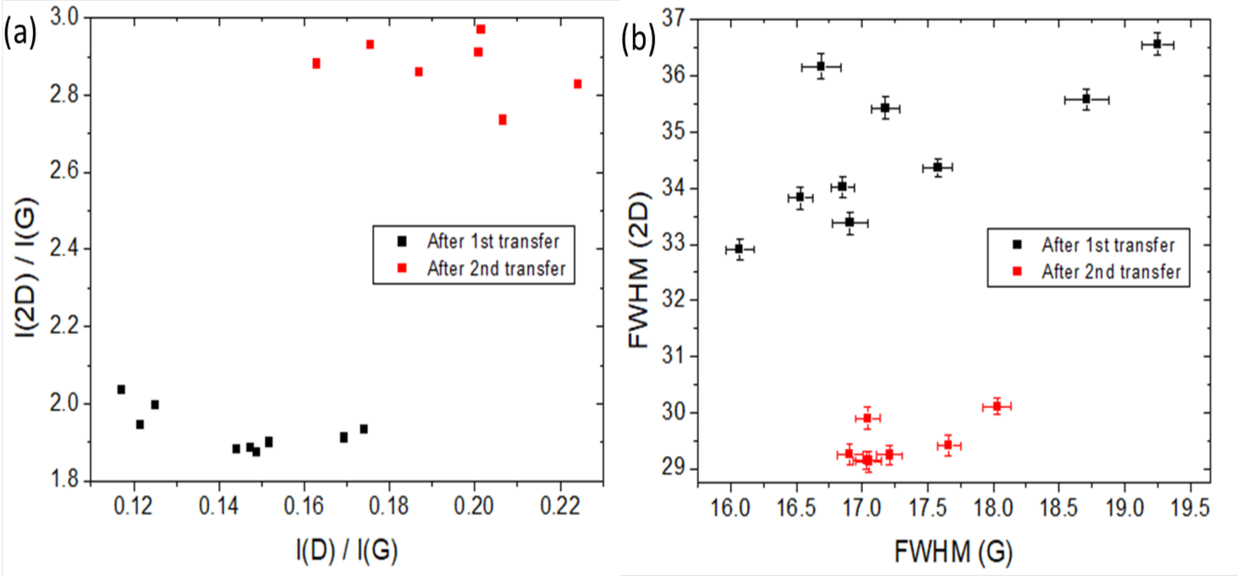


Figure 7: Comparison of the peak height of the 2D and D peaks as well as the FWHM of the 2D peaks from Raman spectroscopy for layer I before and after the second transfer. (a) Intensity of the 2D peak divided by that of the G peak as a function of the D peak intensity divided by that of the G-peak. (b) FWHM of the 2D peak as a functionpf that of the G peak.

Summary

Using Raman spectroscopy and magneto-transport measurements we explored the structural and electronic properties of functionalized graphene addressing the effects of co-doping with B and N impurities. Our findings highlight the complexity of the induced changes for the electronic properties of graphene-based devices. While dopants such as B and N separately will introduce charged long-range scatterers the combination of both lead to effective charge neutral and short-range scattering centers affecting the magneto-transport properties. Accordingly, we find an enhancement of the intravalley scattering time via co-doping as well as an increase of the phase coherence time, both pointing to more short-ranged and charge-neutral defects. The main findings of the Raman spectroscopy highlight the significant inhomogeneities across individual samples. This points to the importance of a careful analysis of any Raman data as any individual measurement could point to different conclusions. As such this analysis enforces the requirement to probe a variety of distinct lateral positions in the Raman spectroscopy via even an averaging approach might be misleading.

A similar finding is shown for the lateral joining of differently doped graphene sheets in an effective p-n junction, where the Raman spectroscopy highlights the induced changes during the preparation process. Characterizing the individual sheets will not trivially translate into the understanding of the finally realized device.

Supplementary Material

See supplementary material for more curves of the magnetoresistance at different charge carrier densities for the co-doped sample, an exemplary XPS spectrum for sample A and more details about the methods that were used.

Acknowledgements

M.-L.B thanks Hao Lu for XPS measurements and Amir Tavabi for TEM measurements. M.G. acknowledges the visiting professorship program of the Centre for Dynamics and Topology and Johannes Gutenberg-University Mainz. This work was financially supported by the DFG primarily through the Priority Program Graphene SPP 1459 (Grant No. 130170629), SFB TRR 173 Spin + X, Projects A01 and B02 (Grant No. 268565370), and the Forschunginitiative Rheinland-Pfalz through the Center for Dynamics and Topology (TopDyn).

Bibliography

[1] F. Joucken, L. Henrard, and J. Lagoute, Phys. Rev. Mater. **3**, 110301 (2019).
[2] Y. Wang, Y. Shao, D.W. Matson, J. Li, and Y. Lin, ACS Nano **4**, 1790 (2010).
[3] A.L.M. Reddy, A. Srivastava, S.R. Gowda, H. Gullapalli, M. Dubey, and P.M. Ajayan, ACS Nano **4**, 6337 (2010).
[4] Y. Wang, Y. Shen, Y. Zhou, Z. Xue, Z. Xi, and S. Zhu, ACS Appl. Mater. Interfaces **10**, 36202 (2018).
[5] L.S. Panchakarla, K.S. Subrahmanyam, S.K. Saha, A. Govindaraj, H.R. Krishnamurthy, U.V. Waghmare, and C.N.R. Rao, Adv. Mater. **21**, 4726 (2009).
[6] S. Ullah, Q. Shi, J. Zhou, X. Yang, H.Q. Ta, M. Hasan, N.M. Ahmad, L. Fu, A. Bachmatiuk, and M.H. Rümmeli, Adv. Mater. Interfaces **7**, 2000999 (2020).
[7] Z.-S. Wu, A. Winter, L. Chen, Y. Sun, A. Turchanin, X. Feng, and K. Müllen, Adv. Mater. **24**, 5130 (2012).
[8] M.S. Kim, S. Cho, S.H. Joo, J. Lee, S.K. Kwak, M.I. Kim, and J. Lee, ACS Nano **13**, 4312 (2019).
[9] R. Verma, I. Chakraborty, S. Chowdhury, M.M. Ghangrekar, and R. Balasubramanian, ACS Sustain. Chem. Eng. **8**, 16591 (2020).
[10] W. Ai, Z. Luo, J. Jiang, J. Zhu, Z. Du, Z. Fan, L. Xie, H. Zhang, W. Huang, and T. Yu, Adv. Mater. **26**, 6186 (2014).
[11] S. Ullah, A. Hussain, W. Syed, M.A. Saqlain, I. Ahmad, O. Leenaerts, and A. Karim, RSC Adv. **5**, 55762 (2015).
[12] Y. Ito, C. Christodoulou, M.V. Nardi, N. Koch, H. Sachdev, and K. Müllen, ACS Nano **8**, 3337 (2014).
[13] M.-L. Braatz, L. Veith, J. Köster, U. Kaiser, A. Binder, M. Gradhand, and M. Kläui, Phys. Rev. Mater. **5**, 084003 (2021).
[14] L.M. Malard, M.A. Pimenta, G. Dresselhaus, and M.S. Dresselhaus, Phys. Rep. **473**, 51 (2009).
[15] F. Tuinstra and J.L. Koenig, J. Chem. Phys. **53**, 1126 (1970).
[16] D. Graf, F. Molitor, K. Ensslin, C. Stampfer, A. Jungen, C. Hierold, and L. Wirtz, Nano Lett. **7**, 238 (2007).
[17] G.S. Papanai, I. Sharma, and B.K. Gupta, Mater. Today Commun. **22**, 100795 (2020).
[18] A. Eckmann, A. Felten, A. Mishchenko, L. Britnell, R. Krupke, K.S. Novoselov, and C. Casiraghi, Nano Lett 6 (2012).
[19] M. Bruna, A.K. Ott, M. Ijäs, D. Yoon, U. Sassi, and A.C. Ferrari, ACS Nano **8**, 7432 (2014).
[20] M.M. Lucchese, F. Stavale, E.H.M. Ferreira, C. Vilani, M.V.O. Moutinho, R.B. Capaz, C.A. Achete, and A. Jorio, Carbon **48**, 1592 (2010).
[21] S. Berciaud, S. Ryu, L.E. Brus, and T.F. Heinz, Nano Lett. **9**, 346 (2009).

[22] E. McCann, K. Kechedzhi, V.I. Fal’ko, H. Suzuura, T. Ando, and B.L. Altshuler, Phys. Rev. Lett. **97**, 146805 (2006).
[23] M. Rein, N. Richter, K. Parvez, X. Feng, H. Sachdev, M. Kläui, and K. Müllen, ACS Nano **9**, 1360 (2015).
[24] X. Li, J. Zhuang, Y. Sun, J. Bai, Z. Zafar, Z. Ni, B. Jin, and Z. Shi, Carbon **82**, 346 (2015).
[25] J. Moser, H. Tao, S. Roche, F. Alzina, C.M. Sotomayor Torres, and A. Bachtold, Phys. Rev. B **81**, 205445 (2010).
[26] M. Alattas and U. Schwingenschlögl, Sci. Rep. **8**, 17689 (2018).
[27] D. Sen, R. Thapa, and K.K. Chattopadhyay, ChemPhysChem **15**, 2542 (2014).
[28] M. Riordan and L. Hoddeson, IEEE Spectr. **34**, 46 (1997).
[29] J. Wang, Z. Li, H. Chen, G. Deng, and X. Niu, Nano-Micro Lett. **11**, 48 (2019).
[30] H.I. Wang, M.-L. Braatz, N. Richter, K.-J. Tielrooij, Z. Mics, H. Lu, N.-E. Weber, K. Müllen, D. Turchinovich, M. Kläui, and M. Bonn, J. Phys. Chem. C **121**, 4083 (2017).

2023 International Conference on Power Electronics and Energy (ICPEE) | 978-1-6654-7058-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICPEE54198.2023.10060649

Dynamic Combined Economic Emission Load Dispatch using Perfectly Convergent Particle Swarm Optimization

Devinder Kumar*
Department of Electrical Engineering,
G B Pant institute of Technology,
Delhi Skill and Entrepreneurship
University, Okhla Phase III, New Delhi,
India.
email:devdaksh@gmail.com
*Corresponding author

Narender kumar Jain
Department of Electrical Engineering,
Delhi Technological University,
Delhi, India.
email: vnarender84@yahoo.com

Nangia Uma
Department of Electrical Engineering,
Delhi Technological University,
Delhi, India.
email:uma_nangia@rediffmail.com

Abstract— **Determining the optimum performance from a number of power producing facilities in the short term to fulfil system demand with the aim of power forecasting at the lowest cost possible subject to transmission systems energy losses and operating parameters, is known as optimal power dispatch. Power balance constraints, generator limits, emission dispatch constraints, and valve point effects are a few instances of operational constraints. In order to manage the objective function and the operational constraints simultaneously, the proposed Perfectly Convergent Particle swarm optimization has been suggested in this research which outperformed various algorithms from recent literature in consistently giving excellent optimal solutions. The three dynamic test unit systems with three, ten and thirteen generators are considered in this article. For the accurate and efficient dynamic distribution of power, the effects of valve point loading, with or without emissions, prohibited operating zones and transmission line power loss are also taken into account.**

Keywords— ***dynamic combined economic emission dispatch, Perfectly convergent Particle swarm optimization, price penalty factor, non-smooth cost function, quadratic emission***

I. Introduction

The economic challenge facing the power system has been addressed in a number of ways that have been proposed and enacted. Some of the first approaches utilized were linear programming, Lagrangian relaxation, and the Lagrange multiplier. In order to improve the performance of currently used methods, like genetic algorithm (GA)[8], particle swarm optimization (PSO)[9,14], biogeography based optimization (BBO)[15], lambda iterative method[19], Evolutionary programming algorithm (EPA)[17], opposition based differential evolution (OBDE)[7], Equal embedded algorithm (EEA)[5], Multi objective Evolutionary algorithm (MOEA)[13], Trust region optimization (TRO)[16],Modified shuffled frog leaping algorithm (MSFLA) [18],Evolutionary algorithm base on Decomposition (EABOD)[20],Honey bees and Simulated annealing (HBSA)[21]and Hybrid CSA-JAYA algorithm (CSA-JAYA)[22]. Regrettably, the demand for the power system is always changing, necessitating a commensurate need for generator adaptation. This means that the output of the generator must grow when the demand for the load increases and vice versa. Therefore, in the dynamic load dispatch, the planning of generators associated with the grid is done in accordance with the fluctuating load at regular intervals with the goal of minimizing generation costs. Therefore, developing a population-based heuristic search approach that can prevent premature convergence while keeping the property of rapid convergence is still challenging. This research study takes into account quadratic cost and quadratic emission functions, which explain the appropriate operational cost of producing units using perfectly convergent Particle Swarm Optimization (PCPSO). The application of this approach produced excellent results quickly. The six portions of the paper are as follows: Section II discusses how the Combined economic and emission dispatch (CEED) problem was formulated. Section III discusses the PCPSO strategy. Section IV discusses the PCPSO implementation in the CEED problem. Section V discusses the findings and debate. The conclusion is the VI and last section.

II. Dynamic Combined Economic Emission problem formulation

The dynamic CEED problem is mathematically expressed and presented in this section, which includes the quadratic fuel cost function model, quadratic emission model, and max-max price penalty function.

A. *Quadratic fuel cost function*

As the initial objective of the committed generating units, coupled with equality and inequality requirements, the large component of the operating cost of thermal power plants is described as a second order of quadratic function:

$$Min\ F_{CT} = \sum_{i=1}^{n} a_i P_i^2 + b_i P_i + c_i \quad + \quad \left|\alpha_i \sin\left(\beta_i\left(P_{i,min} - P_i\right)\right)\right| \frac{\$}{h} \qquad (1)$$

Power balancing constraint: The sum of both the total real power generation and transmission losses is equal to the sum of total power demand.

$$\sum_{i=1}^{n} P_i = P_D + P_L \qquad (2)$$

Generator limit constraint: The actual power generation of i_{th} committed generating unit should be within following limit

978-1-6654-7058-2/23/$31.00 ©2023 IEEE

$P_{i\,min} \ll P_i \ll P_{i\,max}$ (3)

Transmission loss restriction: According to George's equation the overall transmission loss P_L should be kept minimum and is given as:

$P_L = \sum_{i=1}^{n}\sum_{j=1}^{n} P_i B_{ij} P_j$ (4)

Where $F_{CT}$is the cost of fuel of all generators in \$/h,P_i is the actual output power in MW of i_{th} generator,P_D , P_L are total demand and transmission losses in MW,$P_{i\,min}$, $P_{i\,max}$are the minimum and maximum power restriction of i_{th} generator, n is the number of committed generating units,a_i, b_i, c_i, d_i are the co-efficient fuel cost curve of the i_{th} generators respectively.$B_{ij}$is the matrix of transmission loss coefficient of generating units.

B. Quadratic Emission function

Hazardous gases such as SO2, NOx, and CO2 are produced by every thermal power station as a result of the combustion of fossil fuels, which add to the overall emissions and must be reduced individually. All three emissions are mathematically defined in this model using quadratic polynomials as follows:

$E_T = \sum_{i=1}^{n}(d_i P_i^2 + e_i P_i + f_i) + \gamma_i exp(\delta_i P_i)\ \ Kg/h$ (5)

Whereas E_T is the overall emission with valve loading effect in ton/h, d_i, e_{i,f_i} are emission coefficients of i^{th}generating unit in ton/MW^2h, ton/MWh and ton/h , γ_i *and* δ_i are the valve point loading effect emission coefficient of i^{th}generating unit.

C. Price Penalty Factors (PPF)

Price penalty factors are calculated by dividing the cost of fuel by the value of emission and are used to transform emission parameters into comparable fuel price. The Max-Max price penalty factor, h_i employed in this paper are listed below.

$$h_i = \frac{(a_i P_{i\,max}^2 + b_i P_{i\,max} + c_i) + \left|\alpha_i \sin\left(\beta_i (P_{i,min} - P_i)\right)\right|}{(a_i P_{i\,max}^2 + b_i P_{i\,max} + c_i) + \gamma_i exp(\delta_i P_i)} \qquad (6)$$

D. Bi-objective CEED

The bi-objective CEED equations are shown below, which incorporate cost of fuel with each emission and are then converted to a mono objective by multiplying a price penalty factor for each of the three pollutants independently.

$$F_T = \sum_{i=1}^{n}[(a_i P_i^2 + b_i P_i + c_i) + \left|\alpha_i \sin\left(\beta_i (P_{i,min} - P_i)\right)\right| + h_i\left(a_i\, P_i^2 + b_i\, P_i + c_i + \gamma_i exp(\delta_i P_i)\right)]\frac{\$}{h} \qquad (7)$$

III. Proposed algorithm as Perfectly Convergent Particle Swarm Optimization (PCPSO)

The purpose of this proposed variant [1-3] in our scenario is to eliminate early premature convergence, which contributes to stagnation, and to allow personal best particles to replace global particles since they allow more search space exploration. . Initialized with candidate solutions of particles moving through the search space, each particle having a position and velocity, and updates as follows:

$$x_j(k+1) = x_j(k) + v_j(k+1) \qquad (8)$$

$$v_j(k+1) = \omega v_j(k) + c_1(K) r_1\left(p_j(k) - x_j(k)\right) + c_2(K) r_2\left(g(k) - x_j(k)\right) \qquad (9)$$

Where, j =1, 2, 3 … n

$C_1(K), C_2(K)$are asynchronous learning factors with dynamic non-linear self-adjustable features which have highly likelihood of converging to optimum global solution.

k+1 denotes next iteration, k is the current iteration number, v_j is velocity of the particle j,x_j is position of the particle j, ω is Inertia weight factor,c_1,c_2 are acceleration factors,$p_j$is personal best of particle j, g is the global best of the entire swarm,r_1,r_2 are pseudo random numbers between 0 and 1. ω_{max}, ω_{min} are having maximum and minimum value of 0.9 and 0.4 of inertia weights

I've included an new particle in this new version, identical to the one used in [4] Guaranteed convergence particle swarm optimization (GCPSO), but instead of looking for global position, it will hunt for personal best position. Searching areas close to global position, while taking into consideration the current velocity update, restricts exploration and increase the chance of becoming trapped in multi-modal situations with one or even more local minima.

$$v_j'(k+1) = -x_j{}'(k) + pbest(k) + \omega v_j'(k) + \rho(k)(1 - 2r) \qquad (10)$$

Other particles in the swarm, on the other hand, will adjust their velocity according to this new variant:

$$v_j'(k+1) = \omega(k) x_j{}'(k) + c_1 r_1\left(p_j(k) - x_j(k)\right) + c_2 r_2\left(-x_j(k)\right) \qquad (11)$$

$$\omega(k) = \omega_{max} - k \times (\omega_{max} - \omega_{min}) \div K_{max} \qquad (12)$$

$$C_1(k) = 1.167 \times \omega(k)2 - 1.167 \times \omega(k) + 0.66 \qquad (13)$$

$$C_2(k) = 3 - C_1(k) \qquad (14)$$

Where, $-x_j{}'(k) + pbest(k)$ element will conduct the hunt in the personal best zone, $\omega v_j'(k)$ provides the momentum to search in current trajectory, ρ(k) (1-2r) generates a random search in the vicinity area of personal best particle with mean distance of $2\rho(k)$,where ρ(k) is the diameter of stochastic search space defined as follows:

$$\rho(k+1) = \begin{cases} 2\rho(k) & Successes > sc \\ (0.5)\rho(k) & failure > fc \\ \rho(k) & otherwise \end{cases} \qquad (15)$$

#successes (k+1)> #successes (k) => # failures (k+1) =0 and, # failures (k+1)> # failures (k) => #successes (k+1) =0

Where, #successes & # failures are the number of consecutive successes or failures, s_c= 15 & f_c =5 are the threshold parameters and can be precisely configurable.

This method uses an adaptable to choose the ideal sampling volume in its current variant. The maximum distance travelled in a single movement can be increased if a particular value of consistently produces a favorable outcome. The sampling volume needs to be reduced when on the other hand it delivers numerous failures. There won't be a stoppage at the end of the day if all phases were greater than zero.

In essence, this technique creates a real global search technique by allowing all the particles to compete, regardless of whether they are in the exploratory stage, have a greater personal best than the previous iteration or are on the verge of global optimum. These limitations of GCPSO are overcome by this method.

IV. Execution of PCPSO in CEED

Step1. Specify the lowest and maximum bounds for the generation each unit as well as the load demand.

Step2. Produce particles at random between min and max operating restrictions of the N units for a population size ' S'

in the j^{th}-dimensional space, using the $i^{th}\,particle\ \ P_i = [(P_{i1}^n, P_{i2}^n, P_{i3}^n \dots P_{iN}^n)]$where i=1, 2...S.
The following equation yields results that satisfy the generation restriction condition of (15) and having ‘r is a uniformly distributed random number between 0 and 1 following equation results.

$$P_{ij}^n = P_{min} + r\big(P_{ij\,max} - P_{ij\,min}\big) \qquad (16)$$

Step3: Constraints imposed by no operating zones
The generation value corresponding to the levels (P_{ij}^{lower}) or higher (P_{ij}^{upper}) boundary is adjusted and supplied to any element P_{ij} of the initial population (or updated population) if it is found to be inside the k^{th} restricted operating zone. The midway of the k^{th} restricted zone's midpoint $P_{mid,k}$.

$$P_{ij} = \begin{cases} P_{ij}^{lower} & if\ P_{ij}^{lower} \le P_{ij} < P_{mid,k} \\ P_{ij}^{upper} & if\ P_{mid,k} \le P_{ij} < P_{ij}^{upper} \end{cases} \quad (17)$$

Step4. Set particle velocity in the [$v_i^{min} v_i^{max}$] in N-dimensional space.
Step5. Evaluate the equation to assess fitness of each individually using the equation (1, 5, and 7).
Step6: The parameters are iteratively changed to improve fitness. The parameters of PSO are updated using equations (8-15).
Step7: The evaluation function values for the changed particle positions are found. If the new value is superior to the previous one, PSO sets the new value to pbest. The value of gbest’s value is modified to reflect that it is the best vector among pbest.
Step8. Stop criteria

The position of particles is denoted as Gbest for the optimum solution and stop if equation (17) is less than the stagnation threshold of$\varepsilon = 1x10^{-6}$.

V. Simulation Results and Discussion

The CEED problem was solved using the proposed PCPSO methodology using three different test platforms. To achieve this, we developed a software in the MATLAB 2015a environment on a Compaq 6720s lab-top with 4GB RAM and tested it on three different test unit systems with three units, six units, and thirteen units, and took into account the losses in variability transmission as well as other constraints. Simulation parameters has 20 number of particles in swarm , 250 is the number of iterations , number of trails taken as 5, with linearly decreasing inertia weight with maximum and minimum inertia w=0.9 and 0.4, acceleration constants c1 = c2 =2 of proposed PCPSO.

Table II. Shows the comparison of cost of fuel and power loss with PCPSO and other techniques

Parameter	PCPSO	PSO [9]	OBDE [7]	GA [8]	EEA [5]	Lambda iterative method
P_1 (MW)	72.48	73.81	73.72	73.83	73.97	73.52
P_2 (MW)	69.80	69.94	70.81	69.95	69.66	69.50
P_3 (MW)	76.56	74.99	75.46	75.02	75.16	75.78
P_L (MW)	8.50	8.81	8.80	8.87	8.80	8.81
Fuel cost ($/h)	3151.26	3162.93	3161.41	3163.63	3163.69	3163.9

“Test case system 1”

The CEED difficulties are explored and tested on the three generators system [5] at a base load of 210 MW to demonstrate the efficacy of the PCPSO technique for tackling the CEED problem with line flow constraints. Optimum generator scheduling was completed using the PCPSO method, taking into account fuel cost co-efficient, emission coefficients and B-loss coefficient and all other system constraints. Following table I shows the performance of PCPSO applied to 24 hours daily load pattern mentioned in [6].
The suggested PCPSO algorithm is contrasted with the most recent research articles' algorithms PSO, GA, OBDE,EEA , and lambda iterative method with the lowest fuel cost of 3151.26 $/h, F_{CEED} of 3162.91$/h with a power loss P_L of 8.50 MW, the simulation's findings exhibit excellent convergence properties. Following table II shows the comparison of PCPSO with other techniques.

Table I. shows the fuel cost with power losses applied to 24 hour load pattern using PCPSO.

Hour	Load (MW)	P_1 (MW)	P_2 (MW)	P_3 (MW)	P_L (MW)	Fuel cost($/h)
1	199	75.38	85.73	45.89	7.89	3011.83
2	189	97.76	46.80	52.87	8.01	2827.98
3	168	70.10	56.79	46.89	6.39	2622.68
4	157	53.68	53.01	56.09	6.53	2503.85
5	147	50.21	54.13	55.89	6.38	2312.07
6	206	105.67	45.02	63.76	8.21	3123.45
7	199	75.38	85.73	45.89	7.89	3011.83
8	195	82.35	74.89	45.01	7.89	2956.34
9	210	72.48	69.80	76.56	8.50	3131.26
10	231	91.56	74.67	77.68	10.42	3437.86
11	179	49.58	47.63	90.85	10.01	2756.53
12	189	97.76	46.80	52.87	8.01	2827.98
13	168	70.10	56.79	46.89	6.39	2622.68
14	199	75.38	85.73	45.89	7.89	3011.83
15	206	105.67	45.02	63.76	8.21	3123.45
16	168	70.10	56.79	46.89	6.39	2622.68
17	189	97.76	46.80	52.87	8.01	2827.98
18	157	53.68	53.01	56.09	6.53	2503.85
19	252	82.31	58.34	120.21	12.86	3657.36
20	231	91.56	74.67	77.68	10.42	3437.86
21	210	72.48	69.80	76.56	8.50	3151.26
22	189	97.76	46.80	52.87	8.01	2827.98
23	147	50.21	54.13	55.89	6.38	2312.07
24	147	50.21	54.13	55.89	6.38	2312.07

“Test case system 2”

An IEEE thermal system with six units of generation and valve point effects are present in this scenario. It is [10, 11, 12] that provides the coefficient of fuel cost matrix, generator constraint matrix, pollution coefficient matrix, and transmission loss coefficient matrix. The prize penalty factor for each of the six committed units (23.18, 32.70, 20.42, 25.53, 10.93 and 12.01) has been determined. Table III compares the results of utilizing PCPSO to solve CEED problem for a base load 283MW to those obtained using alternative methods. In comparison to , MOEA, PSO, BBO, TRO and OBDE, the proposed PCPSO algorithm achieves the optimal global minimum solution with a minimal number of iterations and processing time, at a cost of 610.12 $/h, lower fuel cost with emission cost and lowest combined economic emission dispatch. Applying the PCPSO technique to the 24 hour daily load pattern presented in table III has further confirmed its dynamic capacity. Table IV displays the CEED cost, emissions, and transmission loss for the generating units at various times.

“Test case system 3”

Table III. Shows the CEED cost, emission and power loss for 24 hour load pattern using PCPSO.

Hour	Load demand (MW)	P_1 (MW)	P_2 (MW)	P_3 (MW)	P_4 (MW)	P_5 (MW)	P_6 (MW)	P_L (MW)	Emission (ton/h)	F_{CEED} ($/h)
1	269	31.43	48.35	49.96	56.51	48.53	36.02	2.80	18.85	603.50
2	255	29.52	47.61	47.70	52.37	46.26	33.09	2.54	8.97	585.03
3	226	25.57	46.08	43.03	43.80	41.56	27.02	2.07	19.33	537.75
4	212	23.67	45.34	40.80	39.68	39.30	24.08	1.87	19.56	510.13
5	198	21.78	44.60	38.57	35.56	37.04	21.14	01.69	19.82	478.96
6	277	32.53	48.78	51.25	58.88	49.83	37.69	2.96	18.80	608.13
7	269	31.43	48.35	49.96	56.51	48.53	36.02	2.80	18.85	603.50
8	263	30.61	48.04	48.99	54.73	47.55	34.76	2.69	18.90	602.97
9	283	33.35	49.10	52.22	60.66	50.81	38.94	3.08	18.77	610.12
10	311	37.19	50.58	56.77	68.98	55.38	44.79	3.69	18.71	696.58
11	241	27.61	46.87	45.45	48.23	43.99	30.16	2.30	19.13	535.25
12	255	29.52	47.61	47.70	52.37	46.26	33.09	2.54	8.97	585.03
13	226	25.57	46.08	43.03	43.80	41.56	27.02	2.07	19.33	537.75
14	269	31.43	48.35	49.96	56.51	48.53	36.02	2.80	18.85	603.50
15	277	32.53	48.78	51.25	58.88	49.83	37.69	2.96	18.80	608.13
16	226	25.57	46.08	43.03	43.80	41.56	27.02	2.07	19.33	537.75
17	255	29.52	47.61	47.70	52.37	46.26	33.09	2.54	8.97	585.03
18	212	23.67	45.34	40.80	39.68	39.30	24.08	1.87	19.56	510.13
19	340	41.19	52.13	61.51	77.63	60.13	50.82	4.40	18.80	765.89
20	311	37.19	50.58	56.77	68.98	55.38	44.79	3.69	18.71	696.58
21	283	33.35	49.10	52.22	60.66	50.81	38.94	3.08	18.77	610.12
22	255	29.52	47.61	47.70	52.37	46.26	33.09	2.54	8.97	585.03
23	198	21.78	44.60	38.57	35.56	37.04	21.14	01.69	19.82	478.96
24	198	21.78	44.60	38.57	35.56	37.04	21.14	01.69	19.82	478.96

Table IV shows the comparison of PCPSO with other techniques for various parameters.

Parameter	MOEA[13]	PSO [14]	BBO [15]	TRO [16]	OBDE [7]	PCPSO
P_1(MW)	27.52	26.95	26.25	33.56	32.45	33.35
P_2(MW)	37.52	38.12	37.70	32.32	20.57	49.10
P_3(MW)	57.96	54.47	57.60	57.13	66.50	52.22
P_4(MW)	67.70	67.90	67.35	66.37	109.12	60.66
P_5(MW)	52.83	54.47	53.77	57.85	43.39	50.81
P_6(MW)	42.82	43.83	42.70	39.5	14.34	38.94
P_L(MW)	3.35	2.74	2.37	3.73	2.66	3.08
F_{CEED}($/h)	617.37	616.12	615.22	611.45	610.22	610.12
Emission (ton/h)	20.00	20.00	20.02	20.08	20.00	18.77

The test system with 13 generating units and a base load of 1800 MW is the third test case that is being taken into consideration for deployment. In [10,11] and [12], specifics of the test case's co-efficient of fuel cost, pollution co-efficient, and transmission loss B co-efficient are provided. Table VI contains comparisons between the acquired results and results from the literature. The table findings demonstrate that the suggested strategy performs better than the various methods and provided a far more ideal result in all cases. Applying the PCPSO method to a 24 hour daily load pattern has further proven its core competencies. Through Table V, the cost of the committed units and the overall cost at various times are displayed. The Table V shows that, for various loadings, the PCPSO approach has delivered the best results.

VI. Conclusion

PCPSO has been developed in this work to address dynamic CEED problems in power systems. A number of test scenarios were reviewed for the PCPSO's effectiveness, and the results were compared to those of current research studies. It has been established that PCPSO is a better alternative method for resolving dynamic CEED problems, particularly in large-scale power systems with valve point impact. The results of the three test systems shows the effectiveness of this algorithm. Additionally, PCPSO demonstrates avoiding premature convergence in local minima, which has a positive impact on emissions, computational efficiency, and convergence.

Table VI. Shows the CEED cost, emission and power loss comparison of PCPSO with other techniques.

Parameter	PCPSO	EPA [17]	BBO [10]	MSFLA [18]	OBDE [7]
P_1(MW)	270.43	80.69	179.51	540.52	271.91
P_2(MW)	358.65	166.30	299.19	225.07	360.89
P_3(MW)	286.89	166.87	297.57	207.96	288.28
P_4(MW)	185.79	154.77	159.73	69.09	186.26
P_5(MW)	157.08	155.41	159.73	84.96	155.64
P_6(MW)	58.78	154.86	159.73	94.76	60.23
P_7(MW)	59.69	154.72	159.73	106.97	61.54
P_8(MW)	60.00	154.52	60.00	109.00	60.57
P_9(MW)	58.12	154.76	60.00	108.32	60.38
P_{10}(MW)	122.85	119.43	40.00	79.63	120.45
P_{11}(MW)	46.81	119.29	114.76	63.76	46.29
P_{12}(MW)	95.03	109.20	55.00	58.06	96.84
P_{13}(MW)	53.37	109.12	55.00	72.96	55.56
P_L (MW)	18.09	---	---	21.09	18.14
F_{CEED} ($/h)	18069.25	18104.04	18081.48	17944.84	18072.24
Emission (ton/h)	84.29	---	95.30	---	84.33

Table V. Shows the results obtained from PCPSO for 24 hour daily load pattern.

Load (MW)	**1260 (MW)**	**1350 (MW)**	**1440 (MW)**	**1530 (MW)**	**1620 (MW)**	**1680 (MW)**	**1710 (MW)**	**1765 (MW)**	**1800 (MW)**	**1980 (MW)**	**2160 (MW)**
Hours →	**5,23,24**	**4,18**	**3,13,16**	**11**	**2,12,17,22**	**8**	**1,7,14**	**6,15**	**9,21**	**10,20**	**19**
P_1 (MW)	89.78	320.76	401.21	277.89	346.67	402.45	227.34	350.23	270.43	281.67	445.87
P_2 (MW)	341.45	58.02	57.47	23.67	301.78	37.56	350.09	291.79	358.65	210.68	350.58
P_3 (MW)	21.23	45.21	143.34	38.58	33.69	306.78	215.68	324.08	286.89	351.79	309.34
P_4 (MW)	63.89	63.78	61.09	59.35	158.53	58.89	107.83	97.67	185.79	160.07	152.07
P_5 (MW)	71.00	69.34	136.41	165.63	150.23	104.42	161.07	65.56	157.08	58.98	95.78
P_6 (MW)	70.00	159.56	179.92	158.49	66.57	98.71	98.69	75.47	58.78	142.45	178.21
P_7 (MW)	112.12	59.03	105.23	167.45	61.47	123.78	59.13	61.02	59.69	109.49	110.68
P_8 (MW)	181.34	178.68	59.79	135.79	59.51	178.96	167.56	71.23	60.00	165.35	108.89
P_9 (MW)	120.01	154.47	111.48	110.27	180.92	58.13	62.25	110.32	58.12	160.68	62.47
P_{10} (MW)	43.34	37.91	59.63	114.86	52.39	82.09	160.78	96.45	122.85	150.01	45.68
P_{11} (MW)	58.02	47.01	32.14	122.59	85.73	79.67	63.77	76.89	46.81	105.68	118.28
P_{12} (MW)	57.67	108.23	52.17	126.42	57.81	95.78	91.08	73.09	95.03	53.78	81.09
P_{13} (MW)	45.16	52.68	51.89	23.37	73.49	68.98	51.67	78.63	53.37	51.09	97.89
P_L (MW)	12.02	13.96	14.38	14.97	15.64	17.21	17.36	17.89	18.09	18.91	20.27
F_{CEED} ($/h)	14141.68	14854.79	15397.68	16393.62	17103.69	17451.59	18006.48	18047.56	18069.25	19879.98	21500.32
Emission(ton/h)	83.36	83.34	86.39	87.68	85.01	85.89	85.37	84.21	84.29	83.75	88.59

As a result,PCPSO optimization is a practical approach for dealing with complex problems in power systems. Future use of this suggested technique is for multi-area power systems integrated with wind farms and solar power systems, smart micro grid energy management having fossil fuelled generators and distributed generator which emit toxic and harmful pollutants are part of the work's future scope.

REFERENCES

[1] J. Kennedy and R. Eberhart. “Particle swarm optimization”, in Proceedings of the IEEE International Conference on Neural Networks .IEEE Service Centre, Piscataway, NJ, Vol IV, pp.1941-1948, 1995.

[2] Devinder kumar, N. k. Jain, Uma Nangia, ”Perfectly convergent Particle swarm optimization in multi-dimensional space”, International journal of Bio-inspired computation, Inderscience Publications, Vol 18,no 4,p.221228,2021.

[3] Devinder kumar, N .k. Jain, Uma Nangia,” Combined Economic Emission Dispatch using perfectly convergent Particle swarm optimization”, DELCON, IEEE Delhi section Conference, Delhi,DOI:10.1109/DELCON54057.2022.9752941 2022.

[4] N.K. Jain, Uma Nangia, Devinder Kumar, “Machine learning through back propagation networks using OPSO with Cauchy mutation and GCPSO in higher dimensions“ in proceedings of the IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems(ICPEICES) Electrical Engineering Department, Delhi Technological University, Delhi, pp.1-6, 2018.

[5] K. Chandram, N. Subrahmanyam, M. Sydulu, Equal embedded algorithm for economic load dispatch problem with transmission losses, Electrical Power and Energy Systems Vol 33, pp. 500-507, 2011.

[6] Sathis kumar, M., Nirmal kumar, A., Lakshminarasimman, L., Thiruvenkadam S., “A self-adaptive hybrid differential evolution algorithm for phase balancing of unbalanced distribution system”, Electrical Power and Energy Systems Vol 42, pp. 91-97, 2012.

[7] Thenmalar K, Ramesh S and Thiruvenkadam S, “Opposition Based Differential Evolution Algorithm for Dynamic Economic Emission Load Dispatch (EELD) with Emission Constraints and Valve Point Effects”, J Electrical Engineering Technology, Vol. 10(4): pp.1508-1517, 2015.

[8] Amjady N, Nasiri-Rad H,” Economic dispatch using an efficient real-coded Genetic algorithm”, IET Generation Transmission Distribution, Vol 3(3), pp. 266-78, 2009.

[9] Kumar R, Sharma D, Sadu A,” A hybrid multi-agent based particle swarm Optimization algorithm for economic power dispatch”, Int J Electrical Power Energy System, Vol 33(1), pp.115-23, 2011.

[10] S. Rajasomashekar, P. Aravindha babu, “Biogeography based optimization technique for best compromise solution of economic emission dispatch” Swarm and evolutionary computation Vol 7, pp.47-57, 2012.

[11] Hadi Hamedi,” Solving the combined economic load and emission dispatch problems using new heuristic algorithm”, Electrical Power and Energy Systems Vol 46, pp.10-16, 2013.

[12] Provas Kumar Roy, Sudipta Bhui,” Multi-objective quasi-oppositional teaching learning based optimization for economic emission load dispatch problem”, Electrical Power and Energy Systems Vol 53, pp.937-948, 2013.

[13]M.A. Abido, “Multi objective evolutionary algorithms for electric power dispatch problem”, IEEE Transactions on Evolutionary Computation Vol 10(3), pp.315-329, 2006.

[14] Kumar R, Sharma D, Sadu A.” A hybrid multi-agent based particle swarm Optimization algorithm for economic power dispatch”, Int J Electrical Power Energy System, Vol 33(1), pp.115-23, 2011.

[15] S. Rajasomashekar, P. Aravindhababu, “Biogeography based optimization technique for best compromise solution of economic emission dispatch” Swarm and evolutionary computation 7, 47-57, 2011.

[16] El-sobkya, Yousria Abo-elnaga, “Multi-objective economic emission load dispatch problem with trust-region strategy bothina”, Electric Power Systems Research Vol 108, pp.254- 259, 2014.

[17] Sinha N, Chakrabarti R, Chattopadhyay PK, “Evolutionary programming Techniques for economic load dispatch”, IEEE Trans Evolutionary Computation, Vol , 7(1),pp. 83-94,2003.

[18] Priyanka Roya, Pritam Royb, Abhijit Chakrabarti,” Modified shuffled frog leaping algorithm with genetic algorithm Crossover for solving economic load dispatch problem with Valve-point effect”, Applied Soft Computing ,2013.

[19] Chauhan, G., Jain, A., Verma, N,: “Solving economic dispatch problem using mi power by lambda iteration method”, In: Proc. 4th IEEE Int. Conf. Recent Adv. Inf. Managemen,. ICISIM 2017, January, pp. 95–99, 2017.

[20] Zhu Y, Quio B, Dong Y, Qu B, Wu D, ”Multiobjective Dynamic Economic Emission Dispatch Using Evolutionary Algorithm Based on Decomposition” IEEJ Trans,2019.

[21] Vennila H and Rajesh R,” A solution of combined static and dynamic dispatch problems using HBSA algorithm with valve point effects”, ICSSIT 2018.

[22] Dey B, Basak S and Bhattacharya B,” A comparative analysis between price-penalty factor method and Fractional programming method for combined economic emission dispatch problem using novel hybrid CSA-JAYA algorithm, IET Smart Grid, pp. 367-380,2021.

World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Review Article)

Check for updates

Effect of air-fuel ratio and pressure ratio on the exergetic performance of combined cycle gas turbine plant components

Sandeep Kumar [1, 2, *] and Ashutosh Mishra [3]

[1] *Department of Mechanical Engineering, Delhi Technological University, India*
[2] *Department of Mechanical Engineering, KCC Institute of Technology and Management, Greater Noida, India.*
[3] *Department of Mechanical Engineering, Delhi Technological University, India*

World Journal of Advanced Engineering Technology and Sciences, 2023, 08(01), 208–215

Publication history: Received on 18 December 2022; revised on 01 February 2023; accepted on 03 February 2023

Article DOI: https://doi.org/10.30574/wjaets.2023.8.1.0030

Abstract

The performance of a combined cycle gas turbine power plant depends on multiple operating parameters, which are influential in distinct ways. These parameters are classified as gas turbine parameters and steam turbine parameters. This paper deals with the influence of two gas turbine operating parameters, i.e., pressure ratio and air-fuel ratio. The 2nd law of thermodynamics has been taken into consideration along with energy analysis to measure the performance of several components of a combined cycle gas turbine plant. Therefore, a comprehensive exergy analysis for each element and the plant has been carried out to overserve the trend of the given range of aforementioned parameters. It is observed from the results that exergy destruction increases in the compressor at a higher-pressure ratio if the compressor is subjected to increased airflow, but exergetic efficiency remains unchanged. Moreover, a similar increment has been observed in the combustion chamber, but the rate of change varies along with the increase in the Air-fuel ratio. For the lower pressure ratios (5-10), the exergy destruction rate for the steam turbine decreases along with increasing in air-fuel proportion, but the effect becomes nearly opposite for the higher-pressure ratios.

Keywords: Combined cycle gas turbine; Exergy analysis; Exergetic efficiency; Compressor; Combustion chamber; Steam turbine

1. Introduction

The energy demand is increasing day by day at a significant rate. This demand is majorly fulfilled by thermal-based power-generating units [1]. The combined cycle gas turbine (CCGT) plants are one of the advanced powers solving multiple available power plant technologies [2]. As a CCGT plant possesses an upgraded efficiency than a standalone coal-based steam power plant, it also shows lesser emission from plant exhaust [3].

The performance of these plants relies on the complex combination of several parameters. These parameters can be described as environmental parameters and user-based parameters. The user-based parameters are further classified as gas and steam turbine parameters. The gas turbine parameters are the ones that influence the performance of CCGT significantly [4]. The major influencing gas turbine parameters are compressor pressure ratio, air-fuel ratio, turbine inlet temperature, and isentropic efficiencies of compressor and gas turbine [5, 6]. The performance speculation of any energy system is carried out by thermodynamics means.

The most popular and simple technique available is 1st law analysis. Though it is easy to implicate the 1st law concept in the aforesaid system, the study does not show the complete performance information. Here comes the role of the 2nd law of thermodynamics as the manifestation of exergy analysis. The exergy analysis takes the various irreversibilities

[*] Corresponding author: Sandeep Kumar; E-mail: er.sandeepgoyal1988@gmail.com

into account, which are getting generated at different locations of the energy system. Moreover, this technique helps in finding those locations and providing the scope of improvement in a particular component. Hence identifying the individual components with high potential for improvement provides an efficient system [7]. An exergetically improved system is said to have lesser energy wastage with better utilization of fuel. The common terms studied under the exergy analysis are known as exergy destruction rate and exergetic efficiency. Few researchers have studied the effect of the aforementioned gas turbine parameters on this exergy-based performance factor.

Ankit et al. [6] studied the effect of compressor pressure ratio and isentropic efficiency on the thermal efficiency of the gas turbine cycle. It was observed that increasing the pressure ratio and isentropic efficiency increases the thermal efficiency as per a higher value of the air-fuel ratio. Also, increasing the air-fuel ratio and keeping the compressor inlet temperature constant resulted in decreased thermal efficiency. Horlock et al. [8] also presented the influence of these operating parameters on the energy performance of natural gas-based gas turbine cycles. Abdollahian et al. [9] studied the effect of supplementary firing on the energy performance factors of a combined cycle power plant. Implementing the supplementary firing caused an increase of 26.3 MW and 2.43% in power generation and cycle efficiency, respectively.

The literature work discussed above is associated with an energy analysis of gas turbine cycles and combined cycle gas turbine systems. The emphasis on the 2nd law of thermodynamics has not been observed. The present study considers the same operating parameters, but the performance of CCGT will be observed through an exergetic perspective. With the help of available literature work regarding the CCGT, the two parameters considered for the analysis are pressure ratio and air-fuel ratio.

1.1. System Description

The combined cycle gas turbine plant consists of a topping cycle (i.e., Gas turbine working on the principle of Brayton Cycle), Heat recovery steam generator (HRSG), and Bottoming Cycle (i.e., Steam Turbine working on the principle of Rankine Cycle). Figure 1 exhibits the schematic arrangement of a combined cycle gas turbine plant. The Exhaust of the gas turbine is utilized as a heat source for the steam cycle. Therefore, this arrangement is more efficient and environmentally friendly as a power-generating facility.

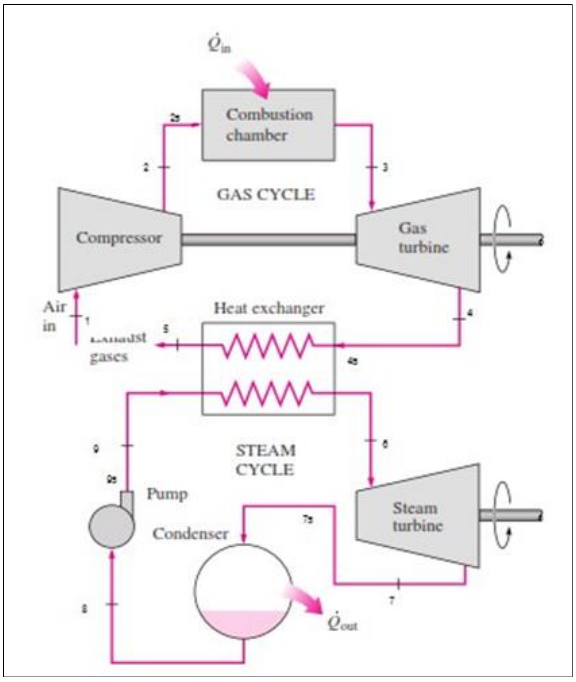


Figure 1 Schematic of combined cycle gas turbine plant

Steam is generated with the help of HRSG as it comprises three heat exchanger packages (Economizer, evaporator, and superheater). Figure 2 displays the heat transfer between exhaust gas and the water/steam line.

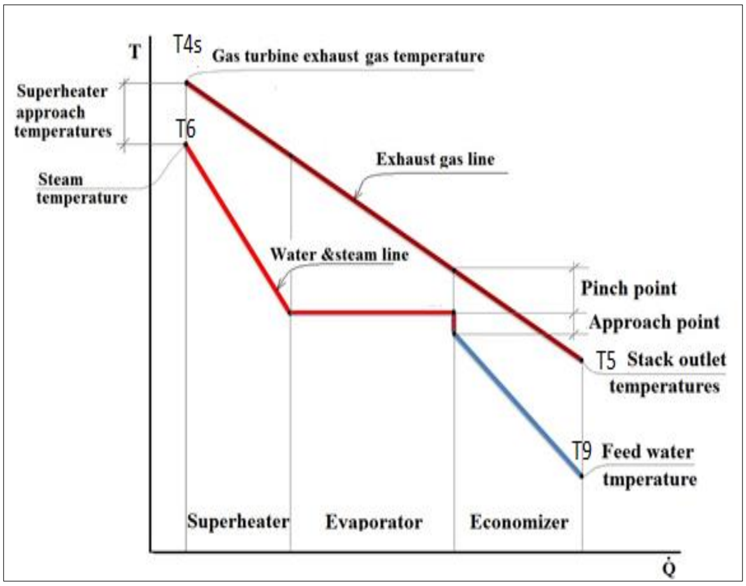


Figure 2 T-Q diagram of HRSG

2. Methodology

The equations for energy analysis of combined cycle gas turbine plants have been taken from Ahmadi et al. [10]. The basic equation employed in the exergy analysis performed on the selected combined cycle power plant is presented in this section. As with the energy analysis, exergy balances for individual components are written, and exergy flows and irreversibilities for each component are found. Then, overall exergy efficiency and exergy destruction are found for the whole system.

The equation for exergy destruction for an energy system can be written as[11]

$$\dot{Ex}_Q + \sum_i \dot{m}_i e_i = \sum_e \dot{m}_e e_e + \dot{Ex}_W + \dot{Ex}_D \qquad 1$$

The combustion chamber is the only component where both physical and chemical forms of exergies are considered. The equation for the chemical exergy is given below:

$$X_{ch} = \dot{m}_f e_{ch} \text{ ....2}$$

where,

$$e_{ch} = \dot{x}_i e_{chi} + RT_0 \sum x_i ln x_i + G_e \text{............ 3}$$

Where Ge is Gibbs free energy which is a negligible quantity in a gas mixture operated at low pressure. So, for the calculation of fuel exergy, the given expression does not hold well. Thus, the fuel exergy can be calculated as the ratio of fuel exergy to the lower heating value of the fuel.

$$\Omega = \frac{e_f}{LCV} \text{.................4}$$

ef is the specific exergy of the fuel.

For gaseous fuel with composition ***CxHy***, the value of Ω can be calculated as

$$\Omega = 1.033 + 0.0169 \frac{Y}{X} - \frac{0.0698}{X} \text{ .... 5}$$

For Methane (CH_4) X=1, Y=4

Then Ω=1.06

$$\boldsymbol{X_f} = \dot{m}_f(\mathbf{1.06}*\mathrm{LCV}) \text{ ....6}$$

The two exergy-based performance factors are considered for the analysis of various components of the CCGT plant. These factors are knowns as exergy destruction rate and exergetic efficiency. The equations for these components are given in Table 1. The input parameters considered for comprehensive exergy analysis are listed in Table 2.

Table 1 Equations for exergy destruction rate and exergetic efficiency

Components	Exergy Destruction Rate	Exergetic Efficiency
Compressor	$X_1 - X_{2s} + W_c$	$\frac{X_1 - X_{2S}}{W_C}$
Combustion Chamber	$X_{2s} + X_f - X_3$	$\frac{X_3}{X_{2S} + X_f}$
Gas Turbine	$X_3 - X_{4s} - W_{GT}$	$\frac{W_{GT}}{X_3 - X_{4S}}$
HRSG	$(X_{4s} + X_{9s}) - (X_5 + X_6)$	$\frac{X_{9S} - X_6}{X_{4S} - X_5}$
Steam Turbine	$X_6 - X_{7s}$	$\frac{W_{st}}{X_6 - X_{7S}}$
Condenser	$X_{in} - X_{out}$	$1\text{-}\frac{\triangle Xdest.,Cond}{\text{Xin}}$
Pump	$X_8 - X_{9s} + W_P$	$\frac{X_8 - X_{9S}}{W_P}$

Table 2 Input parameters consider for analysis

Parameter	Value (Unit)	Parameter	Value (Unit)
Ambient Temp.	298 K	Pressure Ratio	5-30
Ambient Pressure	1.01325 bar	LCV of Fuel	43500 kJ/kg
γ_a	1.4	$\eta_{is,GT}$, $\eta_{is,ST}$	90 %
γ_g	1.33	$\eta_{is,C}$	88%
C $_{pa}$	1.002 kJ/kg.K	Dryness fraction	0.88
C $_{pg}$	1.115 kJ/kg.K	Condenser pressure	0.07 bar
Air-Fuel Ratio	50-130	Pinch point temp. difference	13 ^{0}C

3. Results and Discussion

A MATLAB code has been generated to calculate the factors showing the pattern of varying the multiple operating parameters. The main parameters selected for the study are the air-fuel ratio and pressure ratio. The performance factors for exergetic evaluation are exergy destruction and exergetic efficiency.

Fig. 3 portrays the exergy destruction of the air compressor as a function of the air-fuel ratio at various pressure ratios. The pressure ratio varied from 5 to 30 in a step of 5, while the air-fuel ratio varied from 50 to 130 in a step of 10; with the increase in the air-fuel ratio, the exergy destruction rate of the air compressor increased. Here, the mass of fuel remains constant at 1 kg, and the mass of air increases, so the air-fuel ratio. To compress more air, the compressor has to work more, and it results in an increased exergy destruction rate. At a particular ratio, as the pressure ratio increases exergy destruction rate increases too. This is because more work is required by the compressor and work done required by the compressor is directly proportional to the pressure ratio.

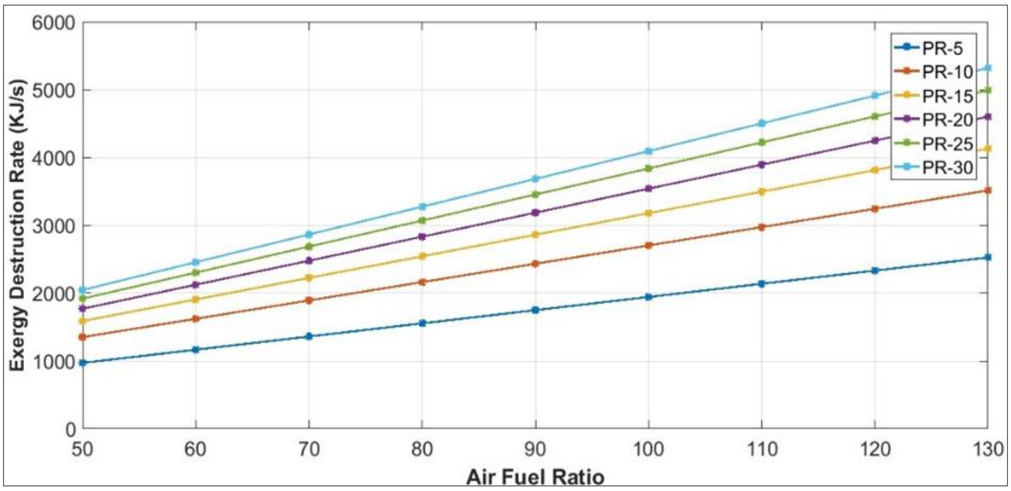


Figure 3 Effect of air-fuel ratio and pressure ratio on the exergy destruction rate of the compressor

Fig. 4 demonstrates the variation of the Exergetic Efficiency of Air Compressor as a function of the air-fuel ratio at various pressure ratios. The air-fuel ratio varied from 50 to 130 in a step of 10. The pressure ratio varied from 5 to 30 in a step of 5. At a particular air-fuel ratio, the exergetic efficiency of the air compressor continuously decreases with an increase in pressure ratio. This is due to the reason that increasing the pressure ratio increases the compressor work. As the air-fuel ratio increases, exergetic efficiency remains constant. This is because exergetic efficiency is not a function of the air-fuel ratio.

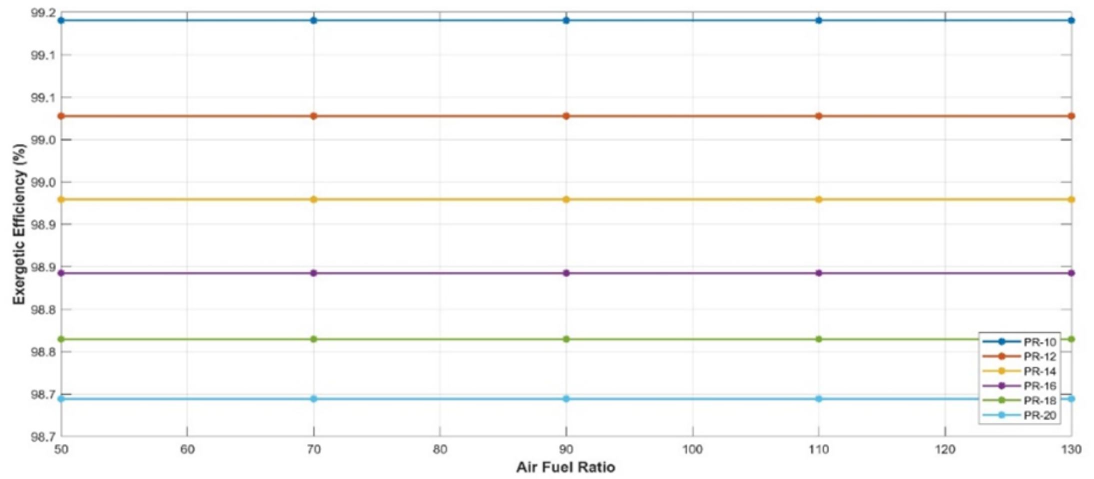


Figure 4 Effect of air-fuel ratio and pressure ratio on the exergetic efficiency of the compressor

Fig 5 demonstrates the variation of the Exergy destruction rate of the combustion chamber as a function of the air-fuel ratio at various pressure ratios. The air-fuel ratio varied from 50 to 130 in a step of 20. The pressure ratio varied from 5 to 30 in a step of 5.

With the increase in the air-fuel ratio, the exergy destruction rate increases too. This is due to the increased amount of heat addition in the combustion chamber, and it results in an increment of exergy destruction rate.

At a particular air-fuel ratio, as the pressure ratio increases, the exergy destruction rate decreases. This happens because, due to the increased pressure ratio, the combustion chamber receives the air with high temperature, so it requires less chemical energy addition.

At a particular air-fuel ratio, with an increasing pressure ratio, the marginal exergy destruction rate decreases. This is due to the reason; the available exergy after the compressor is less for higher pressure ratios owing to the higher destruction rate in the compressor.

At lower air-fuel ratios, the marginal increment in the exergy destruction rate is more rapid as compared to that at higher air-fuel ratios. This is due to the reason at a lower air-fuel ratio, the available exergy is higher since the average temperature of available heat is higher, which goes on decreasing with an increase in the air-fuel ratio.

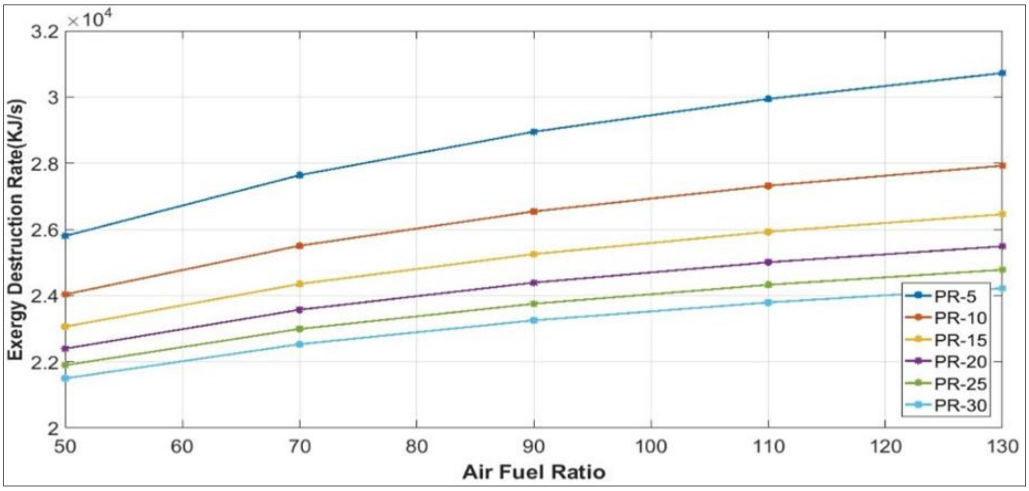


Figure 4 Effect of air-fuel ratio and pressure ratio on the exergy destruction rate of the combustion chamber

Figure 6 displays the Exergy Destruction Rate of the steam turbine at Various Pressure Ratios versus Air Fuel Ratio. The temperature increased by the compressor at a low-pressure ratio is not dominant as compared to the temperature decreased due to the addition of the air-fuel ratio. That's why a specific decreasing trend is visible at the low-pressure ratio. On increasing the pressure ratio, the temperature increase is very much high as compared to the temperature decrease by increasing the air-fuel ratio. That's why the exergy destruction rate continuously increases on increasing air-fuel ratio and at higher pressure ratio.

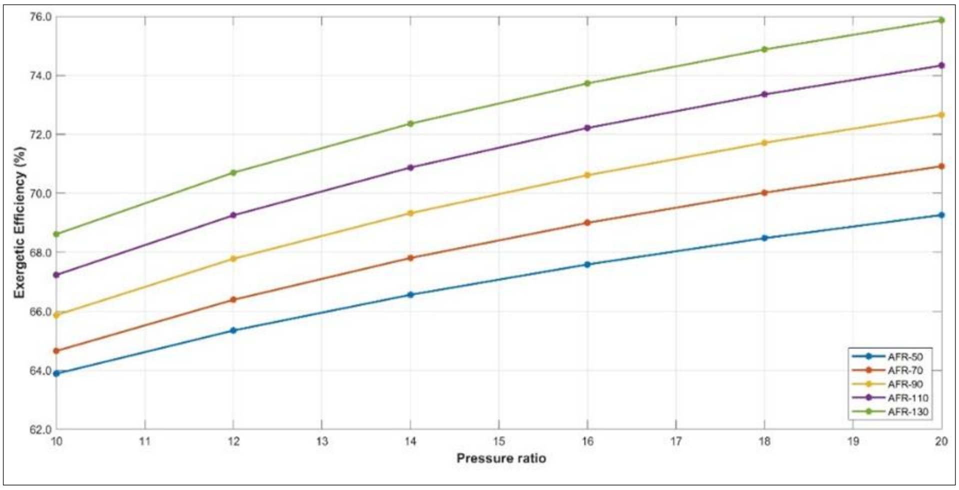


Figure 6 Effect of air-fuel ratio and pressure ratio on the exergetic efficiency of the combustion chamber

Figure 7 displays the Exergy Destruction Rate of the steam turbine at Various Pressure Ratios versus Air Fuel Ratio. The temperature increased by the compressor at a low-pressure ratio is not dominant as compared to the temperature decreased due to the addition of the air-fuel ratio. That's why a specific decreasing trend is visible at the low-pressure ratio. On increasing the pressure ratio, the temperature increase is very much high as compared to the temperature decrease by increasing the air-fuel ratio. That's why the exergy destruction rate continuously increases on increasing air-fuel ratio and at higher pressure ratio.

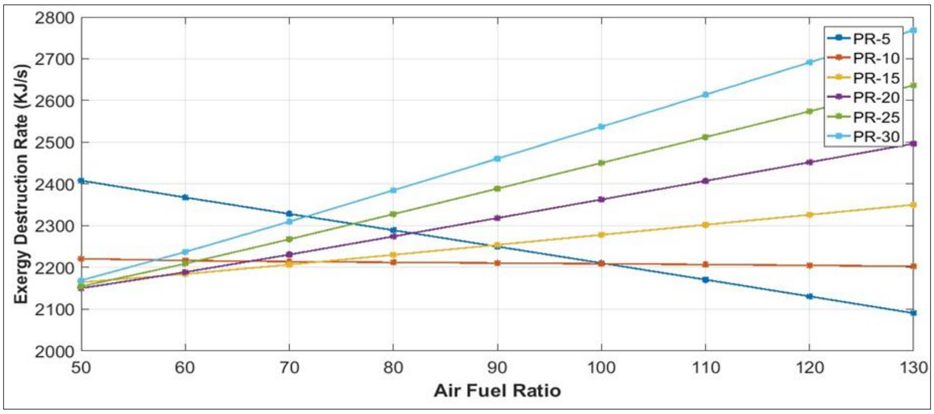


Figure 7 Effect of air-fuel ratio and pressure ratio on the exergy destruction rate of steam turbine

Nomenclature

Symbol	**Description**	**Unit**
X, E	Exergy	KJ
γ	Heat capacity ratio	-
ψ	Specific exergy	kJ/kg
η	Efficiency	-
H, h	Specific Enthalpy	kJ/kg
m	Mass flow rate	Kg/s
C_p	Specific Heat Capacity	kJ/Kg K
T	Temperature	K
Q	Heat supplied	kJ
W	Work	kW
P	Pressure	bar
s	Specific Entropy	kJ/kg.K
Subscript	**Description**	
i, in	Inlet	
e, out	Outlet	
D	Destruction	
f	Fuel	
a	Air	
g	Gas	
is	Isentropic	
0	Ambient condition	
c	compressor	
Abbreviation	**Description**	
LCV	Lower Calorific Value	
GT	Gas Turbine	
HRSG	Heat Recovery Steam Generator	
ST	Steam Turbine	
PR	Pressure ratio	

4. Conclusion

An exergy-based comprehensive analysis is carried out on a CCGT plant. The main parameters selected for the analysis are pressure ratio and air-fuel ratio. The performance has been evaluated in terms of the exergy destruction rate and exergetic efficiency of selected components. The following key points have been concluded from the above analysis are detailed.

- The rate of increment in exergy destruction in the compressor is more at a higher-pressure ratio when associated with an increasing Air-fuel ratio. Moreover, the exergetic efficiency of the compressor remains constant over the range of the air-fuel ratio implying as it is not the function of the air-fuel ratio.
- In the combustion chamber, the marginal increment in exergy destruction rate increases at a greater rate as the pressure ratio is increased. The exergetic efficiency of the combustion chamber is observed to increase with the increase in pressure ratio.
- With the increase in the air-fuel ratio, the exergy destruction rate of the steam turbine decreases for a lower pressure ratio (pressure ratio 5-10). For the pressure ratio of 15 and more, it starts increasing as the pressure ratio increases.

Compliance with ethical standards

Acknowledgments

The authors would like to thank the Department of Mechanical Engineering, KCC Institute of Technology & Management, for providing the resources.

Disclosure of conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] IEA (2019), G.E.C.S.R., IEA, Paris Available from: https://www.iea.org/reports/global-energy-co2-status-report-2019.

[2] Horlock, J.H., Combined Power Plants—Past, Present, and Future. Journal of Engineering for Gas Turbines and Power, 1995. 117(4): p. 608-616.

[3] Liu, Z. and I.A. Karimi, Simulation of a combined cycle gas turbine power plant in Aspen HYSYS. Energy Procedia, 2019. 158: p. 3620-3625.

[4] Hasan, N., J.N. Rai, and B.B. Arora, Optimization of CCGT power plant and performance analysis using MATLAB/Simulink with actual operational data. Springer Plus, 2014. 3(1): p. 275.

[5] Ashutosh Mishra, S.A., EXERGETIC ANALYSIS OF A SINGLE PRESSURE HEAT RECOVERY STEAM GENERATOR, in TAME-2019,. 2019, International Journal of Engineering Sciences Paradigms and Researches (IJESPR). p. 63-69.

[6] Kumar, A., et al., Thermodynamic analysis of gas turbine power plant. International Journal of Innovative Research in Engineering & Management (IJIREM), 2017. 4(3): p. 648-654.

[7] Kotas, T.J., The exergy method of thermal plant analysis. 2012: Paragon Publishing.

[8] Horlock, J.H., Advanced gas turbine cycles: a brief review of power generation thermodynamics. 2013.

[9] Abdollahian, A. and M. Ameri, Effect of supplementary firing on the performance of a combined cycle power plant. Applied Thermal Engineering, 2021. 193: p. 117049.

[10] Ahmadi, P., I. Dincer, and M.A. Rosen, Exergy, Exergoeconomic and environmental analyses and evolutionary algorithm based multi-objective optimization of combined cycle power plants. Energy, 2011. 36(10): p. 5886-5898.

[11] Bejan, A., Fundamentals of exergy analysis, entropy generation minimization, and the generation of flow architecture. 2002. 26(7): p. 0-43.

Flow Measurement and Instrumentation 91 (2023) 102342

Contents lists available at ScienceDirect

Flow Measurement and Instrumentation

journal homepage: www.elsevier.com/locate/flowmeasinst

Effect of the inlet-to-outlet key width ratio of Piano Key Weir on its hydraulic behaviour

Deepak Singh, Munendra Kumar *

Dept. of Civil Engineering, Delhi Technological University, Delhi, India

ARTICLE INFO

Keywords:
Piano key Weir
Relative width ratio
Low head efficient structure
Relative energy dissipation

ABSTRACT

The Piano Key Weir (PKW) is an ungated type of spillway, i.e., a novel evolution over the traditional labyrinth weir. It allows the reservoirs to operate with elevated supply levels without causing any damage to the dam structures, thereby providing additional storage. It is designed to improve the hydraulic performance of linear weirs by increasing pass discharge and energy dissipation. In this study, an experimental investigation has been carried out to assess the effect of the inlet-to-outlet width ratio (W_i/W_o) on PKW hydraulic behaviours viz hydraulic efficiency and energy dissipation. To this end, nine different width proportions ($1 \leq W_i/W_o \leq 2$) type-A PKW models were tested and examined. The findings revealed that the W_i/W_o ratio significantly impacts the hydraulic performance of PKW, and the results indicate that the efficiency of the PKW increases as the width ratio increases at a certain limit and then starts decreasing. The discharge coefficient was the highest for the given discharge and head, resulting in the best hydraulic performance with a W_i/W_o ratio between 1.25 and 1.30. However, the energy dissipation across the PKW decreases as the width ratio increases. Moreover, the discharge coefficient of different width ratios (W_i/W_o) ranging between 1.28 and 1.30 is 7–17% higher than $W_i/W_o = 1$ and 8–13% higher than $W_i/W_o = 2.0$. However, the energy dissipation across the weir for $W_i/W_o = 2.0$ indicates 15–29% less energy dissipation than $W_i/W_o = 1$. It means the energy dissipation across the weir decreases as the W_i/W_o ratio increases.

1. Introduction

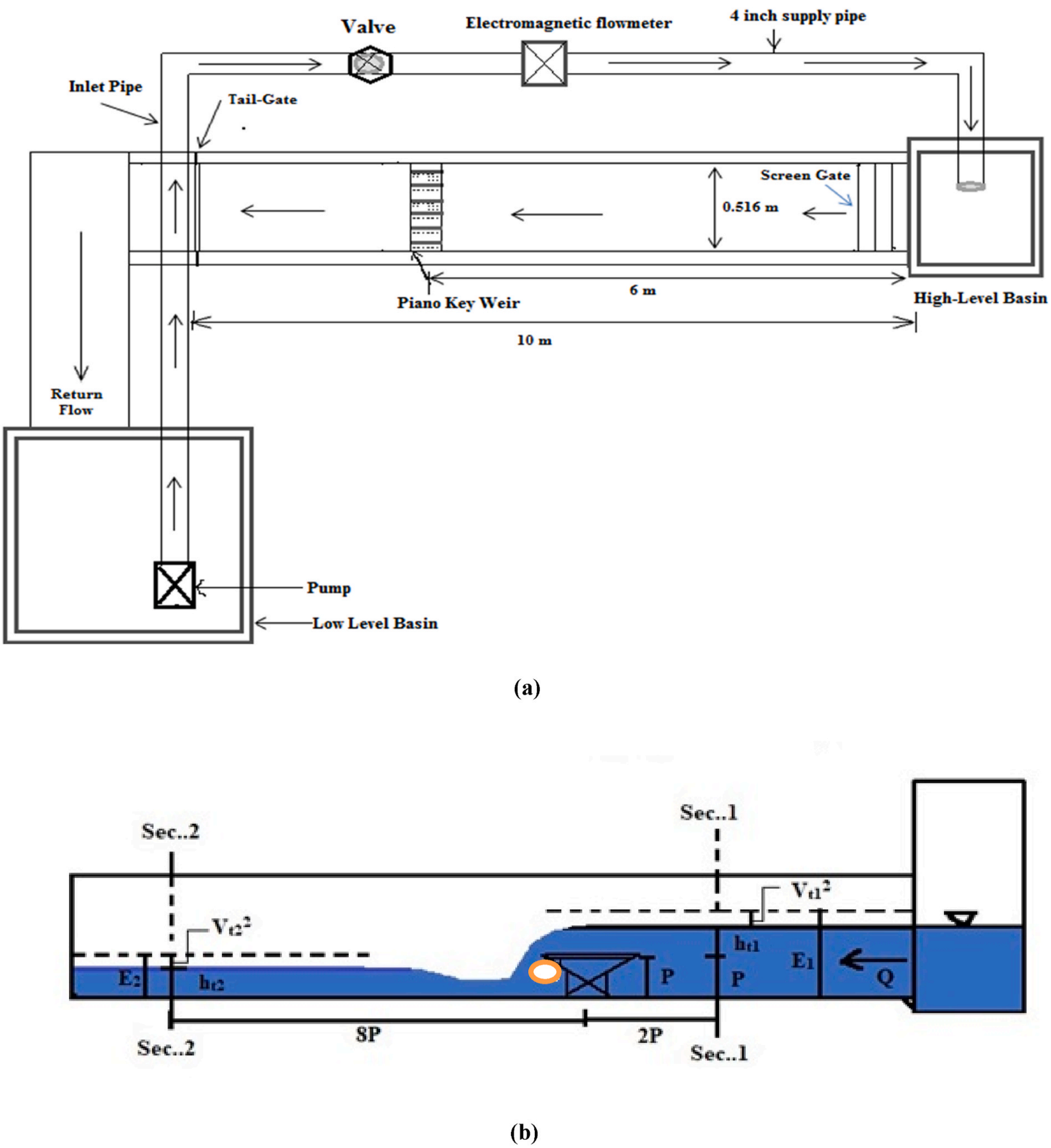
Piano Key Weir (PKW) is an ongoing advancement in spillway hydraulics. It further advances the traditional labyrinth weir, studied first by Lempérière and Ouamane [1]. They aimed to develop a new type of labyrinth weir with a smaller footprint while maintaining a structurally economical and straightforward structure that could be easily constructed. At present more than 35 PKWs have been successfully built worldwide in numerous countries, including India, Sri Lanka, France, Australia, Vietnam, Switzerland, South Africa, the UK, and Algeria [2]. The flow characteristic of the PKW significantly depends on its shape and a large number of geometric parameters that affect the system's hydraulic performance and downstream morphology [3]. Recently [4–6] summarized the PKW's geometrical and hydraulic evaluation over the last decade. Yazdi et al. [7] examined the weir geometry effects on the scour development downstream of the PKW.

Generally, the efficiency of a weir or any hydraulic structure depends on the optimal shape of the geometry. According to Lempérière and Jun [8], the optimal proportion for the inlet-outlet key (W_i/W_o) is near 1.2. Hien et al. [9] studied the ratio $W_i/W_o = 1.5$ and suggested that $W_i/W_o = 1.2$ is likely more effective. However, no information was presented to approve that guarantee. Ouamene and Lempérière [10] studied three different width ratios ($W_i/W_o = 0.67$, 1.0, and 1.5) models of PKWs and found that increasing the ratio W_i/W_o, the efficiency of the weir also increases. However, they did not give a descriptive explanation but gave only a little explanation as to why this occurs. They claimed that $W_i/W_o = 1.2$ increases the efficiency of PKW up to 5% compared to $W_i/W_o = 1$; however, the data for $W_i/W_o = 1.2$ were not presented as part of that study. Later, the width ratio (W_i/W_o) was near-optimal equal to 1.25 by Lempérière [11]. All recent past studies agreed that $W_i/W_o > 1.0$ produces greater discharge efficiency than $W_i/W_o < 1.0$. Anderson [12] and Anderson and Tullis [13] found that the maximum discharge efficiency of PKW, the W_i/W_o ranged between 1.25 and 1.5. Recently, Mero et al. [14] stated that the W_i/W_o ratio equal to 1.25 shows the best hydraulic efficiency.

Nowadays the influence of climate change is now being felt all over the World, and it is exacerbating flood events [15]. It implies that the predicted design of flood discharges in many urban areas will be

* Corresponding author. Dept. of Civil Engineering, Delhi Technological University, Delhi, India.:
E-mail addresses: deepak_phd2k18@dtu.ac.in (D. Singh), munendrakumar@dtu.ac.in (M. Kumar).

https://doi.org/10.1016/j.flowmeasinst.2023.102342
Received 16 September 2021; Received in revised form 4 January 2023; Accepted 27 February 2023
Available online 9 March 2023



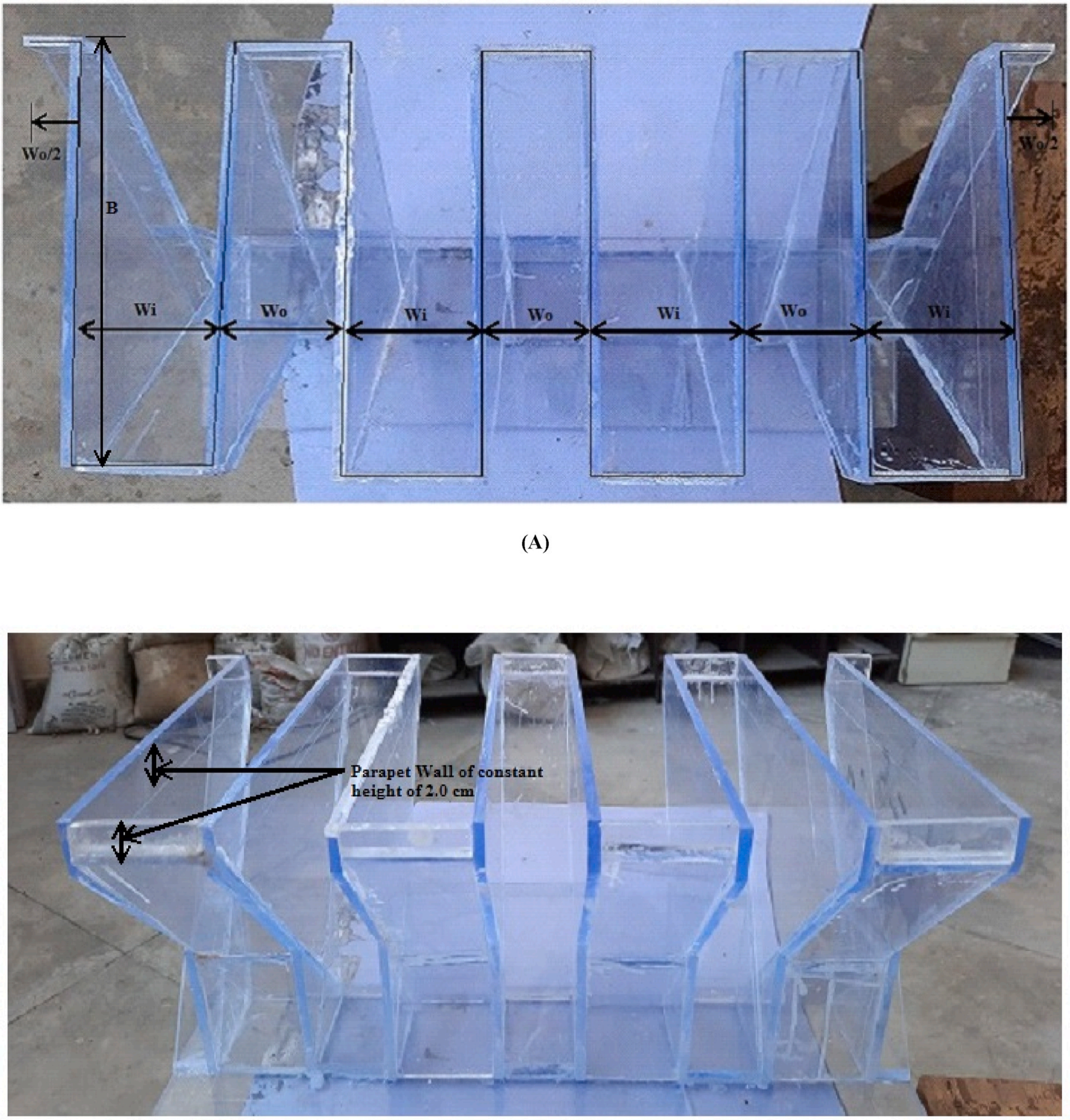
(a)

(b)

Fig. 1. Schematic experimental setup (a) Plan View, (b) specific energy measurement.

significantly increased [16]. According to Abhash & Pandey [5], in prosecutions of dam failures and their causes around the World, insufficient spillway discharge capacity causes approximately 23% of dam failures. Consequently, many existing dams require urgent rehabilitation to improve their safety by enhancing their discharge capacity and, as a result, protecting people in vulnerable areas downstream of these structures [17,18]. Furthermore, the scouring and proper energy estimation downstream of any hydraulic system are extremely difficult to evaluate accurately. So it is important to assess the energy loss downstream of the hydraulic structures. Eslinger and Crookston [19] conducted an experimental study to evaluate the energy dissipation analysis at the base of type-A PKWs. In addition, Silvestri et al. [20] described the vital work of Erpicum et al. [21] by measuring four-stepped spillway lengths. They concluded that shorter stepped spillway lengths downstream of PKWs achieve uniform flow conditions but not those downstream of ogee-crested weirs. Local scour at the toe of PKWs placed in canals and rivers are also related to PKW energy dissipation. Moreover, Singh and Kumar [22,23] presented an experimental investigation and a computational technique based on gene expression programming (GEP) to estimate the residual energy at the base of type-B PKW, respectively.

This paper aims to examine the effects of the W_i/W_o ratio on PKW hydraulic performance and downstream energy dissipation. Although few experimental investigations have been carried out in the past to study the impact of the W_i/W_o ratio, there is insufficient information about the effect of this parameter on the discharge capacity, coefficient, and energy dissipation of the PKWs. Therefore, this experimental investigation was conducted to enhance the understanding of the impact of the W_i/W_o ratio on the hydraulic efficiency and downstream energy dissipation of the type-A PKW.



(A)

(B)

Fig. 2. Fabricated model Geometry of PKW (A) Plan view (B) PKW with constant Parapet Wall.

2. Experimental details

All tests were conducted in a laboratory-scale steel flume (10 m, long X 0.516 m, wide X 0.6 m, deep) in the FM&HE Laboratory at Delhi Technological University, Delhi, India. Water is pumped by a 20 HP pump connected via a pipe network that includes calibrated orifice meter (±0.25% uncertainty), a flow regulating valve to control the discharge, and a 4–20 mA electromagnetic flowmeter (uncertainty ± 0.2%). (see Fig. 1). The flume headbox featured a metal screen gate to improve upstream approach flow uniformity. A 4–20 mA ultrasonic level sensor is attached to the flume (±0.2% uncertainty), and a pointer gauge with the least count of 0.1 mm for the head over the weir crest. The readings were taken (2P upstream and 8P downstream as shown in Fig. 1(b)) after the water surface had reached a steady state for at least 2–3 min. The mean approach flow velocity V_t was calculated as the average velocity measured at the same cross-section for 1-min records at 0.25, 0.5, and 0.75 W across the width W of the flume using a Sontek ADV (Acoustic Doppler Velocimeter). The velocity analysis results revealed agreement between these average cross-sectional velocities and mean approach velocities (by ADV), resulting in a H_t difference of less than 5% for the Q ranges. The geometry of the laboratory-scaled type-A models is shown in Fig. 2, and the data collected in the present study are shown in Table: 1.

3. Result and discussion

The flowing discharge over PKWs is the sum of the flow over the downstream crest, the upstream crest, and the sidelong flow over the side crest (see Fig. 3) [24]. So it creates a complicated three-dimensional flow over the PKW, with splash and spray regions within the outlet keys

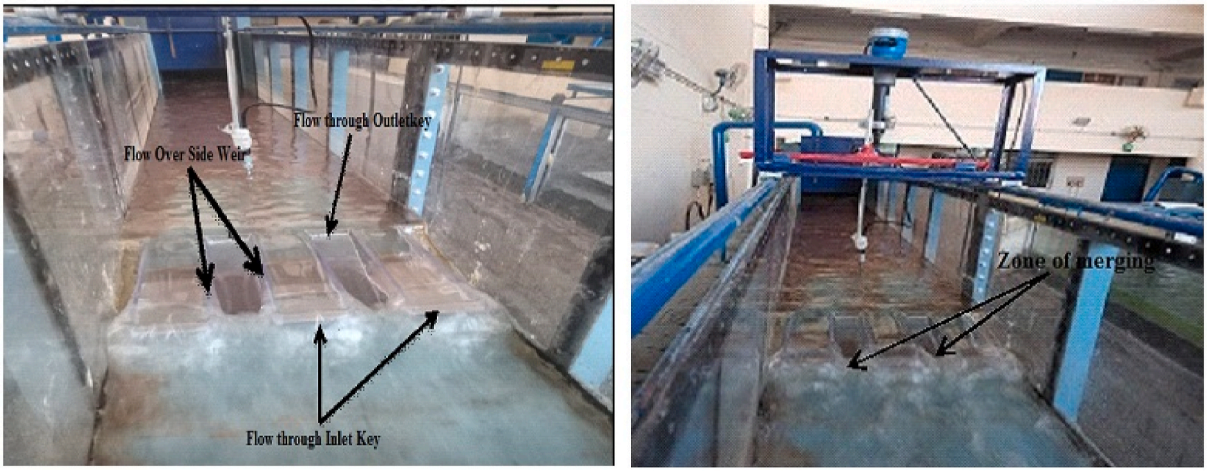


Fig. 3. Flow pattern over PKW.

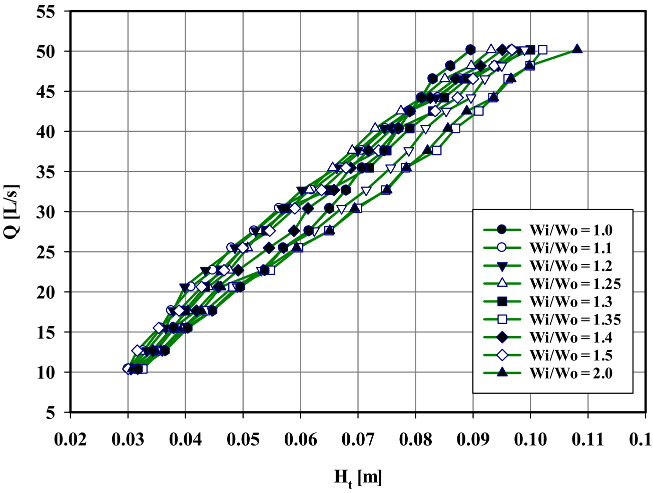


Fig. 4. Variation of discharge Q [L/s] with [H_t].

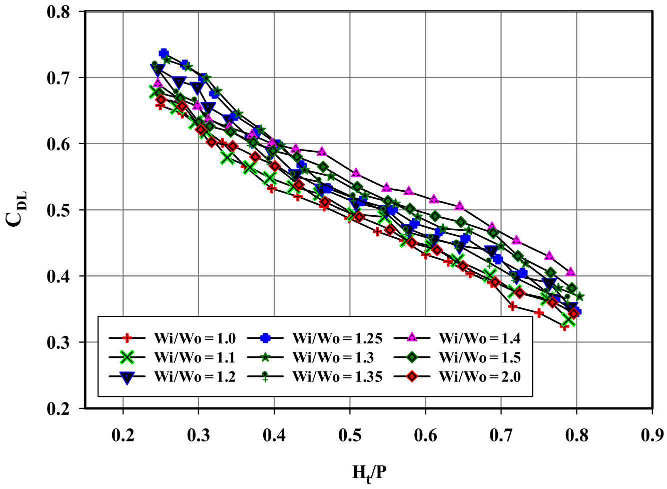


Fig. 5. Variation of discharge coefficient of discharge [C_{DL}] with [H_t/P].

and at the structure's base [25]. The flow over the PKW is highly aerated, and the area of spray and sprinkling only increased marginally proportional to the trajectory of H_t and the planar jet that started downstream on the crest(see Figure: 3). But the aeration region increased significantly with H_t, partly because the local speed increased, resulting in greater advection levels and turbulent mixing [26]. The lateral flow momentum transfer forms a free shear boundary at the inlet key's edge. This boundary layer eventually reconnects with the sidewall some distance downstream, the location of which is determined by the bubble's mass equilibrium and upstream turbulence. The zone behind the free shear boundary is known as a separation bubble and consists of a volume of low-velocity, recirculating flow [27]. Different nature bubbles form in the downstream or outlet key section when water flows. The nappe profile forms a conical cavity of air between itself and the sidewall as the flow discharges over the sidewall crest. The apex of this bubble is highly unstable. It reflects a balance between the longitudinal flow's momentum in the outlet key's upper portions and the transverse nappe's free-fall trajectory [28]. The outcomes of the present study have been analyzed in two ways. The first was to determine the influence of the different W_i/W_o proportions on the discharge efficiency of the PKW. The second aspect is comprehending the effect of different W_i/W_o proportions on the energy dissipation of the PKW. The detailed discussion is described in the following subsections later on.

3.1. Effects of the different W_i/W_o proportions on the discharge efficiency

The discharge coefficient of PKW is computed based on the developed length (see Eqn. (1)) over the scope of $0.24 \leq H_t/P \leq 0.79$. The stage-discharge relationship is the horoscope of each flow measurement structure and is plotted between discharge vs. head, as shown in Fig. 4:

$$Q = Q_{PKW} = \frac{2}{3} C_{DL}\, L\, \sqrt{2g}\, H_t^{3/2} \tag{1}$$

C_{DL} is the coefficient of discharge computed along the crest length, and L is the total developed crest length. To distinguish the optimal W_i/W_o proportion range, relating the most noteworthy C_{DL} esteems or most discharge proficiency, the test outcomes have been introduced in Fig. 5, C_{DL} as a component of H_t/P. From Fig. 5, it is clear that the model W_i/W_o equal to 1.25 and 1.3 delivered the most significant release efficiency, followed by W_i/W_o = 1.35, 1.2, 1.4, 1.1, 1, 1.5, and 2. It means the optimal discharge ranges lie between 1.25 and 1.3. Moreover, the data in Figure: 5 shows that W_i/W_o = 1.25 produce a respectively higher discharge efficiency than $W_i/W_o = 1.3$ for $H_t/P \leq 0.35$, and $W_i/W_o = 1.3$ produce a respectively higher discharge efficiency than $W_i/W_o = 1.25$ at $0.35 < H_t/P \leq 0.44$, however, the model $W_i/W_o = 1.4$ have the highest discharge capacity for the range of $0.44 < H_t/P \leq 0.81$. The PKW of W_i/W_o = 1.25 and W_i/W_o = 1.3 produce (about 7–17%) higher efficiency than $W_i/W_o = 1.0$ and about 8–13% higher efficiency than $W_i/W_o = 2.0$. So from the above discussion, it is clear that the PKW is most sensitive

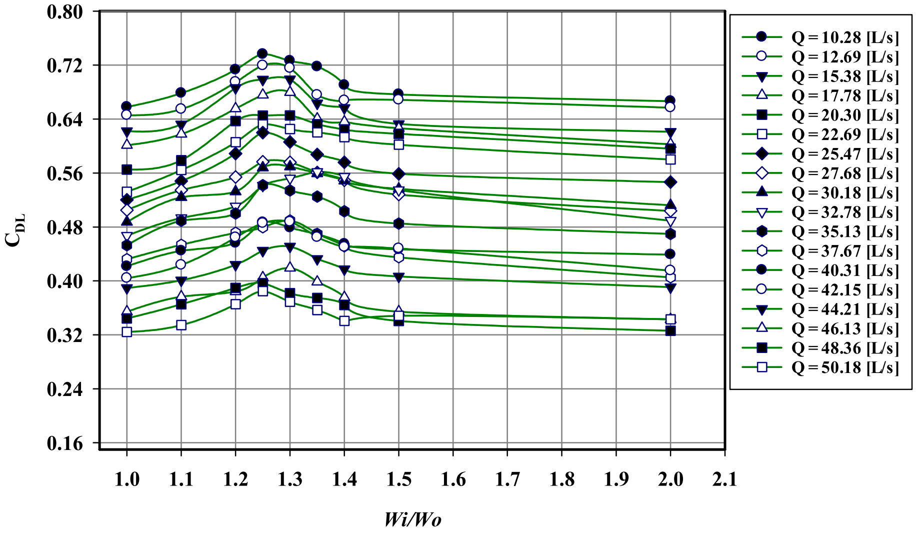


Fig. 6. Variation of discharge coefficient [C_{DL}] with the inlet to outlet width ratio [W_i/W_o].

for the low range of discharges; as the discharge increases, the efficiency of the PKW decreases rapidly or sometimes gradually.

Fig. 6 shows the maximum discharge efficiency at which the relative width ratio W_i/W_o is in the range of 1.25–1.30, and the maximum efficiency observed corresponding to $W_i/W_o = 1.275 \approx 1.28$; however, the result of most of the previous studies varies between ranges 1–1.5 [12, 13]. Some researchers suggested that the relative width (W_i/W_o) ratio ranges from 1.2 to 1.25 is close to optimal, but no data was presented to validate the claim. Therefore, the present study shows the specific width proportion (i.e., $W_i/W_o = 1.275 \approx 1.28$) at which the maximum hydraulic efficiency of the PKW was observed. In the case of the present study, the optimal width ratio range is slightly different than in previous studies, which may be due to the parapet wall. Because in most of the past studies, experiments have been performed on the flat-top PKW models. The present study conducted experiments with a parapet wall of constant height (2 cm) over each model. Recent past studies over PKW suggested that the ratio for the inlet to outlet key width should be greater than one because the relative width of the inlet key determines the unit discharge moving towards its crest, maintaining the subcritical flow as possible, and reducing the losses. On the other hand, the lesser value of the outlet key reduced the occurrence of local submergence.

3.2. Effects of the relative width ratio (W_i/W_o) on energy dissipation

The second part of the study is to examine the energy dissipation over the different W_i/W_o proportions. In order to calculate the potential of the W_i/W_o ratio on PKW energy dissipation efficiency under free flow conditions, the upstream and downstream energy across the PKW was calculated as follows:

$$E_i = P + \left(h_{ti} + \frac{V_i^2}{2g}\right) = P + H_t \tag{2}$$

where E, P, h, and V represent the specific energy, height of the weir in (m) (for d/s $P = 0$ m), head over the weir, and 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)).** E_1 and E_2 were then utilized to determine relative energy dissipation and relative residual energy ($E_r = E_2/E_1$) in the following way:

$$\frac{\Delta E}{E_1} = \frac{(E_1 - E_2)}{E_1} \times 100 = \left(1 - \frac{E_2}{E_1}\right) \times 100 \tag{3}$$

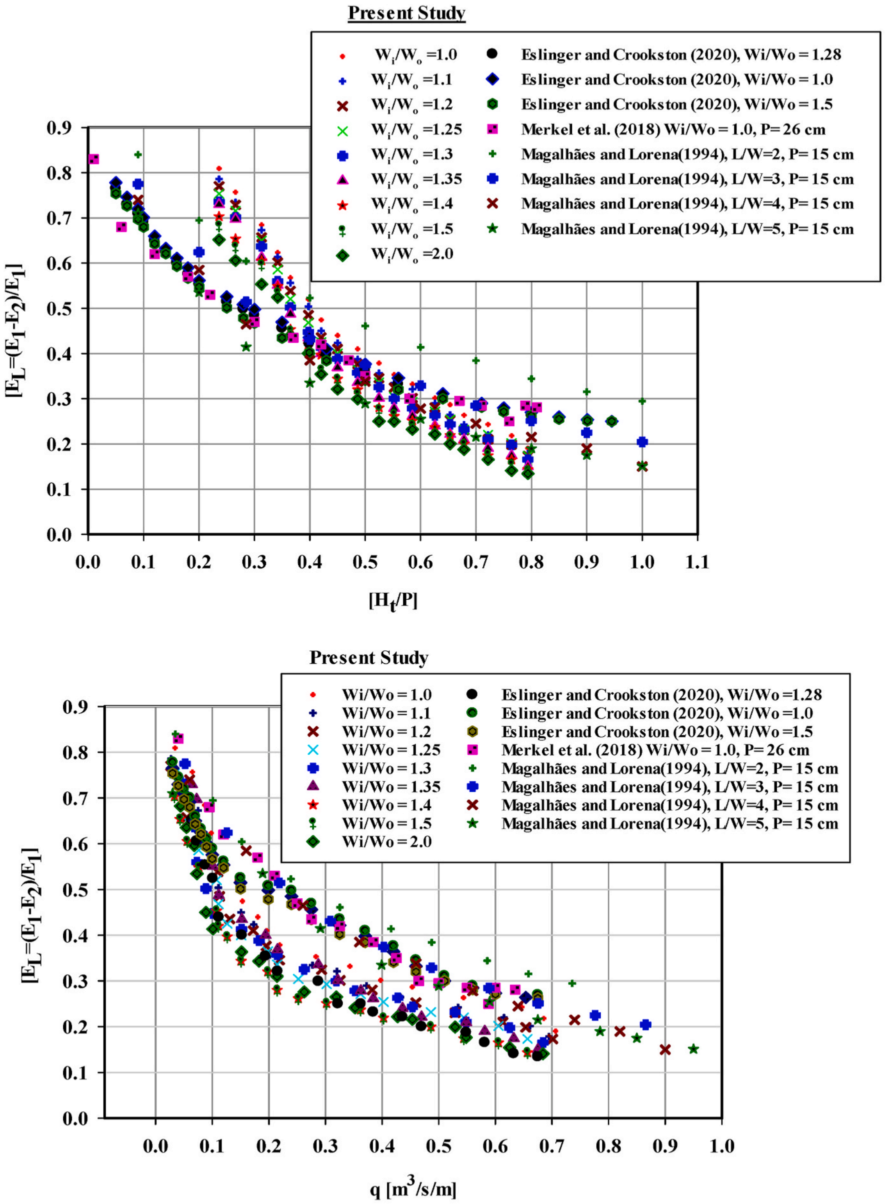
and,

$$E_r = 1 - \frac{\Delta E}{E_1} = \frac{E_2}{E_1} \tag{4}$$

where $\Delta E/E_1$ denotes total relative energy dissipation or energy dissipation ratio, E_r ($=E_2/E_1$) symbolizes the PKW's downstream residual energy. Most researchers have similarly calculated the energy dissipation over the linear and non-linear weir structures [29–33]. The ranges of various parameters and data collected in the present study are summarized in Table 1. Following testing, the rating curves for each PKW were established, analyzed, and compared with published data for laboratory-scale trapezoidal labyrinth weirs [34], rectangular labyrinth weirs [35], and PKW [19] included (see Fig. 7). All the models showed very similar trends for relative energy dissipations, while their results were compared with previous studies within the accuracy of the measurement. In the present study, the energy dissipation rate was found to be more when H_t/P is less than 0.42 (approximate for all the models)

Table: 1
Range of Data collected in the present study.

S. No.	W_i/W_o	L/W	$S_i = S_o$	$H_t(m)$	$Q(L/s)$	$B_i/P = B_o/P$	Range of $\left(\frac{E_L}{E_1}\right)$	Range of $\left(E_r = \frac{E_2}{E_1}\right)$	Number of readings
1	1.00	5	1.08	0.0300–0.0971	10.17–50.26	0.69	0.8093–0.1930	0.1907–0.8096	18
2	1.10	5	1.08	0.0304–0.0986	10.14–50.26	0.69	0.7860–0.1785	0.214–0.8214	18
3	1.20	5	1.08	0.0307–0.0989	10.19–50.07	0.69	0.770–0.1731	0.2300–0.8268	18
4	1.25	5	1.08	0.0317–0.1011	10.28–50.18	0.69	0.7533–0.1734	0.2467–0.8265	18
5	1.30	5	1.08	0.0322–0.1004	10.09–50.00	0.69	0.7356–0.1655	0.2644–0.8344	18
6	1.35	5	1.08	0.0310–0.0891	10.16–50.07	0.69	0.7297–0.1501	0.2703–0.8498	18
7	1.40	5	1.08	0.0303–0.0985	10.19–50.13	0.69	0.7031–0.1435	0.2969–0.8564	18
8	1.50	5	1.08	0.0310–0.0992	10.15–50.45	0.69	0.6822–0.1411	0.3118–0.8588	18
9	2.0	5	1.08	0.0313–0.0995	10.29–49.82	0.69	0.6518–0.1342	0.3482–0.8657	18



(i) Relative energy dissipation [E_L=(E_1-E_2)/E_1] with respect to (a) the headwater ratio [H_t/P] and (b) the unit discharge [q] for different Wi/Wo ratios.

Fig. 7. (a) Relative energy dissipation [E_L=(E_1-E_2)/E_1] with respect to (a) the headwater ratio [H_t/P] and (b) the unit discharge [q] for different Wi/Wo ratios. (b) Relative residual energy [E_2/E_1] with respect to (a) the headwater ratio [H_t/P] and (b) the unit discharge [q] for different Wi/Wo ratios.

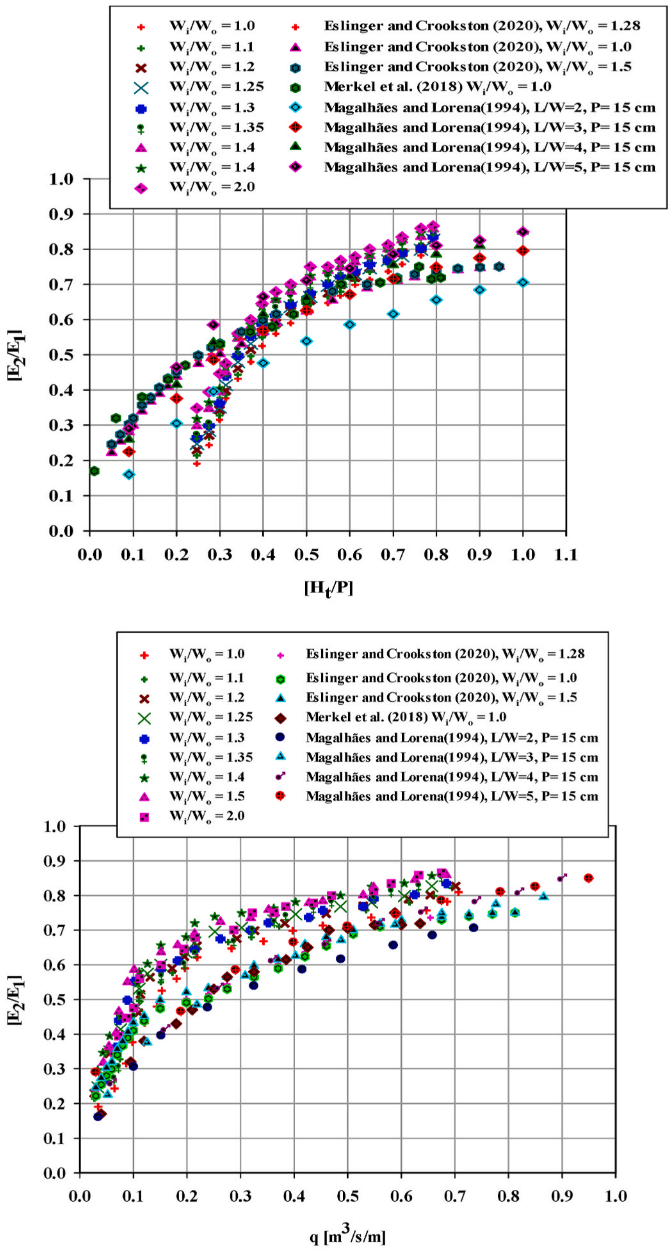
than in previous studies. The rate of relative energy dissipation was found to be less for $H_t/P > 0.55$ and between $0.42 \leq H_t/P \leq 0.55$, the rate of [$E_L$ = (E_1-E_2)/E_1] relative energy dissipation has been observed intermixing in nature.

Fig. 7 (i) & (ii) show the variation of the relative energy dissipation [E_L = (E_1-E_2)/E_1] at the PKWs toe as a function of the upstream head ratio (H_t/P) and the specific discharge (q). The relative residual energy (E_2/E_1) at the base of PKWs increases with H_t/P, particularly for smaller values, regardless of the relative width ratio (W_i/W_o). From Fig. 7 (i) & (ii), it is clear that the maximum relative energy dissipation was observed in the present study corresponding to the lowest width ratio (i.g. E_L = 0.8093 or 80.93% the corresponding W_i/W_o = 1) and the less energy dissipation for the highest width ratio (i.g. E_L = 0.6518 or 65.18% the related W_i/W_o = 2.0). Increasing the inlet key width reduces overall head losses due to the flow entering the inlet key, and increasing the inlet key flow area increases the flow carrying capacity of the inlet key. A high W_i/W_o value improves flow approach and distribution within PKWs inlet keys, whereas a high W_i/W_o value increases submergence effects in outlet keys. Submergence effects in the outlet cycles (regions where the flow depth in the outlet cycle exceeds the weir crest elevation) can reduce the discharge efficiency of the weir.

4. Conclusions

The following conclusions have been drawn from this research:

1. The impact of the inlet-to-outlet key width ratio on its discharge carrying capacity of PKW was investigated systematically, and it was seen that the efficient range of (W_i/W_o) for maximizing discharge efficiency lies between 1.25 and 1.30.



(ii) Relative residual energy [E_2/E_1] with respect to (a) the headwater ratio [H_t/P] and (b) the unit discharge [q] for different Wi/Wo ratios.

Fig. 7. (*continued*).

2. The maximum efficiency was observed corresponding to the width ratio (W_i/W_o) 1.2755 = 1.28; at this width (W_i/W_o) proportion, the PKW has (7–17%) higher efficiency than the $W_i/W_o = 1$; and 8–13% higher than $W_i/W_o = 2.0$. As like, the width of the inlet key section is expended due to the expansion of the inlet stream territory is also expanded; consequently, the energy loss of the water gets in the inlet section is reduced, and the results of the enhancement in the discharge conveying efficiency of PKW; but in the outcome of growing the inlet key width, the outlet key width diminishes (Total width of PKW or Channel width W, is constant). This starts bringing about an increment in nearby submergence of the outlet key (especially at the outlet key pinnacles) and a decline in the outlet key release conveying limit.
3. The energy dissipation across the weir decreases as the W_i/W_o ratio increases, and the maximum relative energy dissipation was observed corresponding to the lowest width ratio (i.g. $E_L = 0.8093$ or 80.93%, the corresponding $W_i/W_o = 1$) and the less energy dissipation for the highest width ratio (i.g. $E_L = 0.6518$ or 65.18% the related $W_i/W_o = 1.5$). It means the energy dissipation across the weir for $W_i/W_o = 2.0$ indicates 15–29% less energy dissipation than $W_i/W_o = 1$.

This study provides appropriate information and guidance to the designer/Engineer for designing an efficient geometry of the PKW. In the present study, the authors did not consider the scale effect. Further possibilities are to perform an experimental analysis or CFD modeling by considering the scaling effects.

Funding information

The Delhi Technological University, India, funded this study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank the Civil Engineering Department's faculty, staff, and the FM & HE laboratory technical team for their invaluable assistance. The authors would like to express their heartfelt gratitude and appreciation to Mr. Kewal Singh, a retired senior mechanic, for his practical assistance during the model fabrication process.

List of Symbols

C_{DL}	Coefficient of discharge along developed crest length
E_i	Specific Energy at section i
E_L	Relative energy dissipation
E_r	Relative residual energy
g	Acceleration of gravity
H_t	Total head
h_t	Piezometric head
i	section
L	Developed crest length
P	Height of Piano Key Weir
Q_{PKW}	Discharges flow over the Piano Key Weir
R	Height of parapet wall
V_t	Average Velocity
W	Channel width/Width of PKW
W_i	Inlet Key Width
W_o	Outlet Key Width
X	Distance of measurement section from the lateral centreline of the PKW

References

[1] F. Lempérière, A. Ouamane, The Piano Keys weir: a new cost-effective solution for spillways, Int. J. Hydropower Dams 10 (5) (2003) 144–149.

[2] B.M. Crookston, S. Erpicum, B.P. Tullis, F. Laugier, Hydraulics of labyrinth and piano key weirs: 1 00 years of prototype structures, advancements, and future research needs, J. Hydraul. Eng. 145 (12) (2019), 02519004.

[3] R.M. Anderson, B.P. Tullis, Piano key weir hydraulics and labyrinth weir comparison, J. Irrigat. Drain. Eng. 139 (3) (2013) 246–253.

[4] A. Mehboudi, J. Attari, S. Hosseini, Experimental study of discharge coefficient for trapezoidal Piano Key weirs, Flow Meas. Instrum. 50 (2016) 65–72.

[5] A. Abhash, K.K. Pandey, 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 J. Hydraul. Eng. (2020) 1–11. Taylor & Francis.

[6] D. Singh, M. Kumar, Hydraulic design and analysis of Piano Key Weirs: A review, Arab. J. Sci. Eng. (2021). Springer Berlin Heidelberg.

[7] A.M. Yazdi, S. Abbas Hoseini, S. Nazari, N. Amanian, Effects of weir geometry on scour development in the downstream of Piano Key Weirs, Water Sci. Technol. Water Supply 21 (1) (2021) 289–298.

[8] F. Lempérière, G. Jun, Low-Cost increase of dams storage and flood mitigation: The Piano Keys Weir, in: Q.53 R. 2.06 International Commissions On Irrigation And Drainage Nineteenth Congress Beijing, 2005.

[9] T.C. Hien, H.T. Son, M.H.T. Khanh, Results of some piano keys weir hydraulic model tests in Vietnam, in: Proc. of the 22nd Congress of ICOLD, Barcelona, Spain, 2006.

[10] A. Ouamane, F. Lempérière, Design of a new economic shape of Weir, Proc. of the International Symposium of Dams in the Societies of the 21st Century, Barcelona, Spain (2006) 463–470.

[11] F. Lempérière, New Labyrinth Weirs Triple the Spillways Discharge, 2009. *Published Feb. 8*, http://www.hydrocoop.org.

[12] R.M. Anderson, Piano key weir head discharge relationships, in: All Graduate Thesis and Dissertations, Utah State University, 2011. *Paper 880*.

[13] R.M. Anderson, B.P. Tullis, Influence of piano key weir geometry on discharge, in: Proc. International Workshop on Labyrinth and Piano Key Weirs, 2011 (Liège, Belgium).

[14] S.K. Mero, D.A.J. Haleem, A.A. Yousif, The influence of inlet to outlet width ratio on the hydraulic performance of Piano Key Weir (PKW-type A), Water Pract. Technol. (2022).

[15] H.O. Phelps, H.M. Azamathulla, G.S. Shrivastava, Hydraulic model study of arena dam spillway works, Trinidad, West Indian J. Eng. 43 (2) (2021) 69–76.

[16] I. Qutbudin, M.S. Shiru, A. Sharafati, K. Ahmed, N. Al-Ansari, Z.M. Yaseen, S. Shamsuddin, X. Wang, Seasonal drought pattern changes Due to climate variability:case study in Afghanistan, Water 11 (2019) 1096, https://doi.org/10.3390/w11051096.

[17] A.M. Sami Al-Janabi, A.H. Ghazali, B. Yusuf, S.S. Sammen, H.A. Afan, N. Al-Ansari, Z.M. Yaseen, Optimizing height and spacing of check dam systems for better grassed channel infiltration capacity, Appl. Sci. (2020), https://doi.org/10.3390/app10113725.

[18] A. Sharafati, Z.M. Yaseen, E. Pezeshki, Strategic assessment of dam overtopping reliability using a stochastic process approach, J. Hydrol. Eng. 25 (2020), 4020029.

[19] K.R. Eslinger, B.M. Crookston, Energy dissipation of type a Piano Key Weirs, Water 12, MDPI, Basel, Switzerland, 2020, p. 1253, 5.

[20] A. Silvestri, S. Erpicum, P. Archambeau, B. Dewals, M. Pirotton, Stepped spillway downstream of a Piano Key weir: critical length of uniform flow, in: International Workshop on Hydraulic Structures, Bundesanstalt fur, Wasserbau, Karlsruhe, Germany, 2013, pp. 99–107.

[21] S. Erpicum, A. Silvestri, B. Dewals, P. Archambeau, M. Pirotton, M. Colombié, L. Faramond, Escouloubre piano key weir: prototype versus scale models, in: Labyrinth and Piano Key Weirs II—PKW 2013, CRC Press, Leiden, The Netherlands, 2013, pp. 65–72.

[22] D. Singh, M. Kumar, Energy dissipation of flow over the type-B piano key weir, Flow Meas. Instrum. 83 (November 2021) (2022 a), 102109, https://doi.org/10.1016/j.flowmeasinst.2021.102109. Elsevier Ltd.

[23] D. Singh, M. Kumar, Gene expression programming for computing energy dissipation over type-B Piano Key Weir, Renew. Energy Focus 41 (2022 b) 230–235, https://doi.org/10.1016/j.ref.2022.03.005.

[24] O. Machiels, Experimental Study of the Hydraulic Behaviour of Piano Key Weirs, PhD Thesis ULgetd-09252012- 224610, University of Liege, 2012 (Belgium).

[25] S.I. Khassaf, L.J. Aziz, Z.A. Elkatib, Hydraulic behavior of piano key weir type B under free flow conditions, Int. J. Sci. Technol. Res. 4 (8) (2015) 158–163.

[26] D. Singh, M. Kumar, Study on aeration performance of different types of piano Key weir, Water Supp. 22 (5) (2022 c) 4810–4821.

[27] C. Tenaud, B. Podvin, Y. Fraigneau, V. Daru, On wall pressure fluctuations and their coupling with vortex dynamics in a separated–reattached turbulent flow over a blunt flat plate, Int. J. Heat Fluid Flow 61 (2016) 730–748.

[28] F. Denys, DigitalCommons @ USU International Symposium on Hydraulic Transient Hydrodynamics of Piano Key Weirs, 2018.

[29] A. Parsaie, A.H. Haghiabi, The hydraulic investigation of circular crested stepped spillway, in: Flow Measurement and Instrumentation, Elsevier Ltd, 2019, 101624, 70(August.

[30] A.H. Haghiabi, M.R. Ghaleh Nou, A. Parsaie, The energy dissipation of flow over the labyrinth weirs, Alexandria Eng. J. (2021). THE AUTHORS, 0–4.

[31] A. Parsale, A.H. Haghiabi, Hydraulic investigation of finite crested stepped spillways, Water Sci. Technol. Water Supply 21 (5) (2021) 2437–2443.

[32] S.B. Sarvarinezhad, M. Bina, E. Afaridegan, A. Parsaie, F. Avazpour, The hydraulic investigation of inflatable weirs, Water Supp. 22 (4) (2022) 4639–4655.

[33] D. Singh, M. Kumar, Computation of energy dissipation across the type-A piano key weir by using gene expression programming technique, Water Supp. 22 (8) (2022) 6715–6727.

[34] A.P. Magalhães, M Perdas de Energiado do Escoamento Sobre Soleiras em Labirinto Lorena, Energy Losses in Flow over Labyrinth Weirs, SILUSBA, Lisboa, Portugal, 1994, pp. 203–211 (In Portuguese).

[35] J. Merkel, F. Belzner, M. Gebhardt, C. Thorenz, Energy dissipation downstream of labyrinth weirs, in: Proceedings of the 7th IAHR International Symposium on Hydraulic Structures, Aachen, Germany, 2018, pp. 15–18.

Deepak Singh is currently a research scholar of the Civil Engineering Department at Delhi technological University, India. He obtained his B.Tech. (Civil) and M.Tech. in Hydraulic Engineering. He has published 6 papers in refereed international and national journals, and 2 chapters in conference proceedings. His area of research is hydraulic engineering, computational hydraulics, and hydraulic structures.

M. Kumar is currently a Professor of Civil Engineering Department at Delhi technological University, India. He obtained his B.Tech. (Civil) and M.Tech. (Hyd) degree from AMU Aligarh, and a Ph.D. degree from IIT Delhi. He has published about 22 papers in referred national and international journals and conference proceedings. His research areas are applied Fluid Mechanics, surface water quality management, computational hydraulics, and hydraulic structures.

Cognitive Neurodynamics
https://doi.org/10.1007/s11571-023-09943-1

RESEARCH ARTICLE

Effectual seizure detection using MBBF-GPSO with CNN network

Dinesh Kumar Atal[1] · Mukhtiar Singh[1]

Received: 2 June 2022 / Revised: 23 December 2022 / Accepted: 5 February 2023

Abstract

EEG is the most common test for diagnosing a seizure, where it presents information about the electrical activity of the brain. Automatic Seizure detection is one of the challenging tasks due to limitations of conventional methods with regard to inefficient feature selection, increased computational complexity and time and less accuracy. The situation calls for a practical framework to achieve better performance for detecting the seizure effectively. Hence, this study proposes modified Blackman bandpass filter—greedy particle swarm optimization (MBBF-GPSO) with convolutional neural network (CNN) for effective seizure detection. In this case, unwanted signals (noise) is eliminated by MBBF as it possess better ability in stopband attenuation, and, only the optimized features are selected using GPSO. For enhancing the efficacy of obtaining optimal solutions in GPSO, the time and frequency domain is extracted to complement it. Through this process, an optimized features are attained by MBBF-GPSO. Then, the CNN layer is employed for obtaining the productive classification output using the objective function. Here, CNN is employed due to its ability in automatically learning distinct features for individual class. Such advantages of the proposed system have made it explore better performance in seizure detection that is confirmed through performance and comparative analysis.

Keywords GPSO-greedy particle swarm optimization · MBBF-modified Blackman bandpass filter · CNN-convolutional neural network

Introduction

EEG is typically a clinical process for monitoring, diagnosing, and determining neurological disorders as same as epilepsy (Issaka et al. 1506). Abnormal electrical discharge and sudden changes in the electrical activity of the brain are the main reason for the neurological disorder epilepsy. Usually, the methodology proceeds with the identification of an epilepsy seizure with the slow spike waveform. Life becomes immobile due to the unexpected nature of these seizures with impermanent damages of memory, perception, speech, and consciousness that leads to increased chances of risk for death (Kamath 2013). Almost four percentages of people in the world are affected by a seizure at a certain period during their life from that one percent are epileptic. Here, they carry out the initiation of Seizures with the help of hyperventilation, photo stimulation, and some other approaches in interictal records. But, the provoked epileptic seizures action that is not similar to natural ones was a disadvantage. No capturing and analyzing ictal events alone a vital milestone for long-term video-EEG recording; it also has an impact over the valued clinical information. When observing EEG analysis based on conventional methods, one could found that all those methods are a time-consuming and tiresome job performed by neurologists.

For these longstanding EEG recordings, visual analysis causes a human fault, and it was inefficient (Abbasi and Esmaeilpour 2017). Furthermore, they also consider the waves of background noise and the artifacts as EEG recordings of an epileptic seizure. Due to these reasons, there was a need to automatically detect epileptic seizures for reducing the time of evaluation and assistance to the neurologists. Single-channel EEG seizure detection is not adequate, owing to a nonlinear and complex dynamic system of the brain. Therefore, the multi-channel EEG processing of seizure detection plays a significant role in detecting abnormalities in the human brain. Though

✉ Mukhtiar Singh
smukhtiar_79@yahoo.co.in

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

Published online: 27 February 2023

multichannel EEG signals carried out the task of reliable mining data effectively so that only a limited number of studies were concentrated on them (Ji et al. 2011). Various ideas of research have been introduced to detect seizures, techniques like Deep Learning (Altan and Karasu 2020; Failed 2021; Sezer and Altan 2021), feature extraction, preprocessing, and categorization are involved.

Conventional researches endeavoured to perform seizure detection. Accordingly, the study (Pattnaik et al. 2022) has considered wavelet-transform for epileptic seizure detection on the EEG signals. Classification has been performed with certain Machine Learning methodologies to detect non-seizure and seizure classes. EEG signals have been gathered and overall 48-events have been regarded for evaluation. EEG signal has been decomposed into varied sub-bands through the use of TQWT (Tunable Q-Wavelet Transform) and features relying on time frequency namely entropy, further, temporal measures have been retrieved for making huge dataset for detecting epilepsy. Dataset preprocessing has been performed to classify epilepsy through RF (Random Forest) and SVM (Support Vector Machine). It has been perceived that, RF classifier has shown better outcomes with regard to accuracy at a rate of 93% in comparison to SVM classifier that has shown 90.4% accuracy. Though better outcomes have been attained, accuracy rate has to be further enhanced to be made applicable in clinical-practices.

On contrary, (Saidi et al. 2018) has used AntMiner+ and has exposed satisfactory results. Further, (Tzimourta et al. 2018) presented an approach for the evaluation of window size in automated seizure detection. In the Open Vibe scenario, statistical and spectral features were extracted and utilized for training four diverse classifiers. It could achieve the accuracy beyond 80% for the Decision Tree classifier. Likewise, from the results, it was recognized that the different window sizes offered classification accuracy with small differences. Similarly, (Sharmila and Geethanjali 2018) elucidated an epileptic signal detection with time-domain features. In this approach, they utilized an SVM and a naive Bayes for TD features that are collective and separate. This work utilized TD features like SSC, WL, and the number of ZC, and it was the first effort by the scholars. Also, they calculated remaining TD features such as average power (AVP) and standard deviation (SD) along with MAV for clarified EEG data. The performance of the classifier was enhanced when utilizing the filtered EEG TD features as input. In spite of the better performance, there still existed scope for improvement due to certain pitfalls comprising of ineffective feature selection that negatively affected the prediction rate. For differentiating normal and epileptic EEG signals, feature selection is a crucial task. Motivated by this, the present study intends to perform optimized feature selection for effectual seizure detection using below objectives.

Objectives

The main contribution of the study is,

- To eliminate noise and select optimized features using the proposed MBBF-GPSO (Modified Blackman Bandpass Filter—Greedy Particle Swarm Optimization) for improvising classification performance.
- To predict seizure using the proposed Novel CNN (Novel Convolutional Neural Network).
- To analyze the performance of the proposed system through performance and comparative analysis for assessing the effectiveness of the proposed system in detecting seizure.

Organization

The remaining sections of the paper are organized as follows: the existing feature extraction techniques related to EEG signal classification are surveyed in Sect. “Related works”. The description of the proposed methodology and feature extraction results are presented with its working flow and pseudocode in Sect. “Proposed work”. The results of the proposed work are compared by using various performance measures in Sect. “Performance analysis”. Finally, the paper is concluded, and future work is mentioned in Sect. “Conclusion”.

Related works

Sharma et al. 2018 presented an Orthogonal Wavelet Filter Bank OWFB design to minimize the frequency bands. In this work, a new SDP was utilized by the researchers without the participation of parameterization, and the filter coefficients were also acquired openly. The features extracted the designed minimally mean squared frequency localized OWFBs from the suggested automated epileptic seizure identification system. They had considered two classification tasks for identifying ictal, non-ictal, and interictal EEG signal episodes. The first one is the seizure vs. seizure-free EEG signal. Secondly, they utilized seizure vs. non-seizure EEG signals. Dominant datasets for testing the performance of the classification approach. The attained classification performance was better. Further investigative research will take up diagnosing disease and anomaly detection with electrocardiogram, magneto encephalogram, heart rate variability, and photoplethysmogram.

Altın and Er (2016) elucidated an innovative approach through acquiring the EMG signal from arm to identify the elbow gestures. For rehabilitation and assistive prostheses of paralyzed and injured people, this work utilized the EMG signal, and this was the main reason behind this work. The biological signal filtering is one of the main key points, and EMG signals filtered with noise elimination of 50 Hz mains supply. After signal filtering, for both wrist and wrist extension cases, the feature extraction was applied. In the time and frequency domain, many feature extraction methods were applicable. By using the extracted features, they studied the classification of hand movements by making use of the K Nearest Neighbor algorithm. EMG signal acquisition tool is the acquired dataset utilized in this approach. Thus the accuracy yields better in classification due to its K nearest neighbor algorithm. Further, an EEG signal for seizure detection with an innovative classification approach was discussed in this paper (Ahmadi et al. 2018). There was preference given for a wavelet-based cross-frequency coupling for the classification of ictal seizures in which the features were also extracted using WBCFC. Then the optimal features were designated from the wavelet coefficient using a *t*-test, and they completed the classification with the quadratic discriminant analysis. An innovation done in the feature extraction provided better performance in the classification process. The experimental outcome of accuracy was satisfactory.

Tiwari et al. (2017) presented an innovative method to diagnose epilepsy automatically based on EEG. Using the pyramid of the difference of Gaussian (DoG) filtered signal, it detected several measures of key points in the EEG signal. From these detected key points, one could calculate local binary patterns and acquire the feature set from the histogram of these patterns. And then, it was fed up into SVM to classify the EEG signal. This approach examined four types of classification problems, such as epileptic and normal seizure, seizure-free and epileptic seizure, and seizure-free, epileptic seizure and normal, non-seizure, and epileptic seizure EEG signals utilizing the openly offered database of University of Bonn EEG database. They compared the classification accuracy with the existing approach, and they proved that their approach was efficient to analyze the above problem of classification. For categorizing seizure-free EEG signals and seizure, they proved that the current approach was effective with experimental results.

Sriraam et al. (2018) provided a methodology for analyzing the multichannel EEG to detect the epileptic seizure with the use of entropy, PSD, MLPNN classifier, and Teager energy. In this method, firstly, EEG signals were processed previously for noise removal, followed by feature extraction. Extracted features appropriateness with its band difference was tested to classify the normal and epileptic seizures using Wilcoxon rank-sum test and descriptive analysis. This work revealed the performance measure of sensitivity, specificity, and false-positive rate for a multi-feature for achieving comparatively better performance than other existing methods. To afford an automated biomarker for epileptic and normal EEG signals, the 'Aepitect' signification of the graphical user interface was established in MATLAB.

Lu et al. (2018) presented a new technique for the arrangement of the EEG signal for epileptic seizure identification. Kraskov entropy of the envelopes of IMF signals disintegrated by Hilbert-Huang transform was calculated for the innovative feature extraction method. The hybrid twelve-dimensional feature vector was regarded as input for the LS-SVM classifier by combining these kinds of features with the other two features. For evaluating generality and universality of this method, they made use of three different datasets, and then, defined the three binary-classification problems. Having hybrid features makes the system more efficient for the detection of an epileptic seizure. But still some problems were presented in real-time applications, to overwhelm this issue, a new prospect was given to contribute neurophysiologists to diagnose the epileptic seizures accurately. Thilagaraj et al. (2018) presented a new single feature, specifically Tsallis entropy, with five different classifiers. This work compared their proposed method with other preceding methods, and it founded that suggested method takes the minimum time of computation. This method accomplished a peak accuracy using a decision tree classifier designed for the four types of two-class classification problems. It was a simple method, and its computation time was lesser. Thus it was scrutinized that this method offered better detection of epileptic seizures using novel single features names as tsallis entropy.

Guo et al. (2018) introduced an Extended Correlation-Based Feature Selection (ECFS) where the feature selection has a vital role because the efficiency of the system can be affected while selecting the features. The feature space selected over here was fed up into five classification algorithms, such as Support Vector Machine Random Forest, Multilayer Perceptron, RBF Network, and Logistic Model Trees. Among those classifiers, the LMT classifier achieves better than other approaches. Thus to diagnose epilepsy, this selection method was effective. (Ahammad 2014) elucidated a novel method of detecting the epileptic seizure event to achieve better sensitivity by avoiding wavelet decomposition and utilizing some demographic features and wavelet-based features. With the linear classifier, they were able to divide epileptic EEG signals, and a normal were divided into three types as epilepsy patients for the period of epileptic seizures, a healthy volunteer with the eye open, and epilepsy patients in the epileptogenic zone in a seizure-free interval. In this process, it calculated features such as entropy, energy, minimum, maximum, standard deviation, and mean at various sub bands. This work made use of the database of Bonn University EEG for detecting seizure event detection. This work computed parametric measures like sensitivity, specificity and accuracy for seizure detection. In such a case of onset seizure detection, they made use of the CHB-MIT scalp EEG database, including wavelet-based features, mean absolute deviation (MAD), and interquartile range (IQR) without wavelet decomposition were taken out. For analyzing the seizure, this work utilized onset recognition performance, latency.

Furthermore, (Raghu and Sriraam 2017) presented an enhanced structure of MLP-NN (Multilayer Perceptron Neural Network) using a pattern classifier to identify the iEEG (intracranial electroencephalogram epileptic seizures). Even, they involved vast procedure for the qualitative analysis of intracranial recordings, it offers important brain neuronal actions for the medical judgment was essential. (Raghu et al. 2017) proposed a study to detect the epileptic seizure automatically with the highest accuracy comparatively than the existing approaches. The log and norm entropies based on wavelet packet with a REN (recurrent Elman neural network) was preferred as it has more specific characteristics. The epileptic, normal pre-ictal EEG recordings are the three types of classification measured in this study. From the original EEG recordings, 50 Hz power line noise was removed initially using an adaptive Weiner filter; for safeguarding the stationary of the signal, this methodology divided original EEGs to 1s patterns. For indicating the fluctuations in the features, they implemented the non-linear Wilcoxon statistical test. And also, log energy entropy with no wavelets is deliberated. Thus the simulation proves that the classification accuracy was better in this REN classifier consisting of wavelet packet log entropy.

Problem identification

From the existing works, we observe that the current techniques have both advantages and disadvantages; however, it mainly lacks the subsequent limits:

- Inefficient feature selection
- Increased computational complexity and time
- The detection takes more time for finding epilepsy
- The classification accuracy is less for identifying the seizure and seizure of free classification.

For overcoming these issues, this paper aims to develop a new classification based seizure detection methodology.

Proposed work

In the proposed system, we perform the epileptic seizure prediction by pattern recognition based on the time—domain analysis. In this study, the EEG database (2018) from the University of Bonn, Germany, is utilized. This database contains five types of EEG data sets (signified as Set A–E). One hundred single-channels are available for each dataset, and its duration is 23.6 s.

Initially, the modified Blackman bandpass filter is used to suppress the signals which are at the other frequency than predefined frequencies. This filter is used to discriminate the ranges of the frequency a pass along with the passage of signal within specific frequencies wherein the filter attenuates the outside ranged signal.

We are extracting the time domain features to describe the EEG signals for the detection of an epileptic seizure. The listed time domain features are Mean value of the Square Root (MSR), Absolute value of the Summation of Square root (ASS), mean absolute value (MAV), waveform length (WL), Absolute value of the Summation of the exponential root (ASE), zero crossings (ZC), slope sign changes (SSC), auto-regressive coefficients (AR), Wilson Amplitude (WAMP) and root mean square (RMS). Also extract frequency domain features such as Maximum Amplitude, Mean Frequency, Median Frequency, Power Spectrum deformation, and Spectrum Moment.

These above-estimated features are optimized using Greedy Particle Swarm Optimization (GPSO) for selecting the best optimal features. After that, those optimized features are classified using a Novel Convolutional Neural network (CNN). We update the convolutional layer of CNN for training based on the extracted features and labels.

Finally, the degrees of severity of the detected epilepsy are also classified. On comparing the proposed algorithm with the existing techniques on the scale, the parameters such as TP, TN, FN, FP precision, recall, F-Score, Kappa and Dice Coefficients, sensitivity, specificity, and accuracy are found to be holding right (Fig. 1).

EEG input signal and modified Blackman band pass filter

We consider an input signal from the EEG for the filtering process. The EEG input signal is shown in the figure (Fig. 2).

In the window function method, a leakage is dispersed spectrally in different ways based on required specific applications. The Blackman window seems the same as Hanning and Hamming windows. A Blackman window has benefits over other windows by providing better stopband

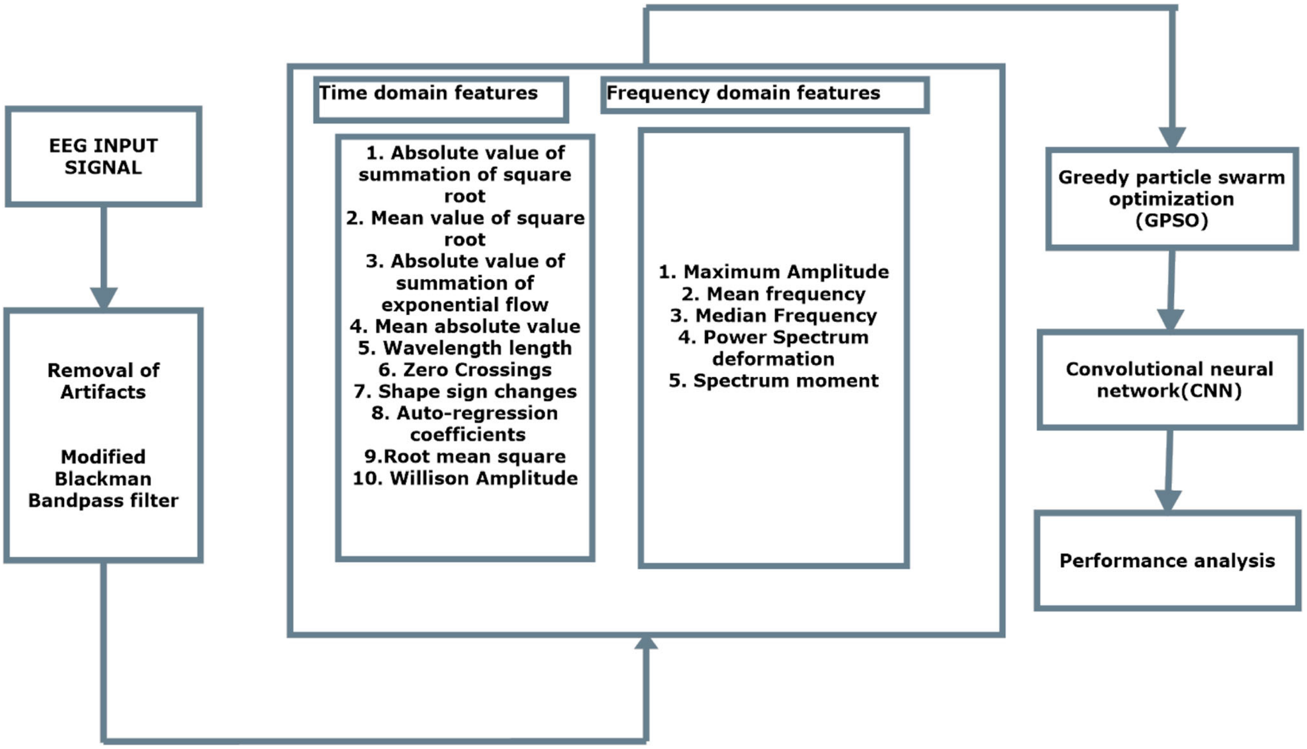

EEG INPUT SIGNAL
Removal of Artifacts
Modified Blackman Bandpass filter
Time domain features
1. Absolute value of summation of square root
2. Mean value of square root
3. Absolute value of summation of exponential flow
4. Mean absolute value
5. Wavelength length
6. Zero Crossings
7. Shape sign changes
8. Auto-regression coefficients
9.Root mean square
10. Willison Amplitude
Frequency domain features
1. Maximum Amplitude
2. Mean frequency
3. Median Frequency
4. Power Spectrum deformation
5. Spectrum moment
Greedy particle swarm optimization (GPSO)
Convolutional neural network(CNN)
Performance analysis

Fig. 1 Overall flow diagram

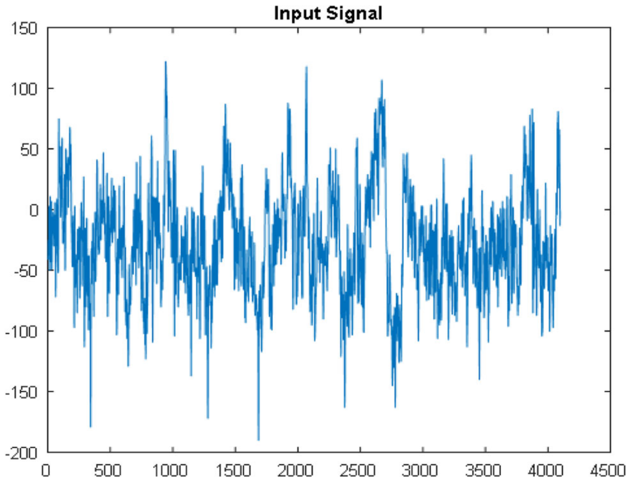


Fig. 2 Input signal

attenuation. With less passband ripple (Tiwari et al. 2014), i.e., Blackman window can show lower Maximum stop-band ripple (about 74 dB down) than the hamming window in the outcome of FIR filter. The mathematical equation is shown below.

The method initiates with an optimal desired frequency–response that is denoted by,

$$H_{id}(W_i) = \sum_{n=0}^{\infty} h_{id}(n_i) e^{-jjw_i n_i} \qquad (1)$$

wherein

$h_{id}(n_i) = \frac{1}{2}\int_{-\pi}^{\pi} H_{id}(W_i) e^{-jjw_i n_i} dw_i$ 1

The below equation states the Blackman-window of length (N).

$$\begin{aligned} \mathrm{W(n)} &= 0.42 - 0.5\cos(2\pi \mathrm{nN} - 1) + 0.08(4\pi \mathrm{nN} \\ &\quad - 1), \text{where n} \\ &= 0, 1, \mathrm{N} - 1 \end{aligned} \qquad (2)$$

To eliminate certain stop band-ripples and pass band ripples, hamming window method is utilized. Coefficients for hamming window computation is given by,

$$w_i(n_i) = 0.54 - 0.46cos(2\pi n_i N_i), 0 \le n_i \le N_i \qquad (3)$$

In our work, initially modified Blackman bandpass filter is utilized to suppress the signals which are at the other frequency range than the predefined frequencies. We use the bandpass filters for distinguishing the ranges of the frequencies and for passing the signal within specific frequencies and attenuates the signal, which is in the outside scope (Fig. 3).

Time and frequency domain feature extraction

After applying the Blackman bandpass filter to the EEG input signal, the feature extraction process is carried out in EEG data. Feature extraction is employed to reduce the dimensionality (cut into feature vector) when we present

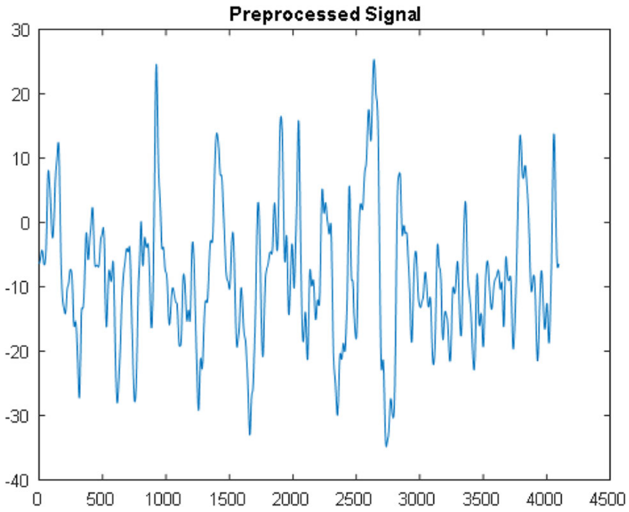


Fig. 3 Pre-processed signal

massive input data. The three sources of information, such as spatial feature, Spectral (Frequency) Domain feature, and we can extract the temporal feature from the feature extraction process. The time and frequency domain features, one from each category, are shown in Tables 1, 2, 3, 4, 5, and 6.

Healthy

Table 1 Time domain features

Absolute value of the summation of square root (ASS)	7.0766e + 03
Mean value of the square root (MSR)	1.5066
Absolute value of the summation of the exponential root (ASE)	1.8246
Mean absolute value (MAV)	6.5056
Waveform length (WL)	246.1464
Zero crossing	76
Slope sign change (SSC)	182
Auto-regressive coefficients (AR)	− 0.3308
Root mean square (RMS)	0.5430
Wilson amplitude (WAMP)	2043

Table 2 Frequency domain features

Maximum amplitude	24.7663
Mean frequency	8.4973e + 03
Median frequency	0.0325
Power spectrum deformation	− 3.2031e + 03

Interictal

Table 3 Time domain features

Absolute value of the summation of square root (ASS)	1.1653e + 04
Mean value of the square root (MSR)	0.5973
Absolute value of the summation of the exponential root (ASE)	4.9285
Mean absolute value (MAV)	13.3298
Waveform length (WL)	238.8708
Zero crossing	52
Slope sign change (SSC)	146
Auto-regressive coefficients (AR)	− 0.3284
Root mean square (RMS)	0.7278
Wilson amplitude (WAMP)	1871

Table 4 Frequency domain features

Maximum amplitude	31.3704
Mean frequency	− 3.9897e + 04
Median frequency	0.0212
Power spectrum deformation	− 3.2031e + 03

Ictal

Table 5 Time domain features

Absolute value of the summation of square root (ASS)	1.7329e + 04
Mean value of the square root (MSR)	3.1375
Absolute value of the summation of the exponential root (ASE)	6.0361
Mean absolute value (MAV)	41.5786
Waveform length (WL)	3.2913e + 03
Zero crossing	174
Slope sign change (SSC)	204
Auto-regressive coefficients (AR)	− 0.3324
Root mean square (RMS)	6.0308
Wilson amplitude (WAMP)	3962

Table 6 Frequency domain features

Maximum amplitude	145.0774
Mean frequency	1.6044e + 04
Median frequency	0.1052
Power spectrum deformation	− 3.2031e + 03

In the present study, the characteristics of time-domain features like ASS (Absolute value of the Summation of Square root), MSR (Mean value of the Square Root), ASE (Absolute value of the Summation of the exponential root), MAV (mean absolute value), WL (waveform length), ZC (zero crossings), SSC (slope sign changes), AR (auto-regressive coefficients), RMS (root mean square) and WAMP (Wilson Amplitude) are extracted from EEG input for characterizing the patterns of EEG for detecting the epileptic seizure. And also, the extraction of the frequency domain features such as Maximum Amplitude, Mean Frequency, Median Frequency, Power Spectrum deformation, and Spectrum Moment are done.

The extraction of the time domain feature is not only possible for the Non-stationary and transient characteristics of the EEG signal. In contrast, the frequency domain features should also be extracted, i.e., Fundamental characteristics of EEG. While acquiring constant frequency on the EEG waveform, it presents the fundamental rhythms of frequency such as theta, beta, gamma, and Delta, which is similar to several brain states, functions, or pathologies.

Greedy particle swarm optimization

A greedy particle swarm optimization is the optimization method where we consider extracted features as an input, and then it is optimized by selecting the best features among them. A PSO (particle swarm optimization) is a computational technique to enhance candidate solutions through optimizing the problem recurrently. By having a population of a candidate solution, the position has solved the issue of the particles so that particles could get moved around the space of search based on theoretical formulas over the velocity and location of the particle. Every movement of particle influences concerning the best local position known. However, we can rely on the best areas known of search space, which we reorganize as a better position with another kind of particle. It probably reached the best solution with the movement of the swarm. Here the greedy method is also involved where the objective function is assumed first, and it has to be optimized (either maximized or minimized) at a specified point. At each step, we made the greedy choices in the greedy algorithm for ensuring that the objective function is optimized; the decision made here could not be reversed back for computing optimal solutions. For enhancing the effectiveness of obtaining optimal solutions, a frequency and time domain feature extraction was announced to complement the greedy system. We get the best solution from the personal best solutions of particles in a greedy manner. Thus greedy particle optimization is the best method for optimizing the problems towards making a better solution through updating the velocity and the position of the particle. The greedy algorithm has the advantage of analyzing the runtime quickly.

Novel convolutional neural network

Applying CNNs to EEG signal are becoming familiar because of its vast usage and effectiveness towards predicting the seizure. The CNN method is one of the fruitful ones and established better results in this research community.

A CNN contains an output and input layer, and also numerous hidden layers, whereas hidden layers of CNNs comprised the pooling layers, fully connected layers, and convolutional layers. The convolution operation is applied as an input by the CNN layers, transporting the outcome to the subsequent segment. The convolution rivals each neuron's response for the visual stimuli, networks including local or global pooling layers where we give the one layer's output of neuron cluster to the next neuron. We utilize an average value taken from the previous layer containing individual group for mean pooling. Entirely neuron network is the network where it links all the neurons from the one layer to the layer that is composed of every single neuron. CNN layer has the same standard as like traditional perceptron neural network that is multilayered. For high dimensional data analysis, CNN is a more efficient one. The parameter sharing scheming is employed and utilized in convolutional layers for reducing the number of parameters, and also quantity of parameters and computation in the network, spatial size of the representation, and consequently control overfitting, are reduced by the design of pooling layer.

The CNN models utilized four layers. Initially, we initialize the CNN parameter. Physically, CNNs have convolutional layers combined with pooling layers, tracked by fully connected layers.

(1) *Convolutional layer*

It links the convolutional layer contains the 64 feature maps that to the input layer through 3*3 kernels, including the kernels which glide over EEG signals diagonally.

Algorithm: 1 Greedy Particle Swarm Optimization

```
Input: $Ex_c$ - Sequences of extracted features
Output: $Op_d$Optimization data
Procedure:
S, the Size of the Total sequences to be features
Initialize the decision variable $nvar_c$)
Initialize $decision\ matrix\ (var_c)$
Maximum iteration $(max_{iter})$
To set the lower and upper bound variables,
$vel_{min}$=-10
$vel_{max}$=10
Let population size $pop_{num}$ =length ($Ex_c$)
Let initialize the greedy_PSO parameters,
 Inertia weight wt=1
Let inertia weight damping ration $wt_{damp}$ =0.99
To find velocity limits,
$vel_{max}$=0.1*($vel_{max}$-$vel_{min}$)
$vel_{min}$=-$vel_{max}$
 for i=1:$pop_{num}$
      Let initialize position, value, velocity and best value,
 To find the particle value($\rho_{value}$)
$\rho_{value}$ (i).position=($Ex_c$(i)),
      Let initialize velocity,
$\rho_{value}$ (i).velocity=$zeros(var_c)$,
          Let evaluation greedy process,
end
 for  i=1:$max_{iter}$
    for i=1:$pop_{num}$
Clustered extraction from the sequences
Say S= $(er_c)$size
For i=1 to S
      Let input data=$er_{c\ i,j}$
$An_{d(i,j)}$=$\left(\sqrt{er_{c\ i}} - \sqrt{er_{c\ j}}\right)^2$-$\left(\sqrt{er_{c\ j}} - \sqrt{er_{c\ i}}\right)^2$
     End
    Let $v_f$=$x_i - x_j/t$ //initialize ant velocity
             Let $r_{nd}$= random numbers choose
      Compute $p_s$=function ($An_d, r_{nd}$ )
  For i=1 to 255 // random numbers
Update the velocity
Let the velocity limits
$\rho_{value}$= wt*($\rho_{value}$. velocity)+rand ($nvar_c$)*$\rho_{value}$. position+ rand($nvar_c$)
      For k=1-$r_{nd} - 1$
$p_s$=($r_{nd(i,k)} + r_{nd(i,k+1))}$- (($r_{nd(i,k+1)} + r_{nd(i,k+1)}$))+$r_{nd}$
$v_f$=ω($r_{nd} - 1$)*$v_f(r_{nd}) - An_d/p_s(\sqrt{r_{nd}})$
End
End
Update $An_d$ values,
Update the position,
Objective function $obj_n$
$Op_d$=$obj_n(An_d)$
Best value=mean ($p_s$)
Fitness value $f_{fn}$
If  $f_{fn} < p_s$
$Op_d$=$f_{fn}$
End
```

$$Y_k = \sum\nolimits_{n=0}^{N-1} Z_n h_{k-n} \tag{4}$$

A kernel includes the matrix to be entwined with the input EEG signal and stride (stride = 1) and controls the degree to which the filter combines diagonal input signals.

(2) *Pooling layer*

The downsampling layer is known as the pooling layer, and it only selects the maximum value for each feature map, and then the output neurons are reduced accordingly.

(3) *Fully connected layer*

Here, it acquires filled association towards the complete activations over the other layer. Then finally, the activation layer the classification output is attained (i.e., ictal, interictal, or preictal) by sigmoid and tanh activation.

Tanh or hyperbolic tangent Activation Fnandri mam Function

$$\mathrm{A} = 1 - (z * z) \tag{5}$$

Sigmoid activation function:

$$A = z * (1 - z) \tag{6}$$

$z - Features\,values$

Algorithm 2 depicts Novel Convolutional Neural Network. This methodology indicates the input as optimized sequences f_{fn} and will initialize the training and testing set size. Extracted the features from the sequences and created a list of feature sets. Then the CNN parameter gets initialized, and we found the backpropagation. The layers like the pooling layer, fully connected layer and convolutional layer, and the activation layer are utilized in combination for obtaining the classified output through sigmoid function and tanh (Fnandri mam Function).

Performance analysis

This performance analysis is the section where the performance metrics can be experimentally analyzed. The features like specificity, accuracy, sensitivity, F score, recall precision, and the Jaccard coefficient are estimated to find the seizure from the EEG signal in an effective manner than the other conventional methods. The classification approach of proposed CNN is superior to acquire better output for classifying the seizures effectively.

Figure 4 shows the performance metrics evaluation for healthy, interictal, and ictal. The accuracy of the healthy interictal and ictal seizure for our proposed method have attained 99.3, 99.2, 98.2, and 99; similarly, the sensitivity reached 99.5, 97, and 98. The specificity of this proposed method is also estimated and specified as 99, 99, and 99.25. Additionally, precision, recall F measures, and the Jaccard coefficient are also evaluated to prove the efficiency of the proposed technique that uses an effective classification approach.

Figure 5 demonstrates the performance metrics evaluation for S-Z, S-F, S-N, and S-ZNF. The EEG dataset includes five subsets denoted as Z, O, N, F and S. In this work, four different types of classification tasks S-F, S-Z, S-N and S-ZNF are considered to evaluate the performance of the proposed system. The accuracy rate of the proposed method for S-F is 0.5% superior to the existing one; for S-Z, it deviates 4.3%, and S-ZNF is 0.3% higher than the traditional one. Similarly, for the sensitivity, the S-Z, S-F, S-N, S-ZNF acquires the range of 95, 100, 98.5, and 99. The specificity of the S-Z, S-F, S-N, S-ZNF are 95, 100, 98.5, and 99. And also, the precision, recall F measures, and the Jaccard coefficient are evaluated where we achieve superiority in our proposed method. The Jaccard coefficient value for S-Z, S-F, S-N, S-ZNF is 90.4545, 100, 97.0043, and 98.6689, which is comparatively better than other approaches.

Figure 6 depicts the Comparison of Performance measure evaluation for proposed and existing method(Yu et al. 2019). The comparison measure shows that the proposed work MBBF-GPSO offers better performance for acquiring better S-F, S-ZNF than the existing KRPCF (kernel robust probabilistic collaborative function). The accuracy of the proposed method for the S-F value is higher than the conventional way. The sensitivity of the suggested technique for S-F, S-Z, and S-ZFN are 100, 95, and 99, which is superior to the traditional technology that yields 99,

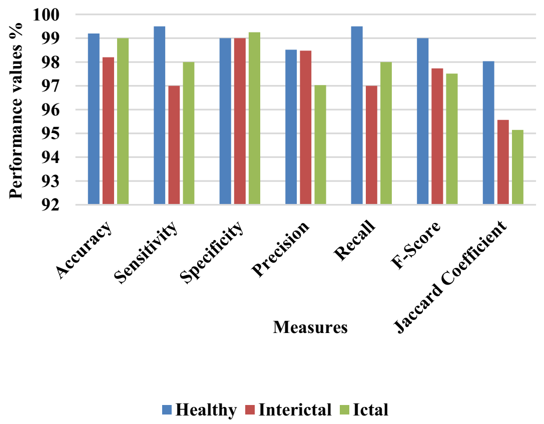


Fig. 4 Performance metrics evaluation for healthy, Interictal and ictal

98.75, and 98. Thus MBBF-GPSO is superior to existing work, and hence our proposed method works better with its brilliant framework. Further, the performance of the proposed system has been assessed with and without cross validation for exposing its performance with and without uncertainty. The corresponding results are shown in Tables 7 and 8.

From Table 7, it is found that, proposed system works better with cross validation and has exposed 99.76% accuracy in comparison to its performance without cross validation. Similarly, the proposed approach works better with cross validation and has exposed minimum error rate (0.423) in comparison to its performance without cross validation. Hence, the performance of proposed system seems to be uncertain when it is considered without cross validation and is certain when it is considered with cross-validation. Further, empirical outcomes of the proposed system with and without optimization are exposed in Fig. 7.

From Fig. 7, it is found that, 300 features are selected with optimization and 500 features exists without optimization. In addition, comparison has been performed with filters and proposed system. The corresponding outcomes are exposed in Table 9 wherein the efficacy of the proposed system is confirmed.

Comparative analysis

Proposed system has been comparatively assessed with conventional classifiers with regard to accuracy. The corresponding outcomes are depicted in Table 10.

From Table 10, it has been revealed that, conventional classifiers like RBFNN has shown better accuracy at a rate of 97.47%, SVM has exposed 98.78%, KNN has shown 97%. However, the proposed system has revealed high accuracy than conventional algorithms at a rate of 99.65% as represented in bold below. Further, analysis has been undertaken with regard to other metrics (recall, precision and F-measure). The corresponding outcomes are shown in Fig. 8.

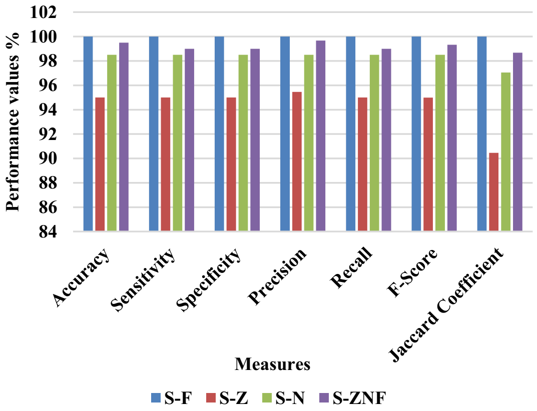


Fig. 5 Performance measure evaluation for S-Z, S-F, S-N, S-ZNF

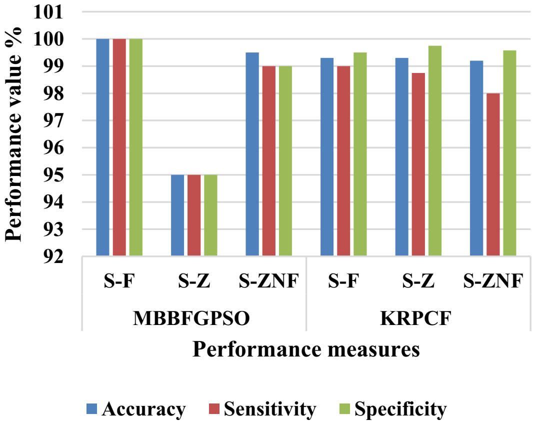


Fig. 6 Comparison of performance measure evaluation for a proposed and existing method

From Fig. 8, it has been found that, conventional approaches like HVD-LSTM has shown better performance in accordance with metrics. In comparison to other traditional algorithms, proposed system has revealed better performance. Furthermore, comparison has been accomplished in accordance with accuracy by considering conventional algorithms like SVM, Naïve Bayes, Logistic Regression and KNN. The respective outcomes are shown in Table 11.

From Table 11, it has been exposed that, existing algorithms like SVM has exposed 74.44% accuracy, while, KNN has shown 94.56%, LR has revealed 94.69% and NB has shown 91.56%. However, the proposed system has revealed high accuracy than conventional algorithms with 99.65%. The proposed MBBF has eliminated unwanted signals due to its better ability in stopband attenuation. Further, only optimized features have been chosen by GPSO. Further, ideal solutions for GPSO are obtained by extracting frequency and time domain features for complementing it. The study also employed CNN due to its innate capability in automatically learning distinct features for individual class. These advantages of the proposed system have made it accomplishing better performance than conventional methods in seizure

Table 7 Analysis with regard to accuracy

	Accuracy
Without cross validation	99.3
With cross validation	99.76

Table 8 Analysis with regard to error

	Error
Without cross validation	0.581
With cross validation	0.423

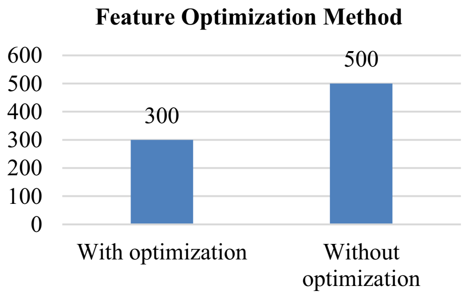


Fig. 7 Features selected with and without optimization

Table 9 Comparison of filters with proposed system (Raza et al. 2014)

Blackman	− 58.2 dB	0.10938
Hamming	− 41.7 dB	0.085938
Kaiser (β = 5)	− 37.8 dB	0.085938
Proposed	− 13.6 dB	0.054688

detection with 99.65% accuracy as confirmed by comparative analysis.

Conclusion

In this innovative study, MBBF-GPSO technique is employed to obtain enhanced results for effective seizure prediction. We utilized the modified Blackman Band Pass filter in which the Blackman window provides better stopband attenuation and less passband ripple. Also, we used the bandpass filter for attenuating the signal, which is out of the range. We extracted the frequency domain and time domain features separately, and it is fed up into the optimization process named GPSO. After this process, to acquire a desired classified output, CNN is utilized. Thus from the experimental outcome, it is clear about the outperformance of the suggested method than other conventional means. Then, we make the cross-validation with traditional methods, and it found that the proposed technique offers better outcomes. The complexity of the proposed model is given by O(n(log(n)) In the future, we look to extend the work by taking different databases like the Freiburg database or some real-time human database with some other optimization techniques.

Table 10 Analysis in accordance with accuracy (Saminu et al. 2021)

Classifier	Accuracy
RELS-TSVM	90.2
RBFNN	97.47
Exponential Energy	99.5
SVM, FFANN	99
SVM	98.78
ANN	91.1
KNN	97
Proposed CNN	**99.65**

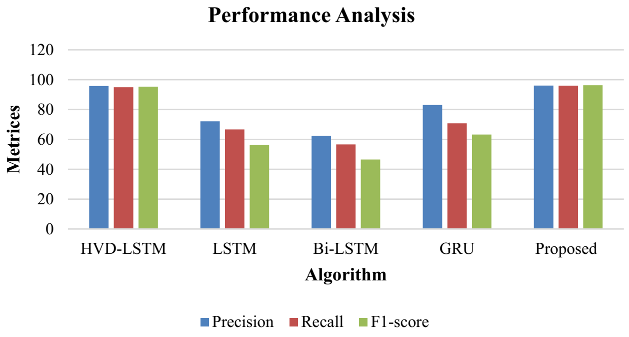


Fig. 8 Analysis in accordance with performance metrics (Khan et al. 2021)

Table 11 Analysis in accordance with accuracy (Karabiber Cura et al. 2020)

Method	Accuracy
SVM	74.44
KNN	94.56
Naive Bayes	91.56
Logistic regression	94.69
Proposed	99.65

Funding The authors declare that this study did not receive any funds from any agency.

Data availability The datasets generated analysed during the current study are available in the [*EEG time series*] repository, [http://epileptologiebonn.de/cms/front_content.php?idcat=193&lang=3&changelang=3].

Declarations

Conflict of interest This is to certify that all authors have seen and approved the manuscript titled "Effectual Seizure Detection using MBBF-GPSO with CNN network" being submitted and authors have no conflict of interest to declare.

References

Abbasi R, Esmaeilpour M (2017) Selecting statistical characteristics of brain signals to detect epileptic seizures using discrete wavelet transform and perceptron neural network. IJIMAI 4:33–38

Ahammad N (2014) Detection of epileptic seizure event and onset using EEG. Biomed Res Int 2014:7

Ahmadi A, Behroozi M, Shalchyan V, Daliri MR (2018) Classification of epileptic EEG signals by wavelet based CFC. In 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), pp 1–4

Altan A, Karasu S (2020) Recognition of COVID-19 disease from X-ray images by hybrid model consisting of 2D curvelet transform, chaotic salp swarm algorithm and deep learning technique. Chaos Solitons Fractals 140:110071

Altın C, Er O (2016) Comparison of different time and frequency domain feature extraction methods on elbow gesture's EMG. Eur J Interdiscip Stud 5:35–44

Guo Y, Zhang Y, Mursalin M, Xu W, Lo B (2018) Automated epileptic seizure detection by analyzing wearable EEG signals using extended correlation-based feature selection

Issaka MA, Dabye AS, Gueye L (2015) Localization of epileptic seizure with an approach based on the PSD with an autoregressive model. arXiv preprint arXiv:1506.00947

Ji Z, Sugi T, Goto S, Wang X, Ikeda A, Nagamine T et al (2011) An automatic spike detection system based on elimination of false positives using the large-area context in the scalp EEG. IEEE Trans Biomed Eng 58:2478–2488

Kamath C (2013) A new approach to detect epileptic seizures in electroencephalograms using teager energy. ISRN Biomed Eng 2013:14

Karabiber Cura O, Kocaaslan Atli S, Türe HS, Akan A (2020) Epileptic seizure classifications using empirical mode decomposition and its derivative. Biomed Eng Online 19:1–22

Khan P, Khan Y, Kumar S, Khan MS, Gandomi AH (2021) HVD-LSTM based recognition of epileptic seizures and normal human activity. Comput Biol Med 136:104684

Lu Y, Ma Y, Chen C, Wang Y (2018) Classification of single-channel EEG signals for epileptic seizures detection based on hybrid features. Technology and Health Care, pp 1–10, 2018.

Pattnaik A, Rout N, Sabut S (2022) Machine learning approach for epileptic seizure detection using the tunable-Q wavelet transform based time–frequency features. Int J Inf Technol 14:3495–3505

Raghu S, Sriraam N (2017) Optimal configuration of multilayer perceptron neural network classifier for recognition of intracranial epileptic seizures. Expert Syst Appl 89:205–221

Raghu S, Sriraam N, Kumar GP (2017) Classification of epileptic seizures using wavelet packet log energy and norm entropies with recurrent Elman neural network classifier. Cogn Neurodyn 11:51–66

Raza GA, Alam MJ, Ansari MN (2014) Design and performance analysis of band pass filter using Blackman, Hamming and Kaiser windows. IJRET 3:211–214

Rg A EEG data (2018). https://www.ukbonn.de/en/epileptology/workgroups/lehnertz-workgroup-neurophysics/downloads/

Saidi A, Othman SB, Kacem W, Saoud SB (2018) FPGA implementation of EEG signal analysis system for the detection of epileptic seizure. In 2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET), pp 415–420

Saminu S, Xu G, Shuai Z, Abd El Kader I, Jabire AH, Ahmed YK et al (2021) A recent investigation on detection and classification of epileptic seizure techniques using EEG signal. Brain Sci 11:668

Sezer A, Altan A (2021) Detection of solder paste defects with an optimization-based deep learning model using image processing techniques. Solder Surf Mount Technol 33:291–298

Sezer A, Altan A (2021) Optimization of deep learning model parameters in classification of solder paste defects. In 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), pp 1–6

Sharma M, Bhurane AA, Acharya UR (2018) MMSFL-OWFB: a novel class of orthogonal wavelet filters for epileptic seizure detection. Knowl Based Syst 160:265–277

Sharmila A, Geethanjali P (2018) Effect of filtering with time domain features for the detection of epileptic seizure from EEG signals. J Med Eng Technol 42:217–227

Sriraam N, Raghu S, Tamanna K, Narayan L, Khanum M, Hegde A et al (2018) Automated epileptic seizures detection using multi-features and multilayer perceptron neural network. Brain Inform 5:10

Thilagaraj M, Rajasekaran, Kumar NA (2018) Tsallis entropy: as a new single feature with the least computation time for classification of epileptic seizures. Cluster Computing, pp 1–9, 2018

Tiwari A, Thakre V, Markam K (2014) Design technique of bandpass FIR filter using various window function. IJCCER 2:93–99

Tiwari AK, Pachori RB, Kanhangad V, Panigrahi BK (2017) Automated diagnosis of epilepsy using key-point based local binary pattern of EEG signals. IEEE J Biomed Health Inform 21:888–896

Tzimourta KD, Astrakas LG, Gianni AM, Tzallas AT, Giannakeas N, Paliokas I et al (2018) Evaluation of window size in classification of epileptic short-term EEG signals using a Brain Computer Interface software. Eng Technol Appl Sci Res 8:3093–3097

Yu Z, Zhou W, Zhang F, Xu F, Yuan S, Leng Y et al (2019) Automatic seizure detection based on kernel robust probabilistic collaborative representation. Med Biol Eng Comput 57:205–219

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

processes

MDPI

Article

Experimental and Simulation Study of the Latest HFC/HFO and Blend of Refrigerants in Vapour Compression Refrigeration System as an Alternative of R134a

Uma Shankar Prasad [1,2,*], Radhey Shyam Mishra [3], Ranadip Kumar Das [1] and Hargovind Soni [4,*]

[1] IIT(ISM), Dhanbad 826004, India
[2] PIEMR, Indore 452010, India
[3] Department of Mechanical Engineering, Delhi Technological University, New Delhi 110006, India
[4] Department of Mechanical Engineering, National Institute of Technology Delhi, Delhi 110036, India
[*] Correspondence: dp16dp.16dp000005@mech.iitism.ac.in or ushankar@piemr.edu.in (U.S.P.); hargovinds@nitdelhi.ac.in (H.S.)

Abstract: Experimental and simulation investigation of the performance and characteristics of different refrigerants and blends of refrigerants is carried out to replace the existing refrigerant R134a for a vapour compression refrigeration system. The performance of VCRS systems was improved by several researchers by introducing the concept of mixing the family of refrigerants with low GWP in the working circuit. This research paper presents the performance results of different refrigerants and blends of refrigerants that can replace the R134a it is also an attempt to cover the mechanism and possible combination of different blends of refrigerants to improve the effectiveness as well as efficiency of the refrigeration system. Detailed analysis of different parameters of heat transfer and predictions of low-GWP refrigerants, including the HFO (hydro fluoro-olefin) class and the HC (hydrocarbon) class through energy and exergy analysis of commercial refrigerants such as R134a is performed. Results are obtained by using an experimental test rig and the input parameters of the experiments are kept the same with the simulation software (CYCLE_D-HX 2.0) and validated with the results to replace R134a.

Keywords: hydrofluoroolefins; hydrofluorocarbons; refrigeration system; energy technology; environmentally friendly; alternative refrigerants; low GWP

check for updates

Citation: Prasad, U.S.; Mishra, R.S.; Das, R.K.; Soni, H. Experimental and Simulation Study of the Latest HFC/HFO and Blend of Refrigerants in Vapour Compression Refrigeration System as an Alternative of R134a. *Processes* **2023**, *11*, 814. https://doi.org/10.3390/pr11030814

Academic Editors: Kai-Shing Yang, Jin-Cherng Shyu and Ming-Tsang Lee

Received: 6 February 2023
Revised: 27 February 2023
Accepted: 1 March 2023
Published: 9 March 2023

1. Introduction

The process of maintaining and achieving the temperature below the surrounding is known as refrigeration. The main aim is to lower the temperature of the product or space to the desired one. Maxwell et al. [1] discussed the history of refrigeration with the basic concept of modern refrigeration systems using ammonia, carbon dioxide and aqua-ammonia as a refrigerant. Several researchers investigated and found that efficiency can be enhanced by the blending of refrigerants with different proportions. During several investigations, it was noted that these blended refrigerants have several novel properties that make them very useful in various types of applications such as fuel cells, heat sinks, heat exchangers, heat transfer, hybrid engines, pharmaceutical processes, microelectronics, grinding, machining, the cooling system of different components, chillers, domestic refrigerators, solar applications, greenhouse applications. They can improve the thermal conductivity of the base fluid as well as it was noted that convective heat transfer is also significantly improved. According to an international agreement (Montreal Protocol, 1987), the use of CFC halogenated refrigerants such as R-11, R-12, R-113, R-114, and R-502 that have high ODP have been phased out. The most widely used refrigerants R-11 and R-12 are replaced because of their long-term greenhouse effects with R134a. HFC -134a is a non-flammable hydrofluorocarbon and is the best alternative to R-12 CFC refrigerant, but due to high GWP of 1430, insolubility and incompatibility with organic mineral oils, the Environmental Protection Agency (EPA) listed R134a in the category of unacceptable refrigerant. In this paper, the behaviour of suitable refrigerants and blended mixture of refrigerants is investigated experimentally and simulation (CYCLE_D-HX 2.0) results of software, to show the new possibilities of utilisation of different blended refrigerants in the field of refrigeration and HVAC systems to replace R134a refrigerant. The mixing and use of different blended refrigerants were started in the third generation of refrigeration in the 1990s, i.e., HFC refrigerants. Mixing of some refrigerants reduces the flammability and toxicity, it was also observed that some refrigerants have very high pressure and by mixing of refrigerants the working pressure is reduced. The normal boiling pressure is also affected and changed by the blending process resulting in a change in the cooling capacity of the system. Spauschus et al. [2] discussed the adoption of R134a as a replacement for R12 in compressor and refrigeration systems for commercial purposes to comply with the Montreal Protocol. The desired physical and chemical properties of R134a were reviewed, including its behaviour with lubricants. However, a complete assessment, including refrigerant process development, toxicological validation, lubricant development, material screening tests, and refrigeration product engineering, would be required before commercialisation. Butterworth et al. [3] tested propane and a mixture of propane and isobutene and found that the mixture can be used as a "drop-in" replacement for R12, with improved COP. Havelsky et al. [4] conducted a study to compare different refrigerants as a replacement for R12 in terms of energy efficiency, COP, and TEWI. They compared R134a, R401A, R409A, R22, and a mixture of R12 and R134a. They concluded that R134a, R401A, and R409A showed better COP results than R12 and reduced TEWI levels. Domanski et al. [5] compared the performance of R134a and CO_2 refrigerants using semi-theoretical cycle models CYCLE-11.UA and CYCLE-11.UA-CO_2 and found R134a has better COP results than CO_2. Sekhar et al. [6] investigated the replacement of CFC12 refrigerant in a household refrigerator with an eco-friendly blended refrigerant mixture of HFC134a/HC290/HC600a and found refrigerant mixture is a better replacement for CFC12. Gigiel et al. [7] performed various tests, including a pressure test, scratch test, leak test for protected circuits, leak test for unprotected circuits, leak test for external joints, and measurement of concentration with R600a. Hosoz et al. [8] compared the performance of a single stage refrigeration system and a cascade refrigeration system, both using R134a as the refrigerant. The cascade system consumed more compressor power overall due to the second compressor in the high-temperature loop. The volumetric efficiency of the single stage system was lower than the low temperature section of the cascade system. The overall COP of the cascade system was low due to the second compressor in the high-temperature loop. Gang et al. [9] conducted a performance analysis on a domestic refrigerator using different ratios of HFC152a/HFC125 refrigerant mixture and found a mass fraction of 0.85 of HFC152a with prior used refrigerant CFC12 yielded the same results. Chimres et al. [10] investigated the performance of refrigerants R290 (propane), R600 (butane), and R600a (isobutane) in a 239 L capacity refrigerator with a 53 L freezer and 100-watt power consumption and the best results were obtained with a mixture of 60% propane and 40% butane. Fatouh et al. [11] analysed and concluded that hydrocarbon refrigerants have zero ODP and GWP and the best results were obtained with a mixture of 60% propane with n-butane and iso-butane. The use of 70% propane in the mixture increased the volumetric efficiency by 15.5% and the coefficient of performance can be improved by 2.3% by using 60% propane. Ding et al. [12] conducted a comprehensive review of the use of simulation models for vapour compression refrigeration systems to optimise their design and predict performance. Jwo et al. [13] performed experiments with R134a working refrigerant replaced by R290 and R-600a hydrocarbons of 50:50 ratios and found that the refrigerating effect is improved. The total energy consumed by 4.4% also the mass of refrigerant for charging the system is reduced by 40%. Mohanraj et al. [14] conducted an experimental study using a mixture of R290 and R600a hydrocarbon refrigerants under different ambient temperatures ranging from 24 to 43 °C resulted in a reduction in power consumption of the refrigerator. Padilla et al. [15] conducted an analysis of energy balance on a household domestic refrigerator using R12 and found that the overall performance of the refrigerator using the zeotropic mixture was better than that of R12, provided that the evaporator temperature was maintained between 15 °C and −10 °C. Agarwal and Srivastava et al. [16] conducted experiments to test the performance of the system with different eco-friendly hydrocarbon refrigerants and observed that these refrigerants resulted in a desired reduction of CFC emissions. Wongwises et al. [17] compared the performance of CFC12, CFC22, and HFC134a refrigerants with alternative refrigerants HC290, HC1270, HC600, and HC600a at different ratios. Tiwari et al. [18] conducted experiments using refrigerants R404a and R134a and found R134a consumed less energy but in all working conditions, the performance of R404a was significantly better than R134a. Bolaji et al. [19] investigated the performance of HFC refrigerants R32, R134a, and R512a and found that R32 had a low COP and very high operating pressure. The performance of R152a and R134a was similar at different temperatures, but the COP of R152a was higher than both R32 and R134a. In addition, R152a had zero ODP and very low GWP. Liu et al. [20] conducted experimental investigations on two different vapour compression systems with the use of a mixed blend of R290 and R600 in the 20-cubic feet system resulted in 6% energy savings, the use of hydrocarbon blended mixtures in the 18-cubic feet system resulted in energy savings of up to 17.3%. Mishra et al. [21] performed numerical computations to analyse the thermal performance of a three-stage cascade vapour compression refrigeration system and analysis showed that R-600a refrigerant yielded the best system performance at an ultra-low temperature of −155 °C. Domanski et al. [22,23] presented the simulation software developed by the national institute of standards and technology (NIST) which imports the most accurate input parameters of thermos-physical properties of different fluids and fluid mixtures, as it has standard reference data and is validated by more than 200 countries. Ian H. Bell et al. [24] investigated the blend of the 23 best refrigerants of low GWP through simulation software REFPROP and CYCLE_D-HX to compare the results as a replacement for the R134a refrigerant. Domanski et al. [25,26] did an investigation on simulation tool CYCLE11 used for the preliminary evaluation of refrigerants and refrigerant mixtures in the vapor-compression cycle. The program is based on the Carnahan-Starling-DeSantes equation of state and assumes an isenthalpic expansion process. It includes a simple model of the compressor and considers heat exchange in the suction and liquid lines. Domanski et al. [27] developed a simulation model called CYCLE_D-HX to evaluate the transport properties and optimise heat exchange in heat exchangers. They conducted an experiment to evaluate the performance of R134a, R600a, and R-32 refrigerants and validated the data with the CYCLE_D-HX model (Figure 1). Gil et al. [28] investigated the efficiency of HFO/HCFO refrigerants in the ejector cooling cycle with three different levels of condensation and evaporation temperatures. The results showed that hydro-fluoro olefins, particularly HFO-1234zf and HFO-1234ze(E), can achieve high efficiency in the ejector cooling cycle. Adelrajafi et al. [29] developed a CSA-LSSVM model to predict the behaviour of a low GWP binary mixture of refrigerants, R-1234yf and R-1234ze(E). They compared their model's predictions to those of the PREOS and PC-SAFT models and found that their CSA-LSSVM model had better performance. Van Vu Nguyen et al. [30] investigated the performance of six refrigerants, R-1234ze(E), R-32, R-152a, R-290, R-600a, and R-1234yf, in terms of COP, operating pressure, and sensitivity of ejector geometry under different working conditions and found HFC R-152a and HFO R-1234ze(E) performed the best, R-600a was the most favourable, and R-1234yf was compatible with R-290. Emmi et al. [31] presented the configurations and monitoring data to study the behaviour of a two-stage heat pump operated with R-744 in the ejector system and secondary refrigerant R-1234ze is used in the high-temperature stage. The study aimed to find out the effective energy performance of the system. Taweekum et al. [32] analysed R-463A as an alternative to R-404 in a NIST vapour compression cycle model. Using CYCLE_D-HX software, they found that R-463A has a higher normal boiling point and a significantly lower GWP than R-404A. R-463A can also be used in high ambient temperature environments due to its higher critical pressure and temperature. At low temperatures, R-463A has a 10% higher COP than R-404A. Gil et al. [33] proposed new three-component refrigerants with a 10% step in mass fraction, using a triangular design. The researchers found that the best performing mixture was R1234yf-R152a-RE170 with a weight share of 0.1/0.5/0.4. Andreas et al. [34] studied the two-phase condensation heat transfer process and pressure drop characteristics of R-513A and found that the pressure drop of R-513A was similar to R-1234yf and 10% lower than that of R134a at higher mass flux. However, the pressure drops of R-1234ze(E) were 20% higher compared to those of R134a at higher mass flux. Bharanitharan et.al [35] conducted a study on the hydrodynamics of an oscillating Stirling regenerator at various speeds and compared the experimental and numerical results. They found that the numerical results were able to predict the flow behaviour in the regenerator, and the Ergun correlation performed well at high flow rates. Kumar et al. [36] conducted a review on low GWP refrigerants such as R-1234ze(E), R-1234ze(Z), R-1234yf, R-513A, and R-450A as substitutes for R134a. They analysed the thermodynamic and transport properties of these refrigerants using experimental, numerical, and simulation studies. Nikitin et al. [37] conducted an investigation on the performance of a heat pump on the soil at different temperatures and variable depths. They used a mixture of R-41 and R-161 to understand the effect of ice thickness and snow cover by employing computational fluid dynamics and thermo-economic–environmental analysis on the cascade system. Deyni et al. [38] developed a model to evaluate six pairs of refrigerants for use in a cascade refrigeration system. The refrigerant pairs evaluated were R41-R161, R41-R1234yf, R41-R1234ze, R744-R161, R744-R1234yf, and R744-R1234ze. The results showed that R41-R161 and R41-R1234ze had the highest COP. Nikitin et al. [39] conducted a comparative study of energy, exergy, economic, and environmental analysis of the 10 coldest Russian cities using the Pareto front curve and found Saint Petersburg would benefit from using air-source heat pump (ASHP) systems, while Khabarovsk city would benefit from using ground-source heat pump (GSHP) systems. Dashtebayaz et al. [40] studied the efficiency of five HFC refrigerants on geothermal heat pumps to optimise system design, finding R-134a had the highest efficiency and R-125 the lowest. Dashtebayaz et al. [41] studied the use of an air source heat pump as a waste heat recovery system in a data centre to reduce energy consumption and emissions. Their results demonstrated significant energy and cost savings as well as improved efficiency, with a projected payback period of 2.5 years. Honda et al. [42] conducted experiments to investigate the effects of mass velocity and condensation temperature difference on local heat transfer during R407C condensation in a horizontal microfine tube to obtain the superficial heat transfer coefficient for the vapour phase, and the combined prediction agreed with the measured values with an error of 9.2%. Rossetto et al. [43] presented a new simple model for predicting heat transfer coefficients in horizontal micro fin tubes during condensation of halogenated and natural refrigerants, validated against a data bank of 3115 experimental heat transfer coefficients. Hargovind et al. [44] used a genetic algorithm to optimize velocity and surface roughness to improve product quality by exploring the effect of pulse on time, wire span, and servo gap voltage on cutting velocity, surface roughness, recast layer, and microhardness of the surface produced. Teng et al. [45] investigated the frictional pressure drop and heat transfer performance of de-ionized water flowing through rectangular microchannels with longitudinal vortex generators (LVGs) results show that heat transfer performance was improved by 12.3–73.8% for microchannels with aspect ratios of 0.0667 and 0.25, respectively, while pressure losses increased by 40.3–158.6% and 6.5–47.7% Hsieh et al. [46] examines the spreading thermal resistance of centrally positioned heat sources and the thermal performance of a flat vapor chamber used for electronic cooling. Parametric studies were conducted, and the results showed a heat removal capacity of 220 W/cm^2 with a thermal spreading resistance of 0.2 °C/W for the vapor chamber heat spreader. The study highlights the potential of using flat vapor chambers for efficient electronic cooling. Uzair [47] conducted both experimental performance analysis and deep learning-based modeling to analyze the performance of a closed-loop heat pump dryer that uses R-134a as a secondary fluid and moist sodium polyacrylate material, also known as Orbeez, as the drying material. The study seeks to understand the behavior of the Orbeez material and its interaction with the heat pump dryer system to improve the efficiency and effectiveness of the drying process

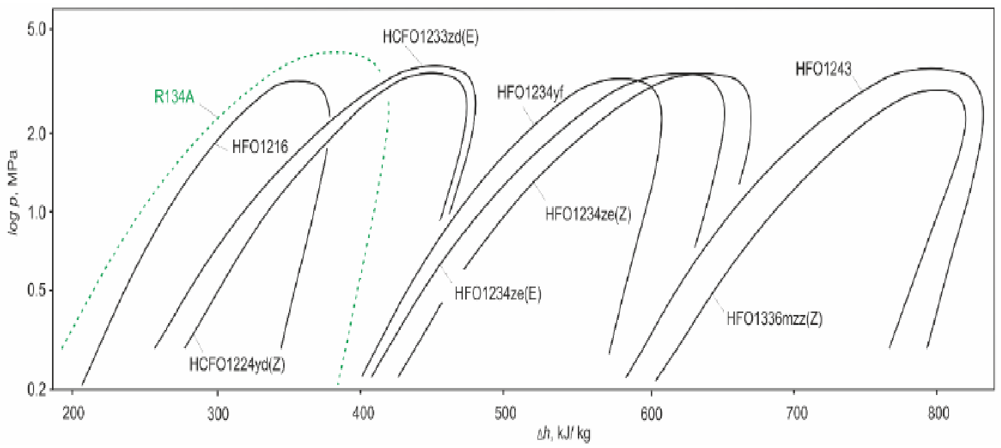


Figure 1. Saturation curves of HFO/HCFO refrigerants, collected in p-h diagram [28].

Based on the literature review, refrigerants and blends are used in the simulation investigation as an alternative to R134a and thermodynamic properties and environmental properties are given below in Tables 1–3.

Table 1. Thermodynamic and environmental properties of HFC and HFC blends (REFPROP, version 10.0 [23]).

S. No	Refrigerants/ Properties	Unit	R134a	R404A [R-125/143a/134a]	R407A [R-32/125/134a]	R32	R152a	R245fa	R227ea	RS50
1	Name of the refrigerant	-	1,1,1,2-Tetrafluoroethane			Dichloromethane	1,1-Difluoroethane	1,1,1,3,3-Pentafluoropropane	1,1,1,2,3,3,3-Heptafluoropropano	
2	Molecular Formula	-	C2H2F4	C2HF5/C2H3F3/ C2H2F4	CH2F2/C2HF5/ C2H2F4	CH2F2	C2H4F2	C3H3F5	C3HF7	CH2F2/C2HF5/CH2FCF3/ C3HF7/C2H4F2
3	Composition (weight share)	-	100	0.44/0.52/0.04	20/40/40	Dichloromethane	1,1-Difluoroethane	1,1,1,3,3-Pentafluoropropane	1,1,1,2,3,3,3-Heptafluoropropano	HFC-32 HFC-125 R134a HFC-227ea HFC-152a
4	Category (type)	-	HFC	HFC blend	HFC Blend	HFC	HFC	HFC	HFC	HFC Blend
5	GWP	-	1430	3922	2107	675	124	1030		1888
6	ODP	-	0	0	0	0	0	0	0	0
7	Critical temperature	°C	101 °C	72.12	82	78.1	113.26	153.86	101.75	82.4
8	Critical pressure	bar	13.6	37.35	44.94	57.82	45.16	36.5	29.25	47.5738
9	Normal boiling point	°C	−26.1 °C	−45.74	−45	−51.62	−24.9	15.3	−16	−46.5
10	Molar weight	g/mol	102.03	97.6	90.1	52.02	66.05	134.05	170.03	81.8

Table 2. Thermodynamic properties of HFC + HFO blends (REFPROP, version 10.0 [23]).

S. No	Refrigerants/ Properties	Unit	R32/R41/ R1234ze(E)	R134a	R161/R41/ R1234ze(E)	R448A	R449A	R449B	R449C	R450A	R452A	R452B	R454B	R454C	R515A
1	Name of the refrigerant	-		1,1,1,2-Tetrafluoro-ethane		HFC32—HFC125—HFC134a—HFO 1234ze—HFO 1234yf	HFC32—HFC125—HFC134a—HFO 1234yf	HFC32—HFC125—HFC134a—HFO 1234yf	HFC32—HFC125—HFC134a—HFO 1234yf	HFC134a—HFO 1234ze (E)	HFC32—HFC125—HFO 1234yf	HFC32—HFC125—HFO 1234yf	HFC32-HFO 1234yf	HFC32-HFO 1234yf	HFCR227ea—HFO 1234ze (E)
2	Molecular Formula	-	CH2F2/CH3F/ C3H2F4	C2H2F4	C2H5F/ CH3F/ C3H2F4	CH2F2/ C2HF5/ CH2FCF3/ C3H2F4/ C3H2F4	CH2F2/ C2HF5/ CH2FCF3/ C3H2F4	CH2F2/ C2HF5/ CH2FCF3/ C3H2F4	CH2F2/ C2HF5/ CH2FCF3/ C3H2F4	CH2FCF3/ C3H2F4	CH2F2/ C2HF5/ C3H2F4	CH2F2/ C2HF5/ C3H2F4	CH2F2/ C3H2F4	CH2F2/ C3H2F4	C3HF7/ C3H2F4
3	Composition (weight share)	-	0.1/0.9/0	100	0.8/0.1/0.1	(26/26/21/ 7/20)	24/25/26/25	25.2/24.3/ 23.2/27.3	20/20/31/29	42/58	11/59/30	67/7/26	68.9/31.1	21.5/78.5	88/12
4	Category (type)	-	HFC + HFO	HFC	HFC + HFO	HFC + HFO	HFC + HFO	HFC + HFO	HFC + HFO	HFC + HFO	HFC + HFO	HFC + HFO	HFC + HFO	HFC + HFO	HFC + HFO
5	GWP	-	608	1430	20	1387	1397	1412	1251	605	2140	698	466	148	393
6	ODP	-	0	0	0	0	0	0	0	0	0	0	0	0	0
7	Critical temperature	°C	80.58	101 °C	95.99	82.68	82.07	82.2	84.21	104.47	75.05	77.1	78.1	85.6	108.71
8	Critical pressure	bar	58.07	13.6	53.1	45.94	44.9	45.3	43.98	38.22	40.14	52.2	52.66	43.18	35.65
9	Normal boiling point	°C	−50.23	−26.1 °C	−39.77	−46	−46	−46.1	−44.6	−23.4	−47	−51	−50	−46	−18
10	Molar weight	g/mol	55.02	102	48.87	189.9	87.2	86.3	90.3	109.0	103.5	63.53	62.6	90.8	117.4

Table 3. Thermodynamic properties of HFO and PFO and HC (REFPROP, version 10.0 [23]).

S. No	Refrigerants/ Properties	Unit	R134a	R1216	R1224yd(Z)	R1233zd(E)	R1234yf	R1234ze(E)	R1234ze(Z)	R1243zf	R1336mzz(Z)	R290	R600a	RE170
1	Name of the refrigerant	-	1,1,1,2-Tetrafluoro-ethane	Hexafluoro-propylene	1-Chloro-2,3,3,3-tetrafluoro-propene	Trans-1-chloro-3,3,3-Trifluoro-propene	2,3,3,3-Tetrafluoro-propene	1,3,3,3-Tetrafluoro-propene	CIS-1, 3,3,3-Tetrafluoro-propene	3,3,3-Trifluoro-propene	1,1,1,4,4,4-Hexafluoro-2 butane	Propane	Isobutano	Dimethyl ether
2	Molecular Formula	-	C2H2F4	C3F6	(Z)-CF3-CF=CHCl	C3H2ClF3	C3H2F4	C3H2F4	C3H2F4	C3ClF3H2	cis-CF3CH=CHCF3	CH3CH2CH3	C4H10	C2H6O
3	Composition (weight share)	-	100	Hexafluoro-propylene	1-Chloro-2,3,3,3-tetrafluoro-propene	Trans-1-chloro-3,3,3-Trifluoro-propene	2,3,3,3-Tetrafluoro-propene	1,3,3,3-Tetrafluoro-propene	CIS-1, 3,3,3-Tetrafluoro-propene	3,3,3-Trifluoro-propene	1,1,1,4,4, 4-Hexafluoro-2 butane	Propane	Isobutano	Dimethyl ether
4	Category (type)	-	HFC	PFO	HFO	HFO	HFO	HFO	HFO	HFO	HFO	HC	HC	HC
5	GWP	-	1430	17,340	4	1030	1	7			2			
6	ODP	-	0	0	0	0	0	0	0	0	0	0	0	0
7	Critical temperature	°C	101	85.8	155.54	166.45	94.3	109.36	150.2	103.7	171.35	96.7	134.7	127.2
8	Critical pressure	bar	13.6	31.49	33.37	36.23	33.82	35.34	35.3	35.17	29.03	42.5	36.3	53.37
9	Normal boiling point	°C	−26.1	−29.6	15	18.31	−29.48	−19	9.8	−25.42	33.4	−42	−12	−24.78
10	Molar weight	g/mol	102.03	150.03	148.5	130.5	114	114	114.04	96.05	164	44.1	58.12	46.07

2. Experimental Apparatus and Test Conditions

Compressor, condenser, evaporator, and expansion valves are the main components of any simple vapour compression refrigeration system to sustain the cooling load, whereas some applications require different temperatures for different sections. Tests have been carried out in a controlled environment with an environment temperature of 18 °C and evaporator and condenser air-flow discharge conditions. The pressure of refrigerant in the condenser and the evaporator, the temperatures in the refrigeration loop, and the compressor power consumption data for each of the tests were recorded with a period of 10 s per measurement in the dynamic cooling process from 15 °C to −10 °C as measured as the outlet of the evaporator. The experiment started with R134a to set up the base reference, the thermodynamic properties of the refrigerants were obtained from the NIST thermodynamic properties of refrigerants and refrigerant mixtures database [23]. The vapour compression refrigeration cycle is based on the following factors:

- Refrigerant flow rate.
- Type of refrigerant used.
- Kind of application viz air-conditioning, refrigeration, dehumidification, etc.
- The operation design parameters.
- The system equipment/components proposed to be used in the system.

A single-stage vapour compression system was used to generate data to verify the model. The system was equipped with a variable speed reciprocating compressor, variably sized evaporator and condenser, manually adjusted throttling valve, and a liquid-line/suction-line heat exchanger, which could be included or bypassed The evaporator and condenser were of the annular design arranged in the counter-current configuration; the refrigerant flowed in the enhanced inner tube (copper), while the HTF flowed in the smooth annular space. The heat exchangers' size could be adjusted by changing the number of active refrigerant tubes; this feature enabled heat flux control. The apparatus was set to achieve evaporation and condensation saturation temperatures nominal to air-source heat pumps and the HTF inlet and outlet temperatures were used to obtain these evaporation and condensation temperatures using R134a and a mid-range compressor speed, 1800 rev·min^{-1}. Four additional data sets at each rating test (total of 12) were generated by holding the HTF inlet temperature constant as the system capacity was varied via compressor speed, (1400 to 2200) rev·min^{-1}; readings are mentioned in observation Table 4, and care was taken to configure other evaporator and condenser operating conditions (beyond refrigerant saturation temperature) to closely resemble those of a typical air-to-air heat pump. Specifically, the heat fluxes were within (5 to 9) kW·m^{-2} and (5 to 10) kW·m^{-2} for the evaporator and the condenser, respectively. Additionally, the ratios of HTF thermal resistance to total heat exchanger thermal resistance were nominally 0.8 and 0.6 for the evaporator and condenser, respectively; these values are representative of air-to-air heat pumps where the air side (i.e., HTF side) thermal resistance dominates. The thermal resistance ratios were enforced by the selection of HTF mass flow rates; the HTF mass flow rates were held constant for all tests at 0.098 kg·s^{-1} for the condenser and 0.131 kg·s^{-1} for the evaporator. The subcooling and superheat were controlled to (2 to 3) K and (3 to 6) K, respectively. More details about these tests, including the uncertainty calculation (95% confidence level) for the COP (0.35%), capacity (0.2%), and Q_{vol} (1.5%).

Table 4. Observation table for R134a as refrigerant.

S. No	Energy Meter Reading for 10 Rev in Sec.	Compressor Inlet Pressure, P_1 (Bar)	Compressor Outlet Pressure, P_2 (Bar)	Refrigerant Temperature at Inlet of Compressor, T_1 (°C)	Refrigerant Temperature at Outlet of Compressor, T_2 (°C)	Refrigerant Temperature at Inlet of Expansion Valve, T_3 (°C)	Refrigerant Temperature at Outlet of Expansion Valve, T_4 (°C)	Water Temperature in Evaporator, T_5 (°C)
1.	8.5	6.4	8.3	15.2	90.3	49.3	6.3	8.6
2.	9.7	6.3	8.5	16.5	91.6	51.1	5.9	7.9
3.	10.6	6.5	8.7	17.8	92.7	52.3	5.2	7.2

From the above observation table, the calculation is performed to find out the refrigerating effect and work completed by the compressor

Calculation-

$$\begin{aligned}\text{Work done by compressor (WD)} &= \frac{\text{No. of revolutions in energy meter} * 3600}{\text{Time taken in energy meter} * \text{Emc}} \\ &= \frac{10 * 3600}{85 * 750} \\ &= 0.5647\ \text{KW}\end{aligned} \tag{1}$$

In Case of mass of water in the chiller,

$$\begin{aligned}\text{Refrigerating effect(RE)} &= \frac{m_w * c_p * \Delta T}{\text{Time taken for drop in initial and final temperature}} \\ &= \frac{11 * 4.187 * 8.6}{10 * 60} \\ &= 0.660\ \text{KW}\end{aligned} \tag{2}$$

$$\begin{aligned}\text{Actual Coefficient of performance } (\text{COP}_{\text{actual}}) &= \frac{RE}{WD} \\ &= \frac{0.660}{0.5647} \\ &= 1.169\end{aligned} \tag{3}$$

For the variations, the pressure and the temperature ranges are changed,
P_1 = 6.4 Bar
P_2 = 8.3 Bar
T_1 = 24.3 °C, h_1 = 373.13 KJ/Kg
T_2 = 34.7 °C, h_2 = 425.84 KJ/Kg
T_3 = 31.7 °C, h_3 = 243.9 KJ/Kg
T_4 = 15.3 °C, h_4 = 220 KJ/Kg

$$\begin{aligned}\text{Theoretical coefficient of performance } \left(\text{COP}_{\text{theory}}\right) &= \frac{h_1 - h_4}{h_2 - h_1} \\ &= \frac{373.13 - 220}{425.84 - 373.13} \\ &= 2.90\end{aligned} \tag{4}$$

$$\text{Exergy at any point can be calculated as : } \dot{ex} = \dot{m}[(\text{h} - \text{h}_{\text{air}}) - \text{T}_{\text{air}}(\text{s} - \text{s}_{\text{air}})] \tag{5}$$

Accordingly, the total exergy destruction is sum of the exergy destruction in each of components and is written as:

$$\dot{Exd}_{tot} = \dot{Exd}_{eva} + \dot{Exd}_{con} + \dot{Exd}_{com,HTC} + \dot{Exd}_{com.LTC} + \dot{Exd}_{exp,HTC} + \dot{Exd}_{exp,LTC} \tag{6}$$

2.1. Description of CYCLE_D-HX 2.0 Model

The CYCLE_D-HX 2.0 model is a simulation tool that is used to analyse the performance of vapour compression cycles. It is based on the concept of using temperature profiles of the heat sink and heat source, and ΔT_{hx} for the evaporator and condenser. This approach enables the model to account for refrigerant thermophysical properties, pressure drop, and heat transfer coefficient on the cycle performance on a relative basis. The simulated system consists of a compressor, condenser, adiabatic expansion device, and evaporator. The compressor is represented by the isentropic efficiency, volumetric efficiency, and electric motor efficiency. The evaporator and condenser can be either counter-flow, crossflow, or parallel-flow, and are represented by their ΔT_{hx}. The solution sequence starts with estimated values of saturation temperatures in the evaporator and condenser. Based on the established thermodynamic cycle with refrigerant temperature profiles and HTF (heat transfer fluid) temperature profiles, the model calculates ΔT_{hx} and compares them to the values specified as input. The model iterates evaporator and condenser saturation temperatures until it achieves the specified ΔT_{hx} values within a convergence parameter. The CYCLE_D-HX 2.0 model is a comprehensive tool that can be used to analyse the performance of different types of vapour compression cycles. It includes enhanced cycle options such as a liquid-line/suction-line heat exchanger. The model has been extensively tested and validated against experimental data, and its accuracy and reliability have been demonstrated in several research publications.

For each iteration step of saturation temperatures, CYCLE_D-HX 2.0 calculates heat exchangers' ΔT_{hx} using Equation (7)

$$\frac{1}{\Delta T_{hx}} = \frac{Q_1}{Q_{hx}\Delta T_1} + \frac{Q_2}{Q_{hx}\Delta T_2} = \frac{1}{Q_{hx}}\sum \frac{Q_i}{\Delta T_i} \tag{7}$$

The equation presented in the statement calculates ΔT_{hx} as a harmonic mean weighted with the fraction of heat transferred in individual sections of the heat exchanger, assuming a constant overall heat transfer coefficient throughout the heat exchanger. This approach enables the model to account for the variations in heat transfer rate along the length of the heat exchanger. At the beginning of each iteration, the model calculates ΔT_{hx} based on the sections corresponding to the subcooled liquid, two-phase, and superheated regions. The model repeatedly bisects each subsection until the ΔT_{hx} obtained from two consecutive evaluations agree within a convergence parameter. This iterative process ensures that the model achieves a high degree of accuracy in calculating the performance of the vapour compression cycle. Alternatively, the heat exchangers can be characterised by the overall heat conductance UA_{hx}, which is a measure of the heat transfer rate per unit temperature difference across the heat exchanger. If this input option is used, the model calculates the specified ΔT_{hx} from the basic heat transfer relation, which relates the heat transfer rate to the temperature difference and the overall heat transfer coefficient. This approach provides an alternative method to specify the heat exchanger performance, which may be more convenient in some cases. This input option is used, the model calculates the specified ΔT_{hx} from the basic heat transfer relation. If this input option is used, the model calculates the specified ΔT_{hx} from the basic heat transfer relation,

$$\Delta T_{hx} = \frac{Q_{hx}}{UA_{hx}} \tag{8}$$

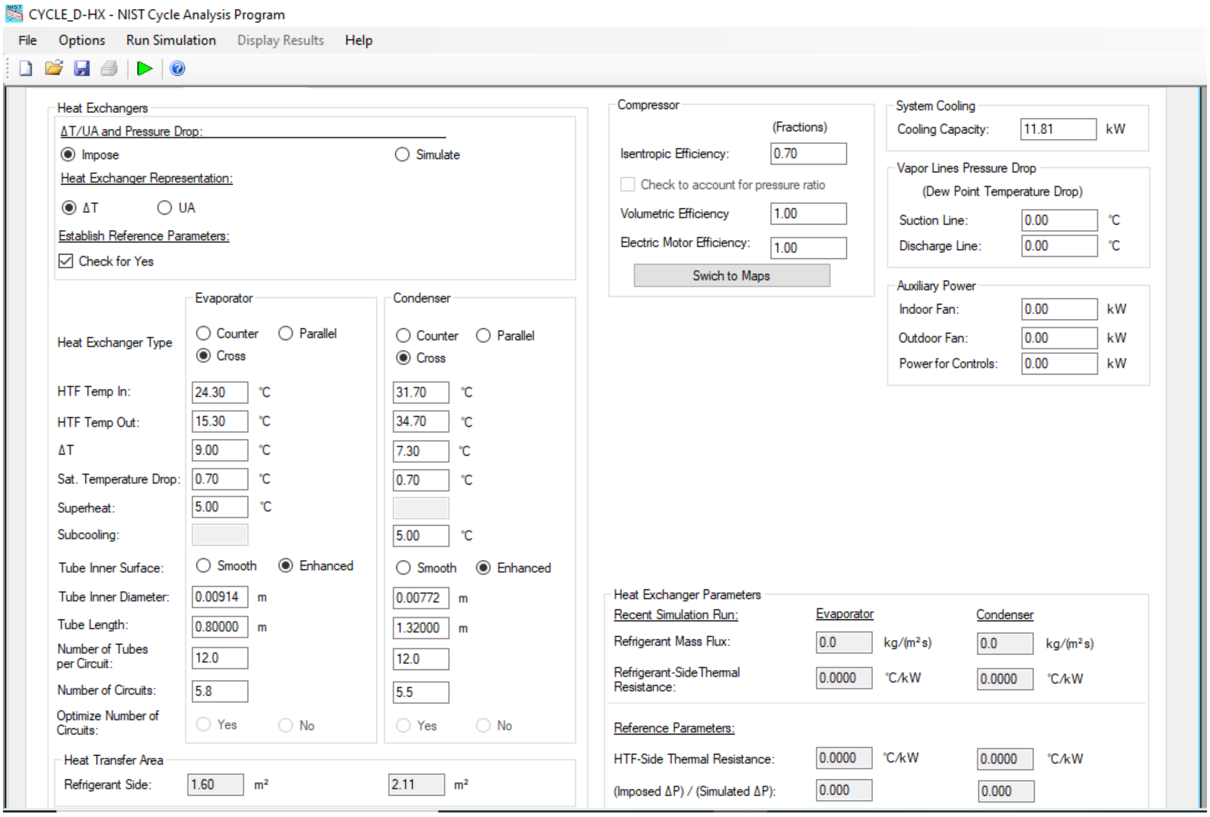
where Q_{hx} is the product of refrigerant mass flow rate and enthalpy change in the evaporator or condenser, as appropriate. Representation of heat exchangers by their UA_{hx} allows for the inclusion of heat transfer and pressure drop characteristics in comparative evaluations of different refrigerants. For this purpose, CYCLE_D-HX 2.0 considers that the total resistance to heat transfer in a heat exchanger, R_{hx}, consists of the resistance on the refrigerant side R_r, and combined resistances of the heat exchanger material and HTF $[R_{tube} + R_{HTF}]$:

$$R_{hx} = \frac{1}{UA_{hx}} = R_r + [R_{tube} + R_{HTF}] \tag{9}$$

$$\text{where, } R_r = \frac{1}{(h_r \, * \, A_{hx})} \tag{10}$$

The resistances $[R_{tube} + R_{HTF}]$ are independent of the refrigerant and are assumed to be independent of operating conditions. Their combined value can be calculated from UA_{hx} and hr values using performance measurements obtained in a laboratory on a system of interest CYCLE_D-HX 2.0calculates $[R_{tube} + R_{HTF}]$ within its "reference run" and stores its value for use in subsequent simulation runs for the calculation of UA_{hx} characterising the heat exchanger with a new refrigerant or operating conditions. CYCLE_D-HX 2.0 requires the following operational input data for the "reference run": Figure 2. Then, several inputs to simulate the performance of the vapour compression cycle. These inputs include the HTF inlet and outlet temperatures for the evaporator and condenser, ΔT_{hx} for the evaporator and condenser to achieve the desired measured evaporator and condenser saturation temperatures, evaporator superheat and pressure drop, and condenser subcooling and pressure drop. The "reference run" inputs include compressor isentropic and volumetric efficiencies and electric motor efficiency. The isentropic efficiency of the compressor can be dependent on the compression ratio, and CYCLE_D-HX 2.0 offers the option of accounting for this dependence. When screening different refrigerants, the model uses a set of thermophysical properties and correlations to simulate their behaviour in the vapour compression cycle. The accuracy and reliability of the model's predictions depend on the quality of the input data and the assumptions made in the simulation. Therefore, it is important to validate the model against experimental data and adjust the inputs and assumptions accordingly. Equation (11) takes into account the change in isentropic efficiency with the pressure ratio in a consistent way.

$$\eta_s = C - 0.05\theta \tag{11}$$



Heat Exchangers

ΔT/UA and Pressure Drop:
● Impose ○ Simulate

Heat Exchanger Representation:
● ΔT ○ UA

Establish Reference Parameters:
☑ Check for Yes

	Evaporator		Condenser	
Heat Exchanger Type	○ Counter ○ Parallel ● Cross		○ Counter ○ Parallel ● Cross	
HTF Temp In:	24.30	°C	31.70	°C
HTF Temp Out:	15.30	°C	34.70	°C
ΔT	9.00	°C	7.30	°C
Sat. Temperature Drop:	0.70	°C	0.70	°C
Superheat:	5.00	°C		
Subcooling:			5.00	°C
Tube Inner Surface:	○ Smooth ● Enhanced		○ Smooth ● Enhanced	
Tube Inner Diameter:	0.00914	m	0.00772	m
Tube Length:	0.80000	m	1.32000	m
Number of Tubes per Circuit:	12.0		12.0	
Number of Circuits:	5.8		5.5	
Optimize Number of Circuits:	○ Yes ○ No		○ Yes ○ No	

Heat Transfer Area
Refrigerant Side: 1.60 m² 2.11 m²

Compressor
(Fractions)
Isentropic Efficiency: 0.70
□ Check to account for pressure ratio
Volumetric Efficiency 1.00
Electric Motor Efficiency: 1.00

System Cooling
Cooling Capacity: 11.81 kW

Vapor Lines Pressure Drop
(Dew Point Temperature Drop)
Suction Line: 0.00 °C
Discharge Line: 0.00 °C

Auxiliary Power
Indoor Fan: 0.00 kW
Outdoor Fan: 0.00 kW
Power for Controls: 0.00 kW

Heat Exchanger Parameters

Recent Simulation Run:	Evaporator		Condenser	
Refrigerant Mass Flux:	0.0	kg/(m² s)	0.0	kg/(m² s)
Refrigerant-Side Thermal Resistance:	0.0000	°C/kW	0.0000	°C/kW

Reference Parameters:

HTF-Side Thermal Resistance:	0.0000	°C/kW	0.0000	°C/kW
(Imposed ΔP) / (Simulated ΔP):	0.000		0.000	

Figure 2. Input parameters (CYCLE_D_HX 2.0).

C is a constant calculated within the "reference run" using the isentropic efficiency.

2.2. CYCLE_D-HX 2.0 Simulation Model Validation

We used the data from the cooling of the R134a test to carry out the CYCLE_D-HX 2.0 "reference run". The "reference run" inputs included the 11.81 kW capacity, the evaporator ΔT_{hx} = 9 °C, the condenser ΔT_{hx} = 7.3 °C, and pressure drops of 33 kPa and 45 kPa for the condenser and evaporator, respectively. We then executed simulations of the remaining Cooling A, Cooling B, and Heating H1 rating tests. The capacities, compressor isentropic and volumetric efficiencies, superheat and subcooling, discharge and suction line pressure drops, and HTF inlet and outlet temperatures were input based on measurements from each test. We evaluated the percentage deviation between the simulation and the experimental results using Equation (12).

$$E = \frac{\Pi\text{Experimental} - \Pi\text{Simulation}}{\Pi\text{Experimental}} \cdot 100\% \tag{12}$$

where Π is any parameter of interest.

The deviations for COP, Q_{vol}, p_{evap}, and p_{cond}. Most of the deviations are within 4%. The largest deviation (7.4%) is for the Cooling B test at the highest (2200 rev·min^{-1})

compressor speed; this operating condition yielded about a 20% increase in refrigerant mass flow rate and capacity over the “reference run”.

2.3. Exergy Analysis

Exergy analysis is a powerful tool used to evaluate the performance of a refrigeration system. It allows us to identify the sources of irreversibility and inefficiencies in the system and helps us to determine where improvements can be made. The vapour compression refrigeration system (VCRS) is a common refrigeration system used in many applications. It consists of a compressor, condenser, expansion valve, and evaporator. The refrigerant circulates through these components and undergoes phase changes as it absorbs and releases heat, resulting in the cooling of the desired space or product. To conduct an exergy analysis of a VCRS, we can use the following steps-

1. Define the system boundary and identify the components within the boundary. This would typically include the compressor, condenser, expansion valve, and evaporator.
2. Calculate the thermodynamic properties of the refrigerant at various points in the system, such as the temperature, pressure, and specific enthalpy. This can be completed using thermodynamic tables or software.
3. Calculate the exergy at each component and at each state point using the following equation-

$$\text{Exergy} = \left(\text{enthalpy} - \text{enthalpy}_{\text{ref}}\right) - \text{T}_{\text{ref}}\left(\text{entropy} - \text{entropy}_{\text{ref}}\right) \tag{13}$$

where enthalpy and entropy are the thermodynamic properties of the refrigerant, enthalpy$_{\text{ref}}$ and entropy$_{\text{ref}}$ are the thermodynamic properties of the refrigerant at a reference state (typically the dead state or environment), and T$_{\text{ref}}$ is the reference temperature.

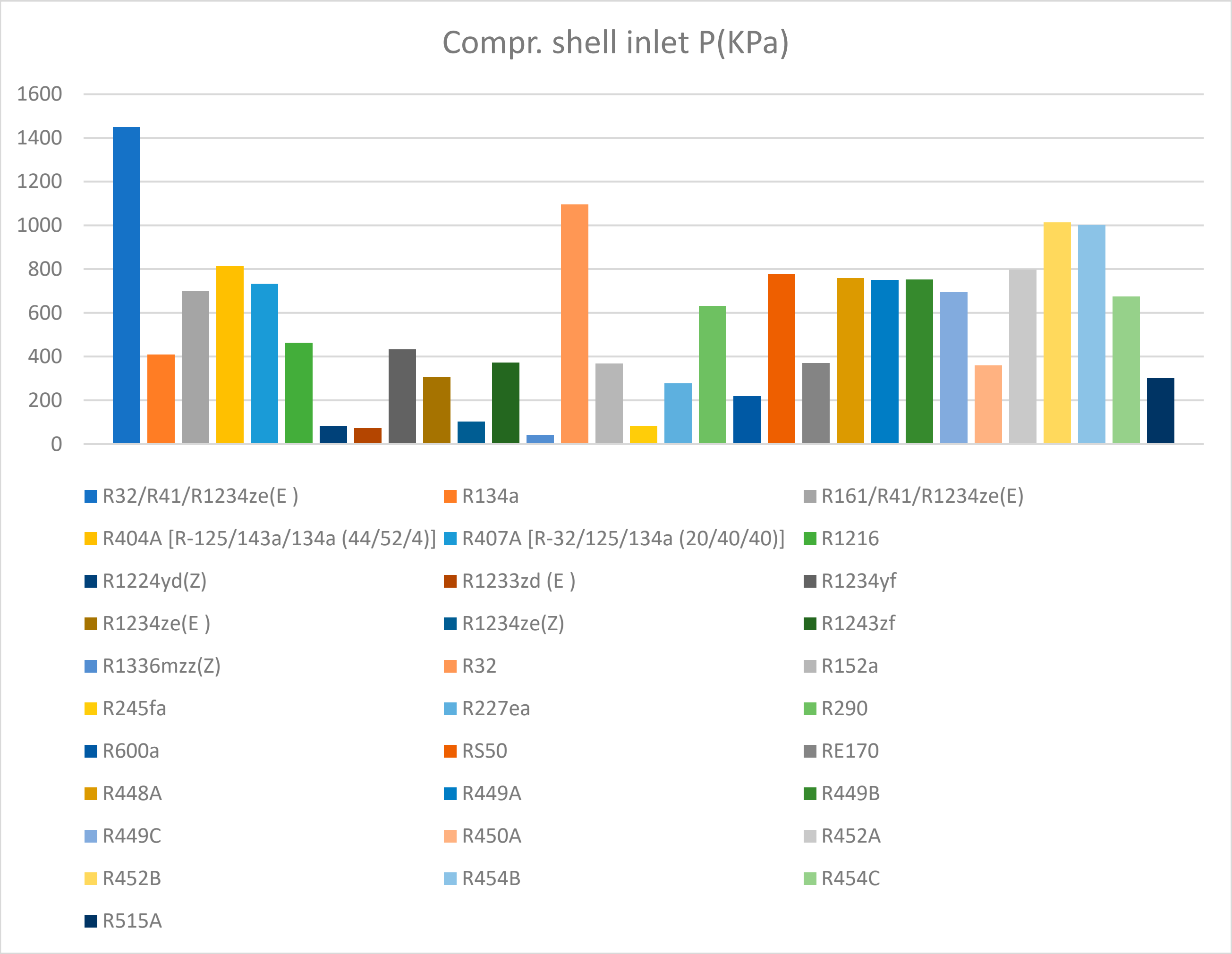
1. Calculate the exergy destruction at each component by taking the difference between the exergy input and output. This represents the amount of exergy lost due to irreversibilities and inefficiencies in the component.
2. Calculate the overall system exergy efficiency, which is the ratio of the exergy output to the exergy input. This represents the percentage of the available exergy that is being used to perform useful work.

By conducting an exergy analysis of a VCRS, we can identify the components and processes that are contributing the most to exergy destruction and inefficiencies. This can help us to make improvements to the system design, such as using more efficient components, optimising operating conditions or implementing waste heat recovery systems to reduce the amount of exergy lost to the environment.

3. Results and Discussion

The above calculations are validated with simulation software CYCLE_D-HX 2.0-NIST cycle analysis program for investigation over different refrigerants the input parameters are taken, and three groups of refrigerants are made for result analysis of variations in different parameters that can affect the system.

Scheme 1 shows pressure variation for 31 refrigerants at the compressor shell inlet. A blended refrigerant mixture of R32/R41/R1234ze(E) has a maximum pressure of 1448.6 KPa, and R1336mzz(Z) has a minimum pressure of 38.7 KPa. Inlet pressure affects refrigerating effect and efficiency, Low inlet pressure reduces refrigerant density and power consumption and is desirable for vapour compression refrigeration systems. Lower pressure can be achieved by installing fouled inlet filters or changing barometric pressure. Reduced weight flow at the inlet decreases power consumption or horsepower. Pressure at the compressor shell inlet for R134a is recorded as 409.5 KPa. The highest point exergy of 559.33 J is recorded by R290.



Scheme 1. Comparison of other refrigerants on compressor shell inlet pressure with R134a.

Figure 3 displays the results of the enthalpy and pressure variation for different refrigerants. The maximum pressure and enthalpy are obtained for R-32, which has a pressure of 1095.1 KPa and an enthalpy of 523.2 kJ/kg. In addition, the HFC blended refrigerant and HFO blended refrigerant R32/R41/R1234ze(E) show the maximum pressure and enthalpy of 1448.6 KPa and 481.3 kJ/kg, respectively. Furthermore, the HC refrigerant R290 exhibits a maximum pressure and enthalpy of 629.9 KPa and 594.4 kJ/kg at the compressor shell inlet temperature of 15 °C.

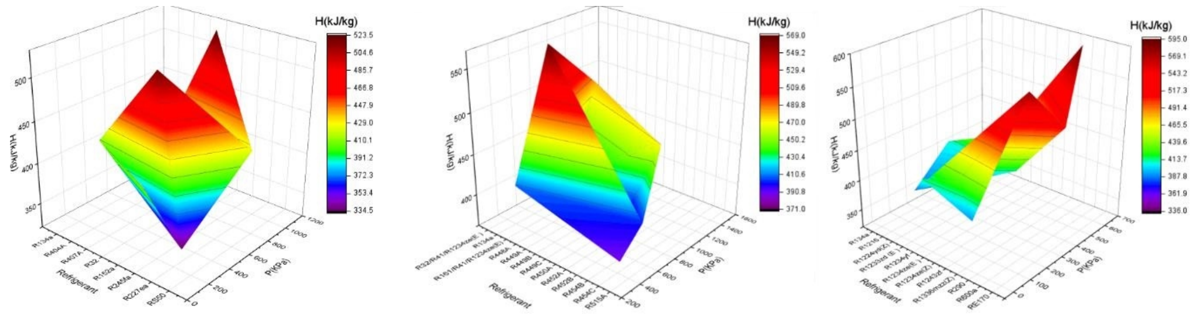
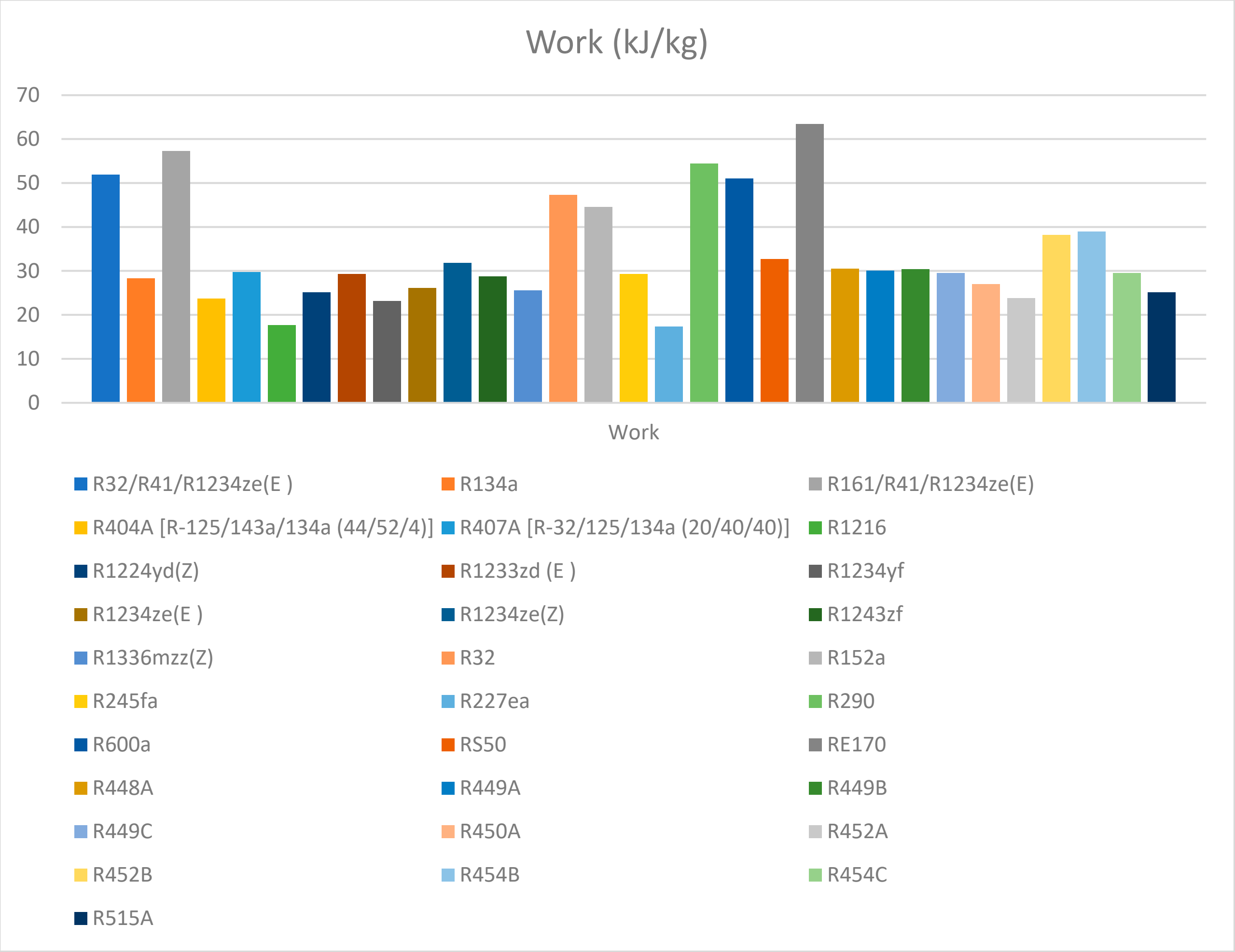


Figure 3. Comparison of other refrigerants on compressor shell inlet pressure vs. enthalpy with R134a.

Scheme 2 illustrates the variation of work completed by the compressor for different refrigerants and refrigerant mixtures in the system. The investigation revealed that R1216 and R227ea refrigerants require the least amount of work in the compressor, at 17.64 kJ/kg and 17.37 kJ/kg, respectively. On the other hand, refrigerant RE170 demands a high amount of work at 63.43 kJ/kg in the compressor. Compressors play a crucial role in the system as they receive low-pressure refrigerant from the evaporator and compress it into the high-pressure refrigerant. The efficiency of the system is highly dependent on the work completed by the compressor, and the lower work completed by the compressor indicates a higher level of efficiency.



Scheme 2. Comparison between eco-friendly refrigerants with R134a.

Figure 4 displays the relationship between the work completed by the compressor and heat supplied in the evaporator for different refrigerants at the same input conditions. The results show that R-32, HFO blended refrigerant R161/R141/R1234ze(E), and HC refrigerant RE170 have the maximum heat supplied in the evaporator and work completed values of 259.3 kJ/kg and 47.34 kJ/kg, 299.43 kJ/kg and 57.22 kJ/kg, and 368.75 kJ/kg and 63.43 kJ/kg, respectively, at the same temperature of 15°C in the evaporator unit. This information is critical in selecting the most efficient refrigerant for a particular refrigeration system.

Scheme 3 depicts the heat transfer variation in the evaporator unit for different refrigerants and refrigerant mixtures. The investigation revealed that the maximum heat transfer was achieved by RE170 at 368.75 kJ/kg, which is 23% higher than that of R134a. On the other hand, refrigerants R1216 and R227ea recorded the lowest heat transfer at 93 and 93.55 kJ/kg, respectively. A higher value of heat transfer in the evaporator is desirable as it indicates faster heat transfer. Highest point exergy of 279.31 J is recorded by R290.

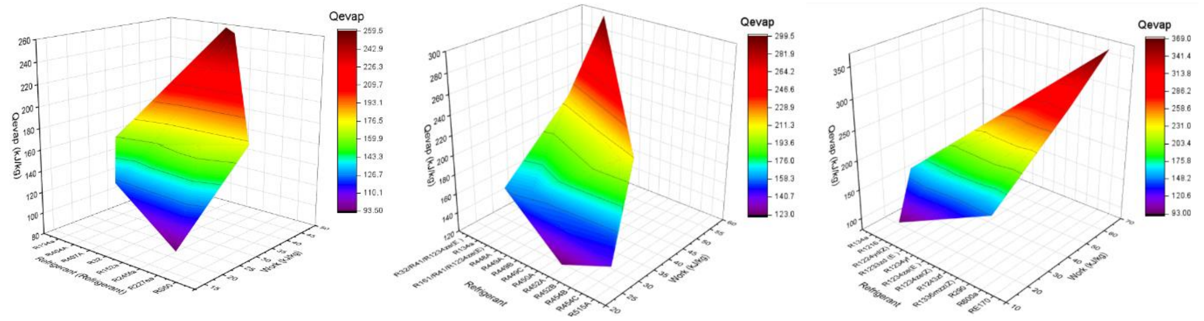
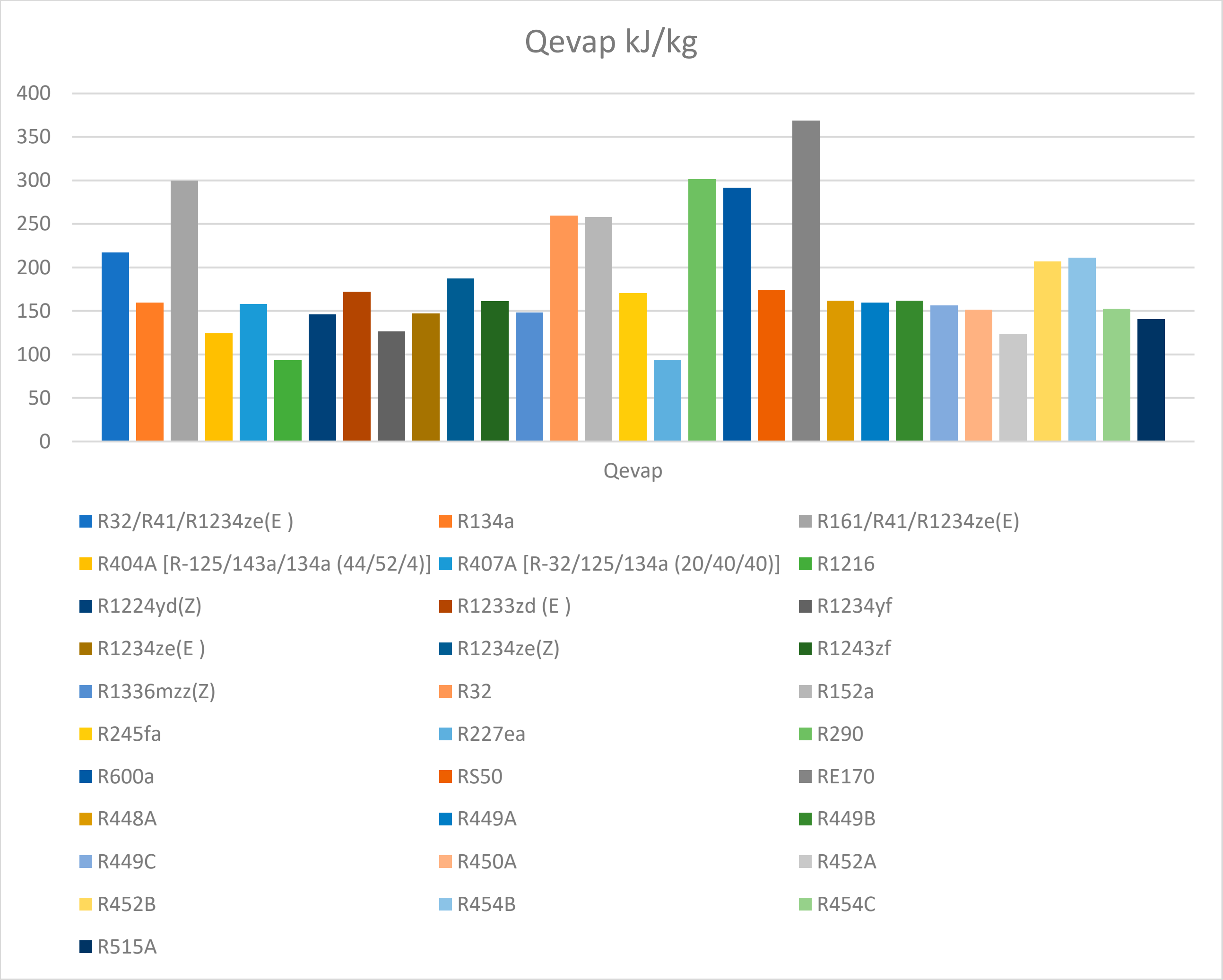


Figure 4. Comparison of refrigerants between work completed and Q_{evap} on with R134a.



Scheme 3. Comparison between eco-friendly refrigerants with R134a.

Figure 5 illustrates the relationship between specific volume and pressure in the evaporator unit for different refrigerants and refrigerant mixtures. The results show that R32 has the highest pressure and enthalpy, with maximum values of 1118.1 KPa and 263.8 kJ/kg, respectively. The blended refrigerant mixture of R32/R41/R1234ze(E) recorded the highest pressure of 1479.8 KPa and the highest enthalpy of 269 kJ/kg, while R161/R41/R1234ze(E) recorded the maximum enthalpy of 269 kJ/kg. Refrigerants R290 and R134a had the highest pressure and enthalpy of 642.5 KPa and 249.1 kJ/kg, respectively. Overall, the graph provides insights into the specific volume–pressure relationship of different refrigerants and their mixtures in the evaporator unit.

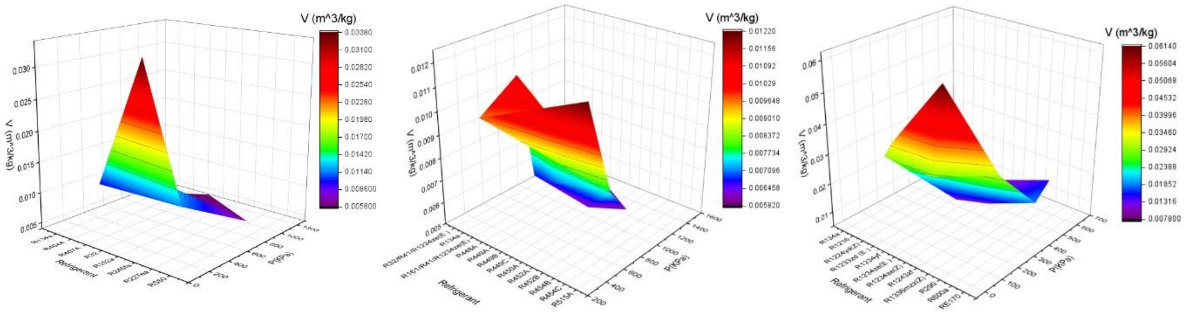
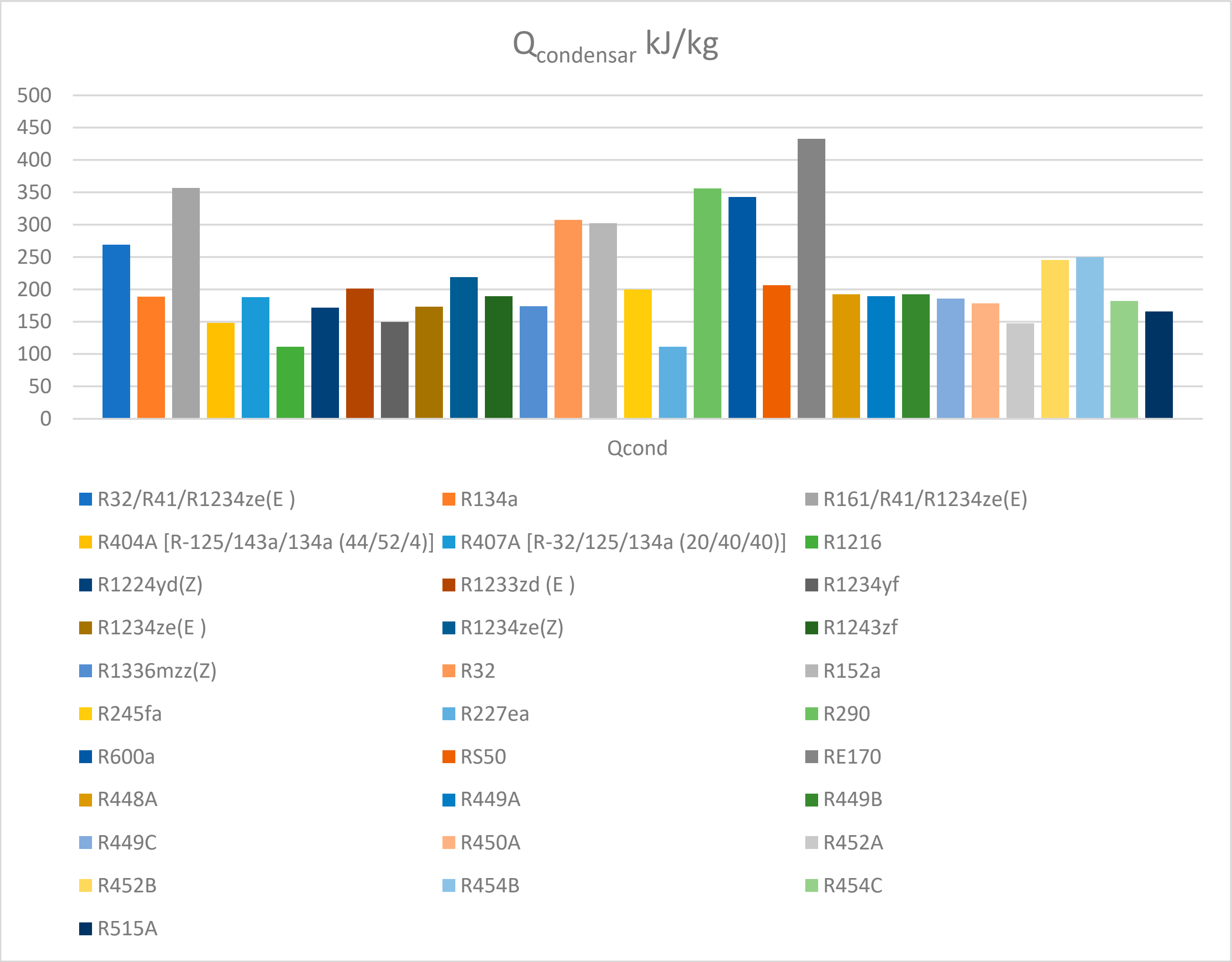


Figure 5. Comparison of refrigerants between inlet pressure vs. specific volume with R134a.

Scheme 4 displays the heat transfer rate variation for different refrigerants in the condenser unit. A high heat transfer rate is desirable to convert high-pressure vapour refrigerant into a high-pressure liquid. The highest rate is recorded for RE170 at 432.18 KJ/Kg, which is 23.5% higher than R134a. A good refrigerant should have a high heat transfer rate during condensation for efficient cooling. The highest point exergy of 511.72 J is recorded by R290.



Scheme 4. Comparison between eco-friendly refrigerants with R134a.

Figure 6 presents the results of the variation in enthalpy to pressure in the condenser unit for different refrigerants and refrigerant mixtures at the same input conditions. The graph shows that the maximum pressure of 1882.7 KPa is recorded for HFC blend RS50, while the highest enthalpy of 531.6 kJ/kg is recorded for HFC refrigerant R152a. The blended HFC and HFO refrigerant of R452b recorded a maximum pressure of 2309.1 KPa, and a maximum enthalpy of 576.3 kJ/kg is recorded for R161/R41/R1234ze(E). Furthermore, refrigerant R290 recorded the highest pressure and enthalpy of 1394 KPa and 614.8 kJ/kg, respectively.

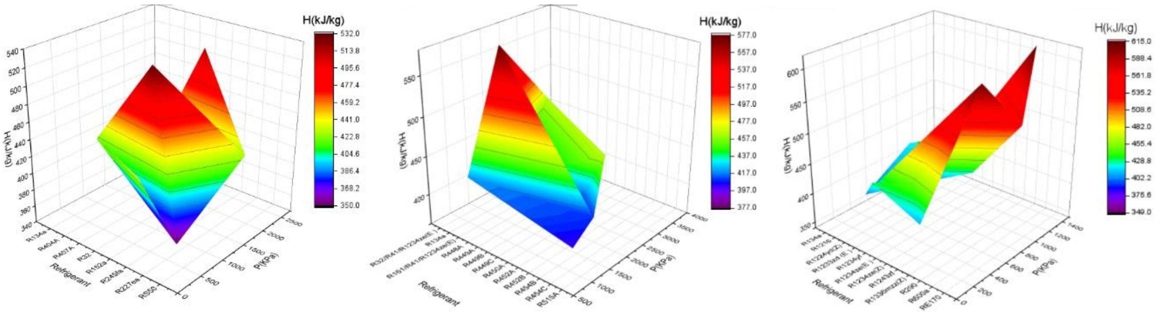
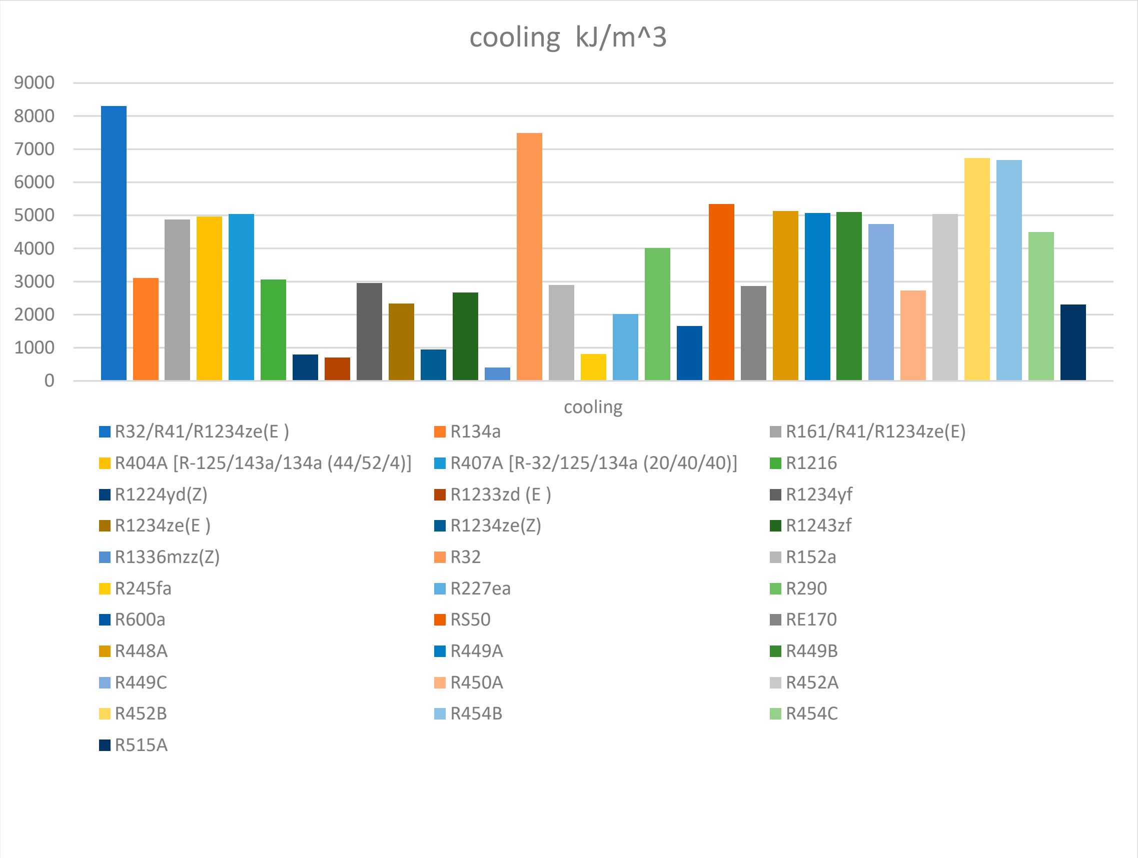


Figure 6. Comparison of refrigerants in condenser between pressure vs. enthalpy with R134a.

Scheme 5 displays the cooling rate of various refrigerants and refrigerant blends in the system. A higher cooling rate is desirable for a specific volume, and the investigation found that a mixture of refrigerants R32/R41/R1234ze(E) provides the highest cooling rate of 8290.5 kJ/m^3, which is 27% higher than the refrigerant R134a. The refrigerant R1336mzz(z) recorded the lowest cooling rate at 403.4 kJ/m^3.



Scheme 5. Comparison of the cooling capacity of other refrigerants with R134a.

Figure 7 illustrates the variations in cooling with the same coefficient of performance (COPc) for different refrigerants and blends. The results of the investigation show that the highest rate of cooling is achieved by HFC refrigerant R32 at 7478.5 kJ/m^3, while the lowest rate of cooling is recorded by HFC refrigerant R245fa at 802 kJ/m^3. For blended refrigerants of HFC and HFO, the highest rate of cooling is recorded by R32/R41/R1234ze(E) at 8290.5 kJ/m^3, and the lowest is recorded by R515A at 2303 kJ/m^3. In the third category of refrigerants from PFO, HC, and HFO, the highest rate of cooling is recorded by R290 at 4000.8 kJ/m^3, and the lowest is recorded by R1336mzz(z) at 403 kJ/m^3.

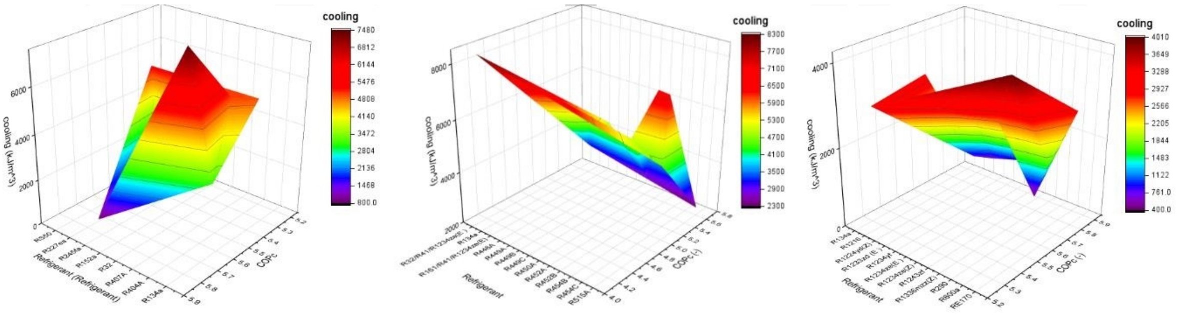
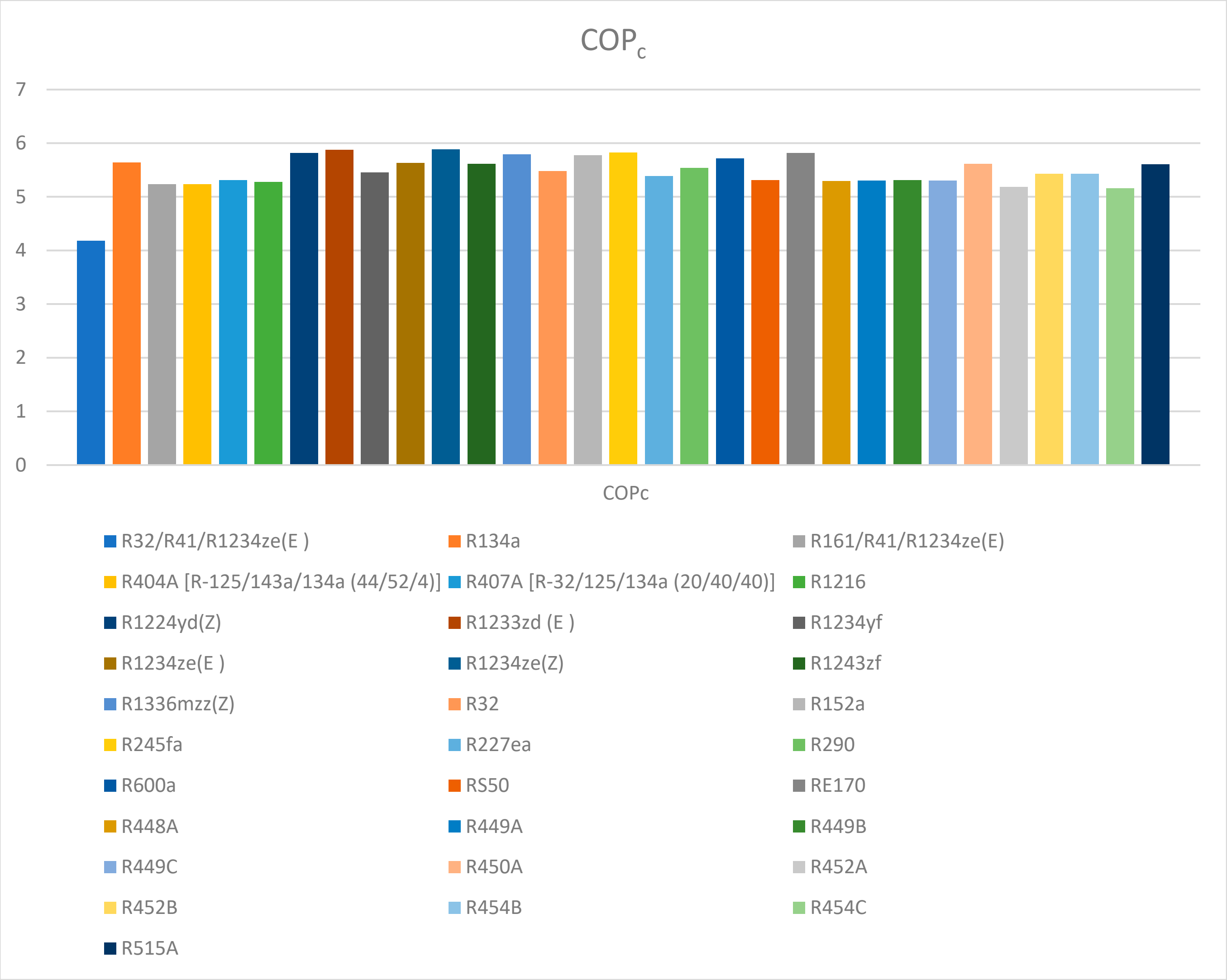


Figure 7. Comparison of refrigerants between cooling effect vs. COPc with R134a.

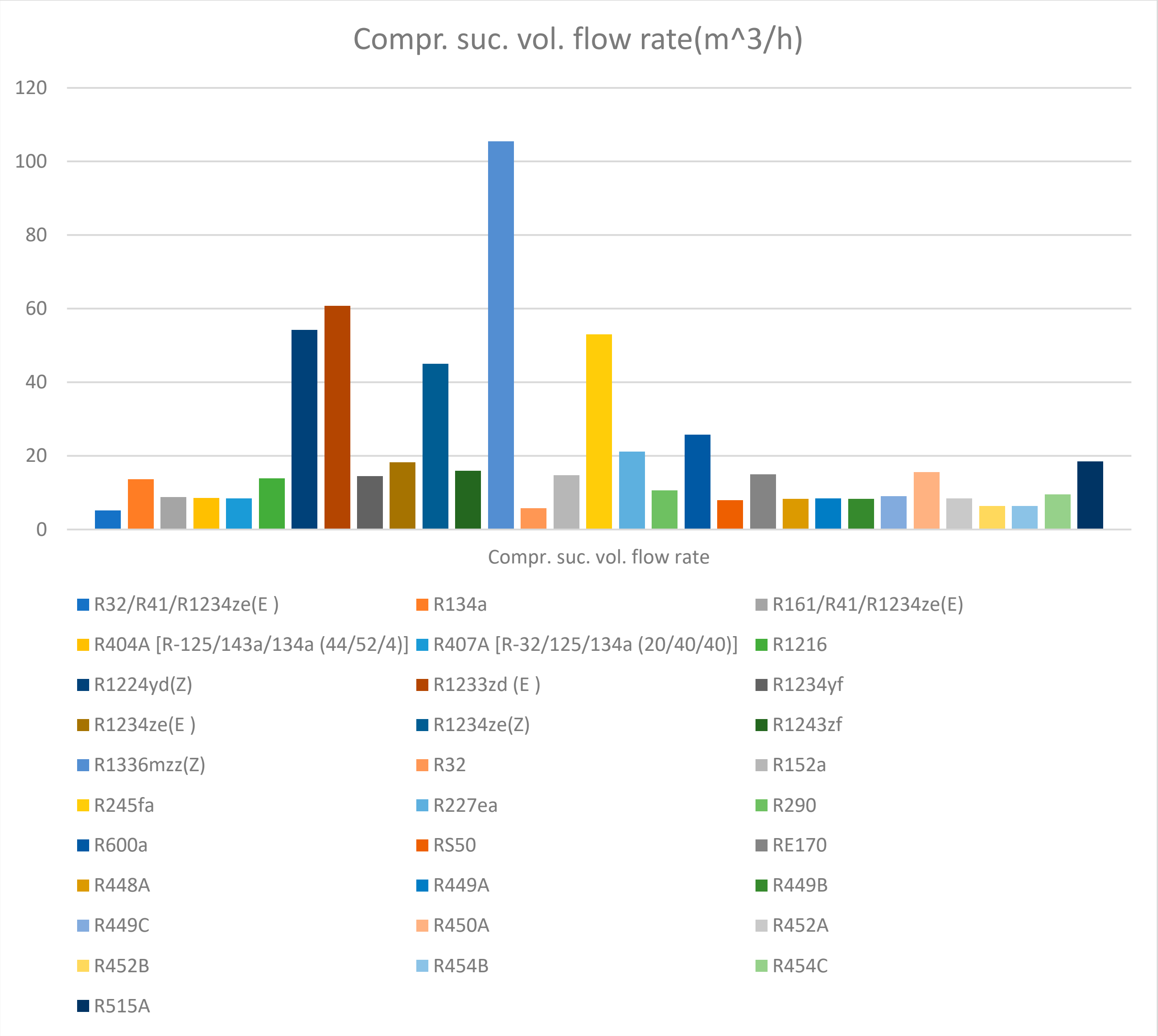
Scheme 6 shows the coefficient of performance (COP$_c$) of different pure refrigerants and blended refrigerants in the cooling system. It is observed that pure refrigerants such as RE170, R245fa, R1234ze, R1233zd(E), and R1224yd(Z) have a higher efficiency of 5.8, which is slightly higher than the efficiency of R134a. The lowest coefficient of performance (COP$_c$) is recorded as 4.178 for the blend of R32/R41/R1234ze(E). The other selected refrigerants and blends have almost the same performance, and they can be easily used as an alternative to R134a.

Scheme 7 depicts the variation in volume flow rate during the compression process for different refrigerants and blended refrigerants. R1336mzz(Z) is found to have the highest volume flow rate of 105.3 m^3/h, which is 80% higher than R134a. On the other hand, the blended refrigerant R32/R41/R1234ze(E) has the least value of 5.128 m^3/h in the graph. The volume flow rate affects the speed of the compressor, the amount of refrigerant, and the refrigerant flow through the evaporator. A thorough investigation revealed that higher flow rates lead to a better distribution of refrigerant in the system. As the refrigerant charge increases, the temperature decreases in the compressor. Consequently, the load on the compressor decreases and the discharge temperature of the refrigerant increases.

Figure 8 represents the variation in compressor suction volume flow rate for different refrigerants and blends at the same compressor power of 2 kW. The graph shows that HFC refrigerant R245fa has the highest compressor suction volume flow rate of 53.013 m^3/h while HFC refrigerant R32 has the lowest compressor suction volume flow rate of 5.685 m^3/h. For blended refrigerants of HFC and HFO, HFC refrigerant R515a has the highest compressor suction volume flow rate of 18.5 m^3/h while the blend of HFC and HFO refrigerants R32/R41/R1234ze(E) has the lowest compressor suction volume flow rate of 5.1 m^3/h. Among HFO refrigerants, R1336mzz(Z) has the highest compressor suction volume flow rate of 105.39 m^3/h and for HC refrigerant, R290 has the lowest compressor suction volume flow rate of 10.6 m^3/h. Compressor suction volume flow rate affects the refrigerant flow through the evaporator and higher flow rates result in better distribution of refrigerant in the system.



Scheme 6. Comparison between eco-friendly refrigerants with R134a.



Scheme 7. Comparison between eco-friendly refrigerants with R134a.

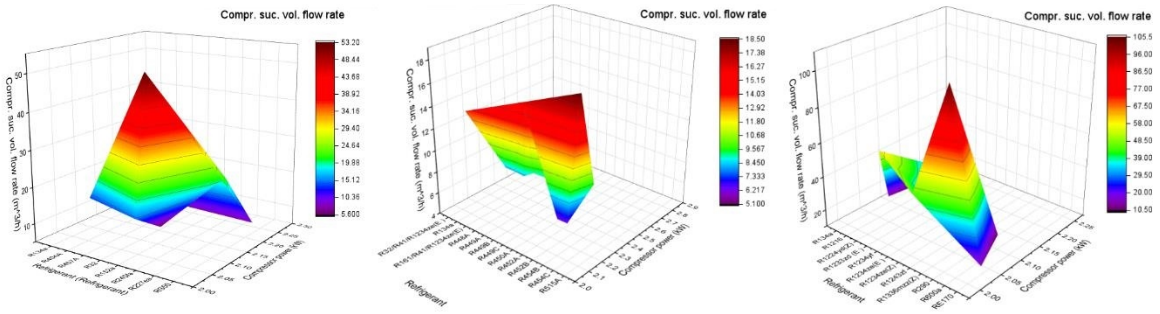


Figure 8. Comparison of refrigerants between power vs. volume flow rate with R134a.

4. Conclusions

On the basis of experimental and simulation investigations, the following conclusions are made

- Maximum pressure at the compressor is recorded by the blended refrigerant mixture of R32/R41/R1234ze(E) at 1448.6 kPa and the minimum value is recorded by R1336mzz(Z) at 38.7 KPa.
- Refrigerants R1216 and R227ea consume minimum work of 17.64 kJ/kg and 17.37 kJ/kg, respectively, and refrigerant RE170 requires a high amount of work of 63.43 kJ/kg in the compressor.
- The maximum rate of heat transfer in the evaporator is recorded by RE170 as 368.75 kJ/kg which is 23% higher than R134a and the least value is recorded by refrigerants R1216 and R227ea at 93 and 93.55 kJ/kg, respectively.
- The highest rate of heat transfer in the condenser unit is obtained by RE170 at 432.18 kJ/kg which is around 23.5% higher than R134a.
- R32/R41/R1234ze(E) recorded the highest rate of cooling of 8290.5 kJ/m3 recorded and it is 27% higher than the refrigerant R134a.
- Pure refrigerants RE170, R245fa, R1234ze, R1233zd(E), and R1224yd(Z) have higher efficiency of 5.8 which is slightly higher than the efficiency of R134a.
- A higher compressor suction volume flow rate is attained by R1336mzz(Z) of 105.3 m^3/h it is 80% higher than the R134a and blended refrigerant R32/R41/R1234ze(E) is noted as the least value of 5.128 m^3/h.

These findings highlight the significant differences between the various refrigerants in terms of their thermodynamic properties and performance. These results have important implications for the selection of refrigerants in various cooling applications, and the study provides valuable insights for researchers and practitioners in the field of refrigeration and air conditioning. Furthermore, the manuscript highlights the importance of the work in addressing the environmental concerns associated with traditional refrigerants such as R134a. As the global focus on reducing greenhouse gas emissions intensifies, the use of new and more environmentally friendly refrigerants is becoming increasingly important. The study provides an important contribution to this effort, and its findings can be used to inform the development of policies and regulations aimed at promoting the adoption of environmentally friendly refrigerants in industrial refrigeration applications.

5. Problems in the System due to Blending or Mixing

During the mixing of two or more refrigerants, each other following problems occur-

- During the running condition of the system, the effectiveness of the system is reduced due to the phase change in refrigerant in the condenser and evaporator unit as the properties of the blended refrigerant change.
- Due to uncertain non-isothermal behaviour and a mixture of refrigerants, the manufacturers are unable to design and select the appropriate component for the system from the catalogue.
- Only specific heat exchangers such as a flat plate and counter flow, concentric tube, shell, and tube heat exchangers perform well due to their geometry.
- Components of refrigeration systems are designed for pure refrigerants; therefore, these designs are not suitable for blended refrigerants.
- It is noted that blended refrigerants can reduce the temperature difference in the heat exchangers due to their non-linearity results as a bigger size of heat exchanger is required.
- Due to the blending of refrigerants the temperature, pressure capacity, and efficiency of the system are changed.

Additional components such as an accumulator and receiver must be added to the circuit for the smooth running of the system as mixed refrigerants can create the problem of choking.

Author Contributions: U.S.P. performed the idea development, methodology, and analysis and wrote down the original document. R.S.M. inspected the study and writing for the draft manuscript. R.K.D. corrected the documents for final submission. H.S. prepared the formatting and documentation of

the paper. All authors have understood and agreed to the published edition of the document. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data available on request due to restrictions eg privacy or ethical. The data presented in this study are available on request from the corresponding author. The data are not publicly available as research is continue.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

The following nomenclature is used in the manuscript.

GWP	Global warming potential
ODP	Ozone depletion potential
HFO	Hydro fluoro-olefin
HC	Hydrocarbon
HCFC	Hydro-chlorofluorocarbon
CFC	Chlorofluorocarbon
TEWI	Total equivalent warming impact
COP	Coefficient of performance
LLSL-HX	Liquid-line/suction-line heat exchanger
P_1	Compressor inlet pressure (bar)
P_2	Compressor outlet pressure (bar)
T_1	The refrigerant temperature at the inlet of the compressor (°C)
T_2	The refrigerant temperature at the outlet of the compressor (°C)
T_3	The refrigerant temperature at the inlet of the expansion valve (°C)
T_4	The refrigerant temperature at the outlet of the expansion valve (°C)
T_5	Water temperature in evaporator(°C)
h_1	Enthalpy at the inlet of compressor (KJ/Kg)
h_2	Enthalpy at the outlet of compressor (KJ/Kg)
h_3	Enthalpy at the inlet of expansion valve (KJ/Kg)
h_4	Enthalpy at the outlet of expansion valve (KJ/Kg)
Emc	Energy meter constant
WD	Work done (KW)
RE	Refrigerating effect (KW)
m_w	Mass of water in the evaporator unit (litre)
c_p	Specific heat at constant pressure
ΔT	Temperature difference (°C)
COP_{actual}	Actual coefficient of performance
COP_{theory}	Theoretical coefficient of performance

References

1. Maxwell, J.C. *Treatise on Electricity and Magnetism*; Clarendon Press: Oxford, UK, 1873.
2. Spaushus, H.O. CF-134a as a substitute refrigerant for CFC 12. *Int. J. Refrig.* **1988**, *11*, 389–392. Available online: https://docs.lib.purdue.edu/iracc/70/ (accessed on 12 April 2022). [CrossRef]
3. Richardson, R.N.; Butterworth, J.S. The performance of propane/isobutene mixtures in vapour-compression refrigerant system. *Int. J. Refrig.* **1995**, *18*, 58–62. [CrossRef]
4. Havelsky, V. Investigation of refrigerating system with R12 refrigerant replacements. *Appl. Therm. Eng.* **2000**, *20*, 133–140. [CrossRef]
5. Brown, J.S.; Yana-Motta, S.F.; Domanski, P.A. Comparative analysis of an automotive air conditioning systems operating with CO_2 and R134a. *Int. J. Refrig.* **2002**, *25*, 19–32. [CrossRef]
6. Sekhar, S.; Lal, D.; Renganarayanan, S. Improved energy efficiency for CFC domestic refrigerators retrofitted with ozone friendly HFC134a. *Int. J. Therm. Sci.* **2004**, *43*, 307–314. [CrossRef]
7. Gigiel, A. Safety testing of domestic refrigerators using flammable refrigerants. *Int. J. Refrig.* **2004**, *27*, 621–628. [CrossRef]
8. Hoşöz, M. Performance comparison of single-stage and cascade refrigeration systems using R134a as the working fluid. *Turk. J. Eng. Environ. Sci.* **2005**, *29*, 285–296.
9. He, M.-G.; Li, T.-C.; Liu, Z.-G.; Zhang, Y. Testing of the mixing refrigerants HFC152a/HFC125 in domestic refrigerator. *Appl. Therm. Eng.* **2005**, *25*, 1169–1181. [CrossRef]

10. Wongwises, S.; Chimres, N. Experimental studies of hydrocarbon mixture to replace HFC134a in a domestic refrigerator. *Energy Convers. Manag.* **2005**, *46*, 85–100. [CrossRef]
11. Fatouh, M.; ELKafafy, M. Assessment of propane / commercial butane mixtures as possible alternatives to R134a in domestic refrigerators. *Energy Convers. Manag.* **2006**, *47*, 2644–2658. [CrossRef]
12. Ding, G.-L. Recent developments in simulation techniques for vapour-compression refrigeration systems. *Int. J. Refrig.* **2007**, *30*, 1119–1133. [CrossRef]
13. Jwo, C.-S. Effect of nanolubricant on the performance of hydrocarbon refrigerant system. *J. Vac. Sci. Technol. B* **2009**, *27*, 1473–1477. [CrossRef]
14. Mohanraj, M.; Jayaraj, S.; Muraleedharan, C.; Chandrasekar, P. Experimental investigation of R290/R600a mixture as an alternative to R134a in a domestic refrigerator. *Int. J. Therm. Sci.* **2009**, *48*, 1036–1042. [CrossRef]
15. Padilla, M.; Revellin, R.; Bonjour, J. Exergy analysis of R413A as replacement of R12 in a domestic refrigeration system. *Energy Convers. Manag.* **2010**, *51*, 2195–2201. [CrossRef]
16. Agarwal, B.; Vipin, S. Retrofitting of vapour compression refrigeration trainer by an eco-friendly refrigerant. *Indian J. Sci. Technol.* **2010**, *3*, 455–458. [CrossRef]
17. Dalkilic, A.; Wongwises, S. A performance comparison of vapour- compression refrigeration system using various alternatives refrigerants. *Int. Commun. Heat Mass Transf.* **2010**, *37*, 1340–1349. [CrossRef]
18. Tiwari, A.; Gupta, R. Experimental Study of R404A and R134A in Domestic Refrigerator. *Int. J. Eng. Sci. Technol.* **2011**, *3*, 8. Available online: https://ijcrr.com/article_html.php?did=2139 (accessed on 14 September 2022).
19. Bolaji, B.O.; Akintunde, M.; Falade, T.O. Comparative Analysis of Performance of Three Ozone-Friends HFC Refrigerants in a Vapour Compression Refrigeration. *J. Sustain. Environ.* **2011**, *2*, 61–64. Available online: http://repository.fuoye.edu.ng/handle/123456789/1231 (accessed on 14 September 2022).
20. Liu, Z.; Haider, I.; Liu, B.; Radermacher, R. *Test Result of Hydro Carbon Mixtures in Domestic Refrigerators Freezers*; Centre Environmental Energy Engineering, University of Maryland: College Park, MD, USA, 1994. Available online: https://p2infohouse.org/ref/14/13828.pdf (accessed on 14 September 2022).
21. Mishra, R.S. Thermal Performance of three stage cascade vapour compression refrigeration systems using new HFO in high and intermediate temperature cycle and R32 ethylene and hydrocarbons in ultra-low temperature cycle refrigerants. *Int. J. Adv. Res. Innov.* **2020**, *4*, 109–123. [CrossRef]
22. Domanski, P.A.; Brignoli, R.; Brown, J.S.; Kazakov, A.F.; McLinden, M.O. Low-GWP refrigerants for medium and high-pressure applications. *Int. J. Refrig.* **2017**, *84*, 198–209. [CrossRef] [PubMed]
23. Lemmon, E.W.; Bell, I.H.; Huber, M.L.; McLinden, M.O. NIST Standard Reference Database: REFPROP Reference Fluid Thermodynamic and Transport Properties, Version 9.4.4.44. 2018. Available online: https://www.nist.gov/srd/refprop (accessed on 8 November 2022).
24. Bell, I.H.; Domanski, P.A.; McLinden, M.O.; Linteris, G.T. The hunt for nonflammable refrigerant blends to replace R134a. *Int. J. Refrig.* **2020**, *104*, 484–495. [CrossRef] [PubMed]
25. Domanski, P.A.; McLinden, M.O. A Simplified Cycle Simulation Model for the Performance Rating of Refrigerants and Refrigerant Mixtures. In *International Refrigeration and Air Conditioning Conference*; 1990; Volume 132. Available online: http://docs.lib.purdue.edu/iracc/132 (accessed on 8 November 2022).
26. CAN/ANSI/AHRI540. *Performance Rating of Positive Displacement Refrigerant Compressors and Compressor Units*; AHRI: Aiken, SC, USA, 2015; Volume 5.
27. Brignoli, R.; Brown, J.S.; Skye, H.M.; Domanski, P.A. Refrigerant Performance Evaluation Including Effects of Transport Properties and Optimized Heat Exchangers. *Int. J. Refrig.* **2017**, *80*, 52–65. [CrossRef] [PubMed]
28. Gil, B.; Kasperski, J. Efficiency Evaluation of the Ejector Cooling Cycle using a New Generation of HFO/HCFO Refrigerant as a R134a Replacement. *Energies* **2018**, *11*, 2136. [CrossRef]
29. Barati-Harooni, A.; Najafi-Marghmaleki, A. Prediction of Vapor-Liquid Equilibrium for Binary Mixtures Containing R1234yf or R1234ze (E). *Int. J. Refrig.* **2018**, *88*, 239–247. [CrossRef]
30. Nguyen, V.V.; Varga, S.; Dvorak, V. HFO1234ze(e) As an Alternative Refrigerant for Ejector Cooling Technology. *Energies* **2019**, *12*, 4045. [CrossRef]
31. Emmi, G.; Bordignon, S.; Carnieletto, L.; De Carli, M.; Poletto, F.; Tarabotti, A.; Poletto, D.; Galgaro, A.; Mezzasalma, G.; Bernardi, A. A Novel Ground-Source Heat Pump with R744 and R1234ze as Refrigerants. *Energies.* **2020**, *13*, 5654. [CrossRef]
32. Saengsikhiao, P.; Taweekun, J.; Maliwan, K.; Sae-Ung, S.; Theppaya, T. Investigation and Analysis of R463A as an Alternative Refrigerant to R404A with Lower Global Warming Potential. *Energies* **2020**, *13*, 1514. [CrossRef]
33. Gil, B.; Szczepanowska, A.; Rosiek, S. New HFC/HFO Blends as Refrigerants for the Vapor-Compression Refrigeration System (VCRS). *Energies.* **2021**, *14*, 946. [CrossRef]
34. Karageorgis, A.; Hinopoulos, G.; Kim, M.-H. A Comparative Study on the Condensation Heat Transfer of R-513A as an Alternative to R134a. *Machines* **2021**, *9*, 114. [CrossRef]
35. Bharanitharan, K.J.; Senthilkumar, S.; Chen, K.-L.; Luo, K.-Y.; Kang, S.-W. Correlations Based on Numerical Validation of Oscillating Flow Regenerator. *Processes* **2022**, *10*, 1400. [CrossRef]
36. Kumar, A.; Chen, M.-R.; Hung, K.-S.; Liu, C.-C.; Wang, C.-C. A Comprehensive Review Regarding Condensation of Low-GWP Refrigerants for Some Major Alternatives of R134a. *Processes* **2022**, *10*, 1882. [CrossRef]

37. Nikitin, A.; Farahnak, M.; Deymi-Dashtebayaz, M.; Muraveinikov, S.; Nikitina, V.; Nazeri, R. Effect of ice thickness and snow cover depth on performance optimization of ground source heat pump based on the energy, exergy, economic and environmental analysis. *Renew. Energy* **2022**, *185*, 1481. [CrossRef]
38. Deymi-Dashtebayaz, M.; Sulin, A.; Ryabova, T.; Sankina, I.; Farahnak, M.; Nazeri, R. Energy, exergoeconomic and environmental optimization of a cascade refrigeration system using different low GWP refrigerants. *J. Environ. Chem. Eng.* **2021**, *9*, 106473. [CrossRef]
39. Nikitin, A.; Deymi-Dashtebayaz, M.; Muraveinikov, S.; Nikitina, V.; Nazeri, R.; Farahnak, M. Comparative study of air source and ground source heat pumps in 10 coldest Russian cities based on energy-exergy-economic-environmental analysis. *J. Clean. Prod.* **2021**, *321*, 128979. [CrossRef]
40. Deymi, M.; Maddah, S.; Goodarzi, M.; Maddah, O. Investigation of the effect of using various HFC refrigerants in geothermal heat pump with residential heating applications. *J. Therm. Anal. Calorim.* **2020**, *141*, 361–372. [CrossRef]
41. Deymi-Dashtebayaz, M.; Valipour-Namanlo, S. Thermoeconomic and environmental feasibility of waste heat recovery of a data center using air source heat pump. *J. Clean. Prod.* **2019**, *219*, 117–126. [CrossRef]
42. Honda, H.; Wijayanta, A.; Takata, N. Condensation of R407C in a horizontal microfin tube. *Int. J. Refrig.* **2005**, *28*, 203–211. [CrossRef]
43. Cavallini, A.; Del Col, D.; Mancin, S.; Rossetto, L. Condensation of pure and near-azeotropic refrigerants in microfin tubes: A new computational procedure. *Int. J. Refrig.* **2009**, *32*, 162–174. [CrossRef]
44. Manoj, I.V.; Soni, H.; Narendranath, S.; Mashinini, P.M.; Kara, F. Examination of Machining Parameters and Prediction of Cutting Velocity and Surface Roughness Using RSM and ANN Using WEDM of Altemp HX. *Adv. Mater. Sci. Eng.* **2022**, *2022*, 5192981. [CrossRef]
45. Chen, C.; Teng, J.-T.; Cheng, C.-H.; Jin, S.; Huang, S.; Liu, C.; Lee, M.-T.; Pan, H.-H.; Greif, R. A study on fluid flow and heat transfer in rectangular microchannels with various longitudinal vortex generators. *Int. J. Heat Mass Transf.* **2014**, *69*, 203–214. [CrossRef]
46. Hsieh, S.-S.; Lee, R.-Y.; Shyu, J.-C.; Chen, S.-W. Thermal performance of flat vapor chamber heat spreader. *Energy Convers. Manag.* **2008**, *49*, 1774–1784. [CrossRef]
47. Hamid, K.; Sajjad, U.; Yang, K.S.; Wu, S.-K.; Wang, C.-C. Assessment of an energy efficient closed loop heat pump dryer for high moisture contents materials: An experimental investigation and AI based modelling. *Energy* **2022**, *238*, 121819. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

2022 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON) | 978-1-6654-6374-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/CENTCON56610.2022.10051370

Exploring Vision Transformer model for detecting Lithography Hotspots

Sumedha
Department of Electronics and Communication
Delhi Technological University
Delhi, India
sumedhachugh27@gmail.com

Rajesh Rohilla
Department of Electronics and Communication
Delhi Technological University
Delhi, India
rajesh@dce.ac.in

Abstract—**In the process of IC design, lithography can be defined as the process of reprinting the pattern of the mask on a Silicon wafer. Lithography is an essential step in this process as it enables feature size to decrease which further helps in decreasing device size. This continuous decrease in feature size may lead to printability issues and hotspots. Presence of hotspots can cause the circuit to fail, so it is very important to detect these hotspots with high accuracy. Previously various simulation, machine leaning and deep learning based techniques have been implemented to solve this issue. In this paper, a method to identify hotspots using Vision Transformers is proposed. Other deep learning techniques, such as CNNs and ANNs have also been used for comparison purposes. All three techniques are implemented on five datasets. ViT gives an overall average accuracy of 98.05% which is 1.39% higher than accuracy of CNNs and 2.04% higher than accuracy given by ANNs. Although the ViTs prove the best in terms of overall accuracy, but at dataset level its performance can be improved. Three out of five datasets have accuracy higher than 99% and for rest two it is slightly above 95%. In future, we wish to improve accuracy for these two datasets by improving the model and reducing imbalance in the datasets.**

Keywords—Deep Learning, Lithography, Hotspot Detection, Vision Transformer, Convolution Neural Network, Artificial Neural Network

I. Introduction

In the process of IC fabrication, patterns are generated on a silicon wafer. These patterns are first obtained on a mask and then transferred on silicon wafer through the process known as lithography. In order to fit more and more transistors in the same area, feature size needs to get smaller. In optical lithography, feature size is directly proportional to wavelength. Mathematically,

$$f = \frac{C.\lambda}{n} \tag{1}$$

Here, f is the feature size, C is the Rayleigh constant which measures how difficult lithography is, λ is the wavelength, and n is the numerical aperture.

The most effective way to reduce feature size is to reduce the wavelength of light. This wavelength reduction leads to printability problems and degradation in resolution [1]. Although Resolution Enhancement Techniques such as Optical Proximity Correction, Sub Resolution Assist Feature (SRAF) are employed to improve the process, but at some locations, differences exist between patterns on mask and wafer. Positions where patterns have dimensions more or less than the defined threshold are known as hotspots [2]. In electron beam lithography, electrons scatter, and these scattered electrons may cover a different path than the one drawn in mask, which may cause hotspots [21]. The hotspots may lead to an open circuit or short circuit [11]; hence detecting them is very important. Various Simulations, Pattern Matching, Machine Learning, and Deep Learning based techniques have been implemented to eliminate of hotspots. Simulation and Deep Learning based techniques are very expensive in terms of computational time and complexity [4,5,26-29]. Although Pattern Matching techniques are faster, they fail to detect previously unseen hotspots [2]. Machine Learning based techniques are faster and more efficient in detecting previously unseen hotspots, but due to an imbalance in dataset, false positives remain a big problem in this method [17,19], which leads to more time for preprocessing the data [11]. These techniques have been discussed in detail in section 2.

In this work, we propose using Vision Transformer (ViT) technique to identify lithography hotspots. Transformers have a wide range of applications in the fieid of Natural Language Processing. However, these are only used to a small extent in Computer Vision based applications due to high number of computations involved, which are not possible to achieve with hardware [16]. ViT aims to overcome these limitations by converting image into patches, passing them through the transformer encoder structure and finally classifying them [3]. This technique, introduced by Alexey Dosovitskiy et al. in June 2021 has not been previously utilized for detecting hotspots. Section 4 discusses details of the datasets and all experiments performed for implementation and comparison of above mentioned techniques. From these trials, it can be observed that ViT gives 1.39% higher overall average accuracy than the accuracy of CNNs and 2.04% higher than the accuracy given by ANNs. While comparing to already existing works, ViT performs the best or comparable to best for three out of five datasets, but it is not able to supplant all the existing methods for all the datasets. In section 5, results are shown, followed by conclusions and future scope in chapter 6.

II. Related Work

Based on methods used for detecting lithography hotspots, all the models can be divided into five categories, these being: Simulation, Pattern Matching, Machine Learning, Deep Learning, and Lithography Hotspot Mitigation based methods. In 1979, the first optical lithography simulation method called SAMPLE was introduced. This technique provided better results for grids with greater sizes [4]. In 1985 another method for simulation called Positive Resist Optical Lithography (PROLITH) was introduced, which made the process of lithography highly accessible as it was the first time when a model could run on a Personal Computer. Various improved versions of PROLITH are still used as simulator for optical lithography process. Simulators used for electron beam are electromagnetic field simulator ProMAX, Monte Carlo, ProBEAM [4]. Full layout simulation is a highly accurate

978-1-6654-6374-4/22/$31.00 ©2022 IEEE

way of recognizing hotspots, but it is very expensive in terms of computation time and complexity [5].

To know if some place is a hotspot or not, design rules which define the minimum distance between two patterns on the same mask so that they don't overlap or define gap between edges in case of same pattern must be followed [5, 6]. Pattern Matching algorithms like string search, tangent space, transitive closure graph, template matching, dual graph etc. first check if each image of the dataset provided follows these rules or not, then a search algorithm is applied to look for hotspots [6]. Graph based techniques create a dual or transitive graph for the layouts provided, layouts with edges or adjacent patterns having lesser spacing or greater width than mentioned in the rules are considered hotspots. In this process, the chance of non-hotspots being detected as hotspots is large [7]. Template matching methods move the pattern to be detected over the entire mask pixel wise and patterns where some disruptions from mask are seen, are termed as faulty or with hotspot [8]. The string based methods covert layouts that are two dimensional into a single dimension; these single dimensional structures are named strings. Then, search operations using distance arrays are done to find strings with hotspots [8, 9, 10]. Although the pattern matching based methods are faster than simulations, but they fail to detect previously unseen hotspots [2].

Lithography Hotspot Mitigation is a process of reducing the risk of hotspots by taking some preventive measures before lithography process, such as during Placement or Routing [11, 12-14]. Adjacent patterns may interfere during the placement process leading to hotspots, which can be avoided by using multiple patterning [11]. Lithography simulation and Edge Placement Error (EPE) guided routing [13] help in optimizing the layout after routing process and reduce hotspots by a significant amount. EPE map compares the edge shapes of the layout to edge shapes that are intended in the form of a matrix and then finds hotspots [14].

In machine learning models, features need to be extracted before performing classification in order to reduce the size of training data, hence leading to an increase in speed. DBLF technique divides the layout into sub-regions and calculates their densities. This information is represented in the form of vectors and can then be classified using machine learning [15]. DBLF based models classify layouts with the same density in the same class, which may cause errors [7]. To overcome weakness of DBLF, HLOP is used, which along with density, can also capture the direction in which light transmits by calculating Histogram Oriented Gradient (HOG) of regions obtained after performing Gaussian Blurring on the image [7]. Topological classification combined with Critical feature extraction is based on deriving geometry based i.e. topological (preserved) and process based i.e. non-topological feature. According to different topologies like the distance between edges of a pattern, the distance between two patterns and non-topologies like number of corners, number of points touching clusters are made, and an SVM kernel related to each critical feature is constructed. These feature specific kernels help to identify hotspots with more accuracy [5]. MCMI is Information Theory based technique that extracts features with high information and takes care that redundancy is less, making the process fast and efficient [4]. Encoder-Decoder feature extractor consists of convolution and de-convolution layers, which helps features to transform and makes it easier to work with CNNs [16]. Feature extraction methods lose the relations among structures; to solve this issue MCCS is used, which stores information in the form of matrix [17]. Machine learning techniques used for hotspot detection purpose are SVM [7, 11, 15, 18, 19], Boosting [7, 20], PCA [7, 8], clustering [11, 21-23], Naive Bayes [24], Bilinear Classifier (Combination of SVM and Ada-boost) [17]. A bilinear classifier is used with MCCS feature extractor in order to preserve the topological relations [17]. Semi-supervised techniques have also been used to identify hotspots with high accuracy [25]. Although ML based detectors overcome the weakness of Pattern Matching, false alarms remain a problem, which leads to exhaustive and costly post-processing [17, 19]. These days Deep Learning based techniques involving CNNs [26, 27], ANN architecture combined with CNNs [4], GANs [28], CNNs with DBSCAN clustering [29], feature extraction followed by CNNs [30] etc. are being used to reduce False Alarms

III. Preliminaries

A. Lithography:

Litho means sculpture of a 2-D or 3-D structure made of metals or stones. Reprinting what is present somewhere else is known as lithography. In the 1960s, circuits were in micrometer range, today these are a few nanometers, and it has been possible with the help of patterning. It controls shapes, dimensions, and placement of various components. Based on the type of radiation, there are basically four types of lithography processes: Optical Lithography, X-Ray Lithography, Electron beam Lithography and Ion Beam Lithography [31].

For optical lithography light from a source is projected through a mask via a focusing lens. That pattern is then focused on the photoresist-coated wafer. Soluble part of the photoresist is rinsed away, leaving an image of the required pattern on the chip. [31] Mask a.k.a. Reticle contains hardcopy of the design that needs to be transferred from mask to the photoresist. For a multilayer design each layer has its own mask. The process of making an original hard mask is complex and may take hours to make, but once it is done, the pattern can be quickly transferred to different wafers, making it easy to get multiple patterns [31, 37]. A photoresist is light sensitive material that is applied on top of a Silicon wafer. For this, one or two drops of photoresist are taken, which are put on wafer, and the wafer is spun really fast so that it uniformly spreads. In positive photoresist, resolution is better, but throughput is high. This photoresist is now dipped in developer solution to remove the softened part, oxide layer is removed which prepares final wafer.

Optical Lithography can further be divided into three types: Contact printing, Proximity printing, and Projection printing. In contact printing wafer is in contact with the mask, hence resolution is high. But life of the mask is reduced due to wear and tear because of contact. In this type risk of contamination is also there because if some dirt is present on the mask, it is transferred on wafer. In Proximity printing mask and wafer are close but not in contact. In this type, resolution decreases a little bit but life increases. In Projection printing, mask and wafer can be as far as one wants. In this, a highly focused image of the mask is projected on the wafer; hence resolution is high and wafer life increases. Because of extra optical setup required, cost also increases.

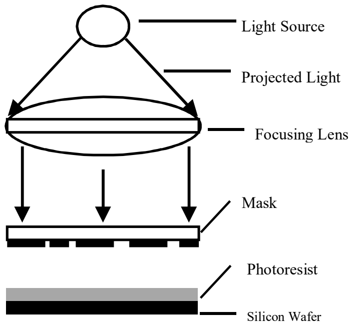


Fig. 1. Lithography Process

There are three figures of merit of lithography: Resolution, Throughput and Depth of focus. If there is a defined feature size in mask, after transferring it may increase or decrease. The closer the feature size to mask, the better the resolution. If original feature size is small, it is required to be very accurate for good resolution. Throughput defines the number of wafers that can be prepared in a given time. It is important because the process needs to be cost effective. Since there are multiple patterns in a design, all of these should be properly aligned with respect to each other as well as base the Silicon wafer. For this alignment to become easier, the depth of focus must be good. e.g., In a simple BJT five masks are required: Active area mask, Junction Isolation mask, Base diffusion mask, Emitter diffusion mask, and Contact metal mask. All these masks must be aligned with each other [31].

For chip area and total cost of chip making to reduce, transistors need to get smaller, [21], so it is desired that the trend of area getting smaller in a cost-effective way continues. Angle of diffraction from mask is sinθ, which is mathematically written as:

$$sin\theta = \frac{c.\lambda}{f} \tag{6}$$

where sinθ is the angle of diffraction, f is the feature size, c is constant and λ is the wave length. Since 0th differential order does not contain any information, the lens needs to capture at least first differential order to recreate pattern of the mask. Hence, the smaller the feature size, the bigger the diffraction angle, the bigger the lens needed.

To reduce f, one can increase n or reduce λ. To increase n, water or oil can be used as medium instead of air. It is known as immersion lithography. X-rays can be used to reduce λ instead of UV light in optical lithography. It is known as X-ray lithography and is done using proximity printing. When electron beam is focused on water cooled palladium target, X-rays are generated. This process is done inside Helium chamber as it doesn't absorb X-rays [31]. Mask used in this technique is a thin membrane made of Aluminium Oxide, Silicon or Silicon Nitride, coated with gold (because gold can absorb X-rays). Most commonly used mask for this lithography is Poly Methyl Meta Crystal (PMMA); electron beam resists can also be used. This technique has various advantages, such as smaller throughput, low proximity effect and high resolution. X-rays do not absorb dirt so, the risk of contamination is also less. One limitation of this process is the blurring of image on the substrate, which depends on the distance between X-ray source and mask and the separation of mask and wafer [21,31].

Another way of obtaining smaller feature size is by using electron beam lithography; in this direct writing on the substrate is possible and mask is not required. Hence, it provides a better depth of focus and resolution; and can be easily automated [31]. The time required for this is very high, leading to a smaller throughput. Another disadvantage of this lithography technique is the proximity effect. When the electron beam is focused on substrate, scattering of electrons take place so; they go large distances away from the original pattern. Because of this broadening of actual pattern and feature size takes place, this is known as proximity effect. Insufficient radiation at corners doesn't let them fully develop; this effect is known as intra-proximity, and overdevelopment of certain areas due to extra exposure is called inter-proximity [31].

Ion-beam can be used in place of electron beam leading to less scattering and hence, less proximity effect. This process is known as Ion-beam lithography. Designers are given rules such as how much edge to edge distance one should keep, how much separation should be there, length and width which must be followed to get least errors possible. However, sometimes because of decreasing feature size, printability issues occur; those places with such issues are known as lithography hotspots [31]

B. Vision Transformer

Today most of the best performing Natural Language Processing models like BERT, GPT are Transformer based, but in computer vision based applications, these are only used to a small extent. Till now, transformers have been used with CNNs, but never at their place. Vision Transformer aims to use transformers for image classification tasks without the involvement of convolutions [16].

Transformers make use of the attention mechanism and parallelization to outperform other NLP models. Mathematically, Attention can be described with the help of function a_{ij}, which can be defined as:

$$a_{<i,j>} = f_{att}(b_{j-1}, h_j) \tag{7}$$

where a_{ij} captures the importance of ith sequence/word in getting jth output, function, f_{att} has many choices depending on the problem to be solved, b_{i-1} captures the state of the encoder so far, and h_j is the input sequence.

These weights are then normalized by using softmax function to get the parameter Γ.

$$\Gamma_{<i,j>} = exp(e_{<i,j>}) \frac{1}{\sum_{i=1}^{M} exp(} \tag{8}$$

The output of self attention layer goes to feed forward network, which converts output attention vector to a form that can be processed by next encoder block or decoder block. The output of feed forward network goes to the next encoder and so on; the output of encoder is fed to decoder. Decoder block consists of three layers, first being self attention layer which generates attention vector for input. These attention vectors and vectors from encoders are passed through the encoder-decoder attention block, which decides how related these two vectors are. Its output is fed to feed forward network, which passes the output to next decoder or linear layer. The linear layer extends dimensions to dimensions required for output, and the softmax layer transforms it to probability distribution at output.

Ideally, transformers operate on sequences or sets by applying attention mechanism on them. i.e., Since attention is a quadratic operation, pair-wise inner product needs to be calculated between each pair of sets; therefore computations and memory required are very high. For images, it is harder because these are composed of many pixels. Even a small image consists of approximately 250*250 pixels. Every pixel needs to attend every other pixel, so even for a small image $((250)^2)^2$ operations are required. This much computations are not possible to achieve with hardware [3].

Vision transformers make use of transformers by including some operations to reduce the number of computations required. First of all, the image is partitioned into patches of same shape, which may or may not overlap; and every patch is a small image. Then these patches are vectorized by reshaping tensors to vectors, letting the output be X_1, X_2, ….., X_n. Next dense layer is applied to these vectors; let the outputs of dense layer be Z_1, Z_2, ….., Z_n, where $Z_n = w*X_n + b$. Then, positional encoding is added to these patches because swapping of patches may lead to information loss. Other than these, a CLS token is passed through the embedded layer, and its output is used to provide classification output. All these vectors are passed through the transformer encoder network. Since we need to perform classification task output of encoder i.e., a feature vector is not passed through a decoder network as it is more efficient for generation related tasks instead, it is passed through MLP head which acts as a classifier and provides image classification output [16].

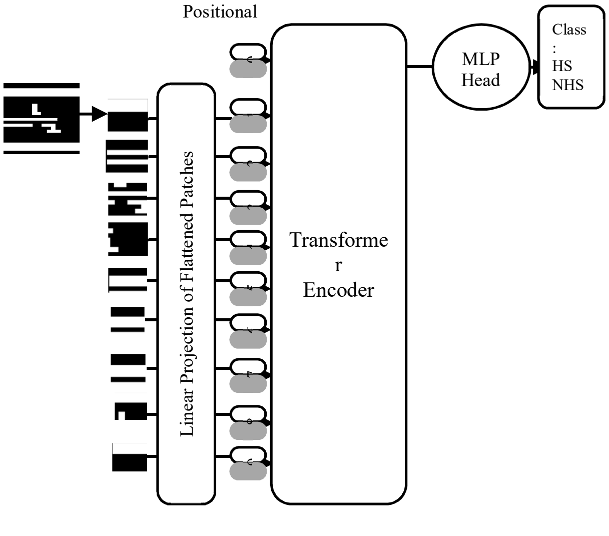


Fig. 2. Structure of Vision Transformer [16]

In transformers, one can pay attention to the things which are far away; this is not possible in CNNs. One feature that has been observed with ViTs is that they perform better than ResNETs only when training data is sufficiently large; otherwise they are equally or less effective. In CNNs, integration is done over a pixel, which connects to its neighborhood. Then that neighborhood connects to its neighborhood and so on. This is known as local attention. ViTs work on the principle of global attention i.e., all the points are connected at once [3].

IV. EXPERIMENTS PERFORMED

We explored ViT model on ICCAD 2012 dataset to see its efficiency in detecting lithography hotspots. For comparison, CNN and ANN models have also been applied. Details of the dataset are as follows:

A. Dataset Used

For this research, ICCAD-2012 dataset is utilized. It is further divided into five datasets with different types of layouts. The first dataset was obtained using 32 nm process and other four were obtained during 28 nm process. Each dataset contains a training and test set; detailed information of each dataset is provided in the following table.

TABLE 1 DETAILS OF DATASET USED

Dataset	**Training Data**			**Test Data**		
	Hotspot	***Non-Hotspot***	***Total***	***Hotspot***	***Non-Hotspot***	***Total***
Dataset1	99	340	439	226	3869	4095
Dataset 2	174	170	5459	498	41298	41796
Dataset 3	909	406.6	5552	1808	46333	48141
Dataset 4	95	16.98	4547	177	31890	32067
Dataset 5	26	2176	2202	41	19327	19368
Dataset 6	1303	16896	8199	2750	142717	145467

The results provided are after passing lithography output through design the rule checker, and hence the number of hotspots is much less than the number of non-hotspots. Many techniques have been used previously to take care of this imbalance, some of them being data augmentation and filtering [32]. Another problem faced by this dataset is false alarms. It has been seen that using synthetic patterns to increase the amount of training data significantly reduces these false alarms [33].

B. Model:

To implement ViT, we loaded a pre-trained transformer model 'vit_base_patch16_224' provided by Hugging Face. This model is trained on the ImageNet21k dataset and then fine-tuned on ImageNet dataset. It uses a patch size of 16*16. Chained transformations are used to resize images to 224*224 resolution. These resized images are then converted to tensors and normalized with mean and standard deviation = 0.5 for all three channels. The model has been created using torch image model library and pre-trained transformer model.

For generating the model, we tried various parameters; batch size set to 100, learning rate with values 1e-2 for datasets 2,3,4,5, and 1e-3 for dataset 1 gave the best results. While compiling the model, Adam optimizer, cross-entropy loss, and accuracy matrix were used, and the output of the model is discussed in the results section. Steps performed for classification using ViT are mentioned in Figure 3.

Similarly, for generating models for CNNs and ANNs, various values of different parameters have been tried. A sequential model with a batch size of 64 and Images reshaped to 224*224 resolution have been used. For generating a model for CNNs, three convolution layers with 12 filters and Kernel Size = (3, 3), two dense layers, and two max-pooling layers with pooling window (2, 2) gave the best results. For ANNs best accuracy has been obtained using 9 dense layers and 8 dropout layers with rate = 0.3. The same parameters have been used while compiling the models for all three techniques. Steps performed for classification using CNN and ANN are the same as in case of ViT, except that we don't use a pre-trained model for CNNs.

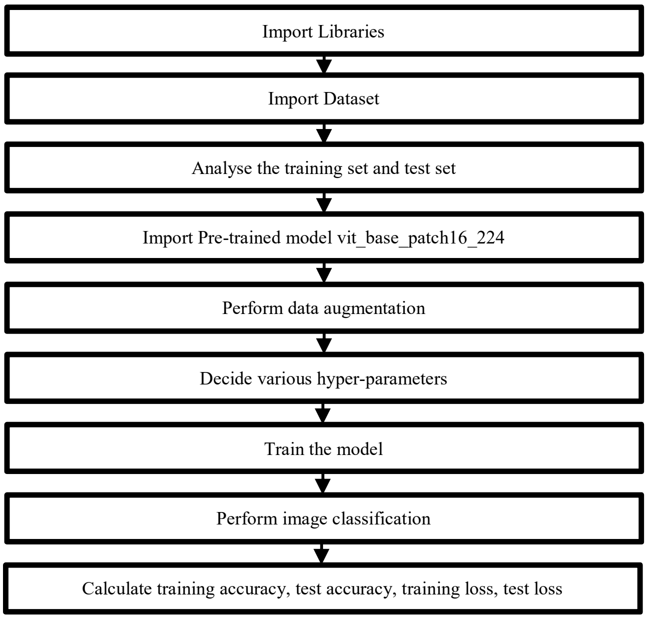

Import Libraries
Import Dataset
Analyse the training set and test set
Import Pre-trained model vit_base_patch16_224
Perform data augmentation
Decide various hyper-parameters
Train the model
Perform image classification
Calculate training accuracy, test accuracy, training loss, test loss

Fig. 3. Steps Performed for Classification using VIT

V. Result

From Table 2, it can be seen that ViT performs better than CNNs and ANNs in terms of overall accuracy.

TABLE 2 ACCURACY FOR ALL SUBSETS USING VIT, CNNS AND ANNS

Dataset	**Accuracy**					***Overall Average Accuracy***
	Dataset 1	***Dataset 2***	***Dataset 3***	***Dataset 4***	***Dataset 5***	
ViT	95.48	99.37	95.77	99.83	99.8	98.05
CNNs	94.37	98.81	90.91	99.45	99.79	96.66
ANNs	89.58	97.73	94.58	96.68	99.48	96.01

Fig. 4 shows that for all the datasets ViT gives the best results.

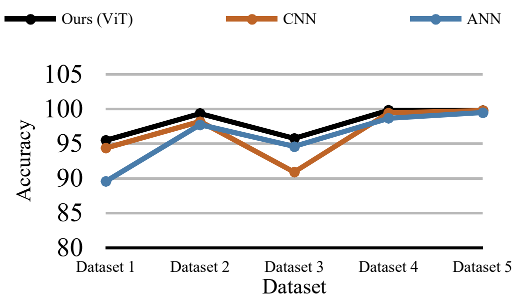


Fig. 4. Comparing accuracies for ViT, CNNs and ANNs

CNNs perform moderately well for datasets 1, 2, 4 and 5 and worst for dataset 3. ANNs perform poorest for sub-datasets 1, 2, 4, 5 and moderately well for sub-dataset 3. In terms of overall accuracy ViTs give 1.39% better accuracy than CNNs and 2.04% better accuracy than ANNs.

Table 4 compares our ViT model with other researches. Figure 5 shows that in terms of overall accuracy ViT model gives the best result. In terms of individual datasets, for dataset 4 and 5, ViT gives the best accuracy. For dataset 2, accuracy is not the best, but it is comparable to best performing models. For dataset 1, [24] and [20] provide the best results and for sub-dataset 3, [27] gives the best accuracy.

TABLE 3 COMPARISON WITH OTHER WORKS

Dataset	**Accuracy**					***Overall Average Accuracy***
	Dataset 1	***Dataset 2***	***Dataset 3***	***Dataset 4***	***Dataset 5***	
Ours(ViT)	95.48	99.37	95.77	99.83	99.8	98.05
Y.Yu et.al, [5]	93.81	99.2	91.88	85.94	92.86	92.53
H.Yang et.al, [27]	99.6	99.8	97.8	96.4	95.1	97.74
H.Zhang et.al, [24]	100	99.4	97.57	97.74	95.12	97.95
T.Matsunanwa et.al, [20]	100	98.6	97.2	87.1	92.68	95.11

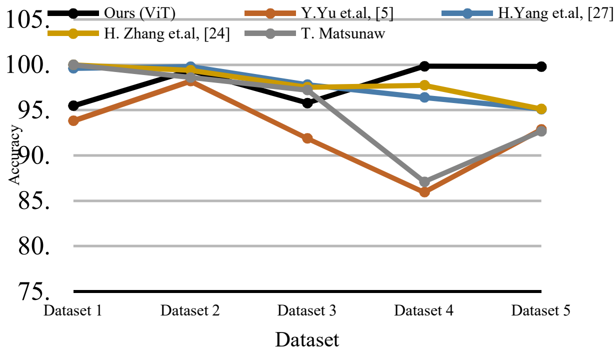


Fig. 5. Comparing accuracy of ViT and other research works

VI. Conclusion and Future Scope

In this paper, lithography hotspots have been detected using Vision Transformers. To see if this proposed technique gives better results than the already existing deep learning techniques, we applied CNNs and ANNs to solve this problem along with ViT. Table 2, shows that in terms of overall accuracy ViT gives 1.39% better accuracy than CNNs and 2.04% better accuracy than ANNs. Considering individual dataset wise accuracies, ViT performs better or as well as CNNs for each dataset. For three out of five datasets ,accuracy on the test set is more than 99%, and for the other two, it is more than 95%. Table 3 shows comparison of ViT model with existing research works, and it can be seen that in terms of overall accuracy, ViT gives the best results. At the individual dataset level for three out of five datasets, it provides the best or comparable results but lags for two datasets. From the results, it can be concluded that although the proposed technique performs better than many already existing state of the art techniques, it can only supplant some of the existing methods for some of the datasets. Since the dataset is imbalanced, accuracy is affected by it. Future research aims to improve accuracy for datasets 1 and 3 by improving the model and modifying the dataset using techniques like data augmentation, mirror flipping and upsampling on training sets to reduce the imbalance in it and increase the training data. Since the technique is very novel and many improvements lie for it in the coming future, ViTs can be seen as a new and alternate method to identify hotspots in lithography.

REFERENCES

1. Sivakumar, S. EUV lithography: Prospects and challenges In 16th Asia and South Pacific Design Automation Conference, 2011, pp. 402-402 doi: 10.1109/ASPDAC.2011.5722221.
2. Robles, J. A.T.; Mostafa, S.; Madkour, K. & Wuu, J.Y. Hotspot Detection Based on Machine Learning. In United States Patent Application Publication, 2013, , [Online], Available: https://patents.google.com/patent/US20130031522. [Accessed: June 2021]
3. Gkelios, S.; Boutalis, Y. & Chatzichristofis, S.A. Investigating the Vision Transformer Model for Image Retrieval Tasks. 11 Jan 2021, [Online], Available: arXiv:2101.03771v1. [Accessed: June 2021]
4. Ye, W.; Lin , Y.; Li, M.; Liu, Q.; & Pan, D.Z. LithoROC: lithography hotspot detection with explicit ROC optimization. In ASPDAC '19: Proceedings of the 24th Asia and South Pacific Design Automation Conference, January 2019, pp. 292–298.
5. Tomioka, Y.; Matsunawa, T.; Kodama, C. & Nojima, S. Lithography hotspot detection by two-stage cascade classifier using histogram of oriented light propagation. In 22nd Asia and South Pacific Design Automation Conference (ASP-DAC), 2017, pp. 81-86 doi: 10.1109/ASPDAC.2017.7858300.
6. Yu, Y.; Lin, G.; Jiang, I. H. & Chiang, C. Machine-Learning-Based Hotspot Detection Using Topological Classification and Critical Feature Extraction In IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 34, no. 3, pp. 460-470, March 2015 doi: 10.1109/TCAD.2014.2387858.
7. Yu, Y.; Chan, Y.; Sinha, S.; Jiang, I. H. & Chiang, C. Accurate process-hotspot detection using critical design rule extraction In DAC Design Automation Conference 2012, 2012, pp. 1163-1168.
8. Reddy, G. R.; Xanthopoulos, C. & Makris, Y. Enhanced hotspot detection through synthetic pattern generation and design of experiments. In IEEE 36th VLSI Test Symposium (VTS), 2018, pp. 1-6 doi: 10.1109/VTS.2018.8368646.
9. Gao, J.; Yu, B.; Ding, D. & Pan, D. Z. Lithography hotspot detection and mitigation in nanometer VLSI. In IEEE 10th International Conference on ASIC, 2013, pp. 1-4 doi: 10.1109/ASICON.2013.6811917.
10. Tamagawa, S.; Fujimoto, R.; Inagi, M.; Nagayama, S. & Wakabayashi, S. A Table Reference-Based Acceleration of a Lithography Hotspot Detection Method Based on Approximate String Search. In CENICS 2017: The Tenth International Conference on Advances in Circuits, Electronics and Micro-electronics, 2017.
11. Borisov, V. & Scheible, J. Lithography Hotspots Detection Using Deep Learning. In 15th International Conference on Synthesis, Modelling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD), 2018, pp. 145-148. doi: 10.1109/SMACD.2018.8434561.
12. Kim, S.K.; Lee, J.E.; Park, S.W. & Oh, H.K. Optical lithography simulation for the whole resist process. In Current Applied Physics, Volume 6, Issue 1, 2006, pp. 48-53 doi.: 10.1016/j.cap.2004.12.003.
13. Mitra, J.; Yu, P. & Pan, D.Z. RADAR: RET-aware detailed routing using fast lithography simulations. In DAC '05: Proceedings of the 42nd annual Design Automation Conference, June 2005, pp. 369–372 doi.: 10.1145/1065579.1065678.
14. Pan, D. Z. Lithography-aware physical design. In 6th International Conference on ASIC, 2005, pp. 1172-1173 doi: 10.1109/ICASIC.2005.1611242.
15. Yang, H.; Lin, Y.; Yu, B. & Young, E.F. Y. Lithography Hotspot Detection: From Shallow To Deep Learning. In IEEE International System-on-Chip Conference (SOCC), pp. 233–238, Munich, Germany, September 5–8, 2017.
16. Alexey Dosovitskiy, Lucas Beyer Et al.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, June 2021 [Online], Available : https://doi.org/10.48550/arXiv.2010.11929 [Accessed: May 2022]
17. Zhang, H.; Zhu, F.; Li, H.; Young, E.F.Y.; Yu, B. Bilinear Lithography Hotspot Detection Share. In ISPD '17: Proceedings of the 2017 ACM on International Symposium on Physical Design, March 2017, Pp. 7–14 doi.: 10.1145/3036669.3036673.
18. Ye, W.; Alawieh, M. B.; Li, M.; Lin, Y. & Pan, D. Z. Litho-GPA: Gaussian Process Assurance for Lithography Hotspot Detection. 2019 Design, Automation & Test in Europe Conference & Exhibition, 2019, pp. 54-59 doi: 10.23919/DATE.2019.8714960.
19. Ding, D.; Torres, J. A. & Pan, D. Z. High Performance Lithography Hotspot Detection With Successively Refined Pattern Identifications and Machine Learning. In IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 30, no. 11, pp. 1621-1634, Nov. 2011 doi: 10.1109/TCAD.2011.2164537.
20. Matsunawa, T.; Gao, J.R.; Yu, B. & Pan, D. Z. A new lithography hotspot detection framework based on AdaBoost classifier and simplified feature extraction. In SPIE vol. 9427, 2015.
21. Swaminathan, P. " Lithography", [Online], Available: https://onlinecourses.nptel.ac.in/noc20_mm25/. [Accessed: July 2021]
22. Ma, N.; Ghan, J.; Mishra, S.; Spanos, C.; Poolla, K.; Rodriguez, N. & Capodieci, L. Automatic hotspot classification using pattern-based clustering. In SPIE Design for Manufacturability through Design-Process Integration II, 4 March 2008 doi.: 10.1117/12.772867.
23. Ghan, J.; Ma, N.; Mishra, S.; Spanos, C.; Poolla, K.; Rodriguez, N. & Capodieci, L. Clustering and pattern matching for an automatic hotspot classification and detection system In SPIE Design for Manufacturability through Design-Process Integration III, 12 March 2009 doi.: 10.1117/12.814328.
24. Zhang, H.; Yu, B. & Young, E.F. Y. Enabling online learning in lithography hotspot detection with information-theoretic feature optimization. In ICCAD '16: Proceedings of the 35th International Conference on Computer-Aided Design, November 2016, pp. 1–8 doi.: 10.1145/2966986.2967032\
25. Chen, Y.; Lin, Y.; Gai, T.; Su, Y.; Wei, Y. & Pan, D. Z. Semisupervised Hotspot Detection With Self-Paced Multitask Learning. In IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 39, no. 7, pp. 1511-1523, July 2020 doi: 10.1109/TCAD.2019.2912948.
26. Xiao., Y; Huang, X. & Liu, K. Transferability from ImageNet to Lithography Hotspot Detection In J Electron Test 37, pp. 141–149 doi.: 10.1007/s10836-021-05925-5.
27. Yang, H.; Luo, L.; Su, J.; Lin, C. & Yu, B. Imbalance Aware Lithography Hotspot Detection: A Deep Learning Approach. In SPIE Intl. Symposium Advanced Lithography Conference, San Jose, CA, Feb. 26–Mar. 2, 2017.
28. Ye, W.; Alawieh, M. B.; Lin, Y.; Pan, D.Z. LithoGAN: End-to-End Lithography Modeling with Generative Adversarial Networks. In The 56th Annual Design Automation Conference 2019, June 2019 doi.:10.1145/3316781.3317852.
29. Shin, M. & Lee, J.H. Accurate lithography hotspot detection using deep convolutional neural networks. In Journal of Micro/Nanolith. MEMS MOEMS 15(4), 18 Nov 2016 doi.: 10.1117/1.JMM.15.4.043507.
30. Yang, H.; Su, J.; Zou, Y.; Ma, Y.; Yu, B. & Young, E. F. Y. Layout Hotspot Detection With Feature Tensor Generation and Deep Biased Learning. In IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 38, no. 6, pp. 1175-1187, June 2019 doi: 10.1109/TCAD.2018.2837078.
31. Dasgupta, N. Lithography Lecture-1 and Lecture-1. [Online], Available: https://nptel.ac.in/courses/117/106/117106093/. [Accessed: July 2021]
32. Torres, J. A. ICCAD-2012 CAD contest in fuzzy pattern matching for physical verification and benchmark suite. 2012 In IEEE/ACM International Conference on Computer-Aided Design (ICCAD), 2012, pp. 349-350.
33. Reddy, G. R.; Xanthopoulos, C. & Makris, Y. On Improving Hotspot Detection Through Synthetic Pattern-Based Database Enhancement. In IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems doi: 10.1109/TCAD.2021.3049285.

Exposing the Vulnerabilities of Deep Learning Models in News Classification

Ashish Bajaj
Biometric Research Laboratory, Department of Information Technology, Delhi Technological University
Bawana Road, Delhi-110042, India
bajaj.ashish25@gmail.com

Dinesh Kumar Vishwakarma
Biometric Research Laboratory, Department of Information Technology, Delhi Technological University
Bawana Road, Delhi-110042, India
dvishwakarma@gmail.com

2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT) | 978-1-6654-9414-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICITIIT57246.2023.10068577

Abstract—**News websites need to divide their articles into categories that make it easier for readers to find news of their interest. Recent deep-learning models have excelled in this news classification task. Despite the tremendous success of deep learning models in NLP-related tasks, it is vulnerable to adversarial attacks, which lead to misclassification of the news category. An adversarial text is generated by changing a few words or characters in a way that retains the overall semantic similarity of news for a human reader but deceives the machine into giving inaccurate predictions. This paper presents the vulnerability in news classification by generating adversarial text using various state-of-the-art attack algorithms. We have compared and analyzed the behavior of different models, including the powerful transformer model, BERT, and the widely used Word-CNN and LSTM models trained on AG news classification dataset. We have evaluated the potential results by calculating Attack Success Rates (ASR) for each model. The results show that it is possible to automatically bypass News topic classification mechanisms, resulting in repercussions for current policy measures**.

Keywords— Adversarial Attack, News Classification, (Natural Language Processing) NLP, Semantic Similarity, Vulnerability, Transformers.

I. Introduction

Machine Learning (ML) models have excelled in various tasks during the past ten years, including classification, regression, and decision-making. Though, it has been discovered that these models are susceptible to adversarial examples, which are actual inputs modified by tiny, frequently undetectable perturbations, as shown in Figure 1. Recent studies have successfully generated adversarial images[1] that render computer vision algorithms useless. There are few studies of adversarial instances in natural language processing applications like topic classification, sentiment analysis, fake news detection, hate content detection, machine translation, etc. Nevertheless, it is a newer topic that is interesting to investigate and has recently received more attention due to the success of adversarial learning in images. Adversarial examples are generated under two adversarial settings. One is a "black-box setting," i.e., creating an adversarial example when the adversary is unaware of the classifier or the training set. On the other hand, there is a white-box setup in which the adversary has complete knowledge of the classifier and the training data. Formally, outputs of a natural language assaulting system should also satisfy three important utility-preserving features in addition to their capacity to deceive the target models:1.) semantic similarity—as determined by humans, the constructed example should have the same meaning as the source, 2.) Adversarial examples generated should appear natural and grammatical, 3.) Consistency of human prediction—human predictions should not change.

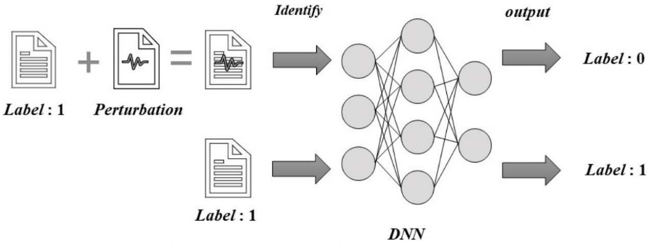


Figure 1 Imperceptible perturbation added to input resulting in giving wrong output

In this work, we present the vulnerability in news topic classification by targeting the widely used long short-term memory and convolution neural networks, including the most potent transformer model, i.e., BERT, for text classification. The models are first trained on the (AG news dataset) a popular dataset for news topic classification. These pre-trained models are then evaluated on their performance degradation by conducting attacks using various *state-of-the-art* adversarial attack algorithms. The results are of potential interest to users who frequently use famous *state-of-the-art* models for their classification tasks. The reader will be able to find out the best fit model for their problem. Also, this motivates the researchers to build models with adversarial robust generalizations instead of standard generalizations. The authors claim that this is the first work raised in the literature that has shown a comparative analysis of the vulnerability of different models on various adversarial attack algorithms for the news classification task.

II. Related Work

A. News Topic Classification

With the rise in the usage of social media applications, the user usually gathers important news from social media sources. Users often seek interest in reading news articles related to them. Based on their interest, the google recommendation system suggests news articles to their users for their benefit. Before the news articles are recommended, the article is first classified into various categories, i.e., sports, entertainment, technical, business, etc. Since news is of multiple types, this is considered a multi-class classification. The classical ML algorithms, such as Naïve

978-1-6654-9414-4/23/$31.00 ©2023 IEEE

Bayes, SVM, etc., have achieved huge success in categorizing news articles[2]. But with the upcoming deep learning techniques, the models are more advanced and intelligent such as Word-CNN, LSTM & transformer models. These models have gained massive attention because of their excellent confidence scores in text classification. In this paper, we have used the three classification models i.e., Word-CNN, Word-LSTM & Bert, for text classification of news articles.

B. Adversarial Attacks in NLP

Adversarial machine learning, which has grown its significance in the area of applied artificial intelligence during the past few years, includes adversarial attacks. In an adversarial attack, a neural network's input data is deliberately changed to see how resilient it is to produce the same results. A corpus of n sentences is provided, $X = \{x_1, x_2, . . ., x_n\}$, and an associated set of n labels $Y = \{y_1, y_2, . . ., y_n\}$, The input text space X is mapped to the label space Y via a pre-trained model called f: $X \rightarrow Y$. A legitimate adversarial example, x_{adv}, for the expression $x \in X$ should meet the following standards:

$f(x_{adv}) \neq f(x)$ (1)

$Sim(x_{adv}, x) \geq \in,$ (2)

here $\in$ is the least similarity between the adversarial and original samples, and Sim: $X \times X \rightarrow (0, 1)$ is a similarity function. Sim (.) is a function often used to notice semantics and syntax similarity in the text domain.

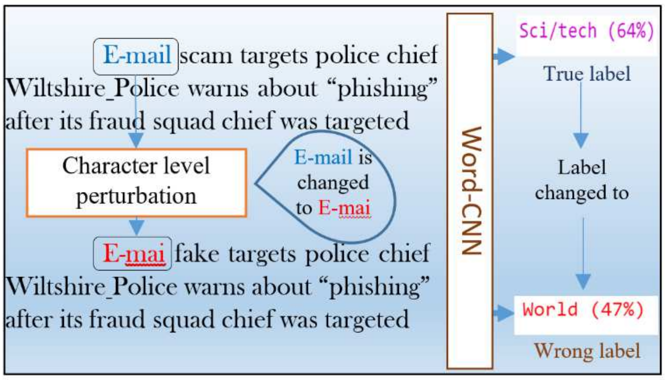

E-mail scam targets police chief
Wiltshire_Police warns about "phishing"
after its fraud squad chief was targeted
Character level perturbation
E-mail is changed to E-mai
E-mai fake targets police chief
Wiltshire_Police warns about "phishing"
after its fraud squad chief was targeted
Word-CNN
Sci/tech (64%)
True label
Label changed to
World (47%)
Wrong label

Figure 2 Adversarial example generated by *perturbing character* in a way such that it holds semantic similarity for humans but fools Word-CNN model by giving wrong output[9]

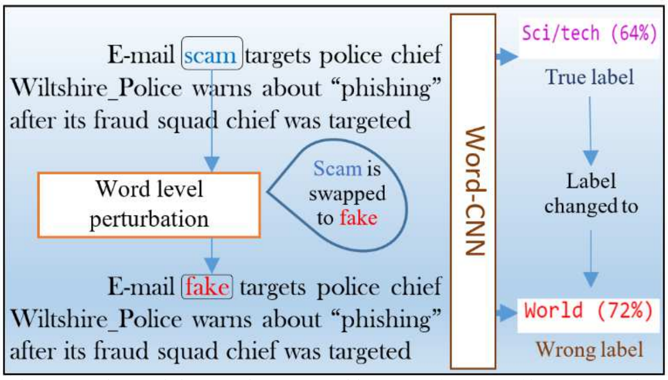

E-mail scam targets police chief
Wiltshire_Police warns about "phishing"
after its fraud squad chief was targeted
Word level perturbation
Scam is swapped to fake
E-mail fake targets police chief
Wiltshire_Police warns about "phishing"
after its fraud squad chief was targeted
Word-CNN
Sci/tech (64%)
True label
Label changed to
World (72%)
Wrong label

Figure 3 Adversarial example generated by *perturbing word* in a way such that it holds semantic similarity for humans but fools the Word-CNN model by giving the wrong output[3]

An adversarial attack in NLP can be conducted via perturbing words and characters, which includes swapping, deletion, insertion, mis-spelling, synonym replacement etc., in a way such that the underlying meaning of the sentence will not change for the human observer but fools the model in giving wrong output which can be seen from Figure 2 and Figure 3. Each adversarial attack contains four essential components, which include transformation, search method, set of constraints, and a goal function as discussed below:

- *Transformation*: A transformation that produces a number of potential perturbations from a single input.
  Examples: homoglyph character substitution, word embedding word swap, and thesaurus word swap.
- *Search method:* A search technique repeatedly probes the model and chooses promising perturbations from various transformations[4].
  Examples: beam search, genetic algorithm, greedy with word importance ranking.
- *Constraints:* A group of restrictions used to assess the validity of a perturbation in relation to the original input.
  Examples: part-of-speech consistency, minimum sentence encoding cosine similarity, grammar checker, maximum word embedding distance.
- *Goal function:* A task-specific goal function that assesses the effectiveness of the attack in terms of model outputs.
  Examples: non-overlapping output, minimum BLEU score[5], targeted classification, untargeted classification

C. Attacks Recipes

The three introduced models were manipulated using the attack recipes shown in Table 1 on the news categorization dataset to find out the vulnerability. Each attack algorithm contains different components, as discussed in Section II of this article.

Table 1 Adversarial Attack Recipes

Attack	Transformation	Search Method	Constraints	Goal function
A2T[6]	Word Swap Embedding	Greedy-Word Importance Ranking (gradient-based)	USE sentence encoding cosine similarity, Part of speech match, Word Embedding Distance	Untargeted Classification
BAE [7]	BERT Masked Token Prediction	Greedy-Word Importance Ranking	USE sentence encoding cosine similarity	Untargeted Classification
Check-List[8]	Word Swap Change Location, Word Swap Contract, Word Swap Change Name Word Swap Extend, Word Swap Change Number	Greedy Search	Repeat word Modification	Untargeted Classification

DWB [9]	{Character Deletion, Character Substitution, Character Insertion, Neighboring Character Swap} *	Greedy-Word Importance Ranking	Leven-shtein edit-distance	{Targeted, Untargeted} Classification
IGA [10]	Counter-fitted word embedding exchange	Genetic Algorithm	Word embedding distance, Percentage of words perturbed	Untargeted {Entailment, Classification }
Kulesh -ov[11]	Counter fitted word embedding exchange	Greedy word replacement	Language model similarity probability, cosine similarity, Thought vector encoding,	Untargeted Classification
Pruthi [12]	{Insertion, Keyboard-Based Character Swap, Character Deletion, Neighboring Character Swap} *	Greedy search	Maximum number of words perturbed, Minimum word length	Untargeted Classification
PSO [13]	How-Net Word replacement	Particle Swarm Optimization		Untargeted Classification
Pwws [14]	WordNet-based similar word replacement	Greedy-WIR (saliency)		Untargeted Classification
Text-bugger [15]	{Neighboring Character Swap, Deletion, Substitution, Insertion for characters} *	Greedy-Word Importance Ranking	USE sentence encoding cosine similarity	Untargeted Classification
Text-fooler [3]	Counter-fitted word embedding replacements	Greedy-Word Importance Ranking	Word Embedding Distance, Part-of-speech match, USE sentence encoding cosine similarity	Untargeted {Entailment, Classification }

III. Proposed Approach

We divided our experiment into two phases to better evaluate the weakness of models developed to categorize different types of news articles. Using the AG news classification dataset, cutting-edge deep learning models were first trained, and then adversarial approaches were used to alter the learned models' predictions. The framework of conducting an adversarial attack on news categorization models is shown in Figure 4. We evaluate the attack success rate ASR (ratio of successful attack samples to the sum of successful and failed samples) of each attack type on all three models by taking hundred test samples along with their average perturbed word percentage. The potential results of the experiment conducted are shown in Table 3, Table 4 & Table 5. Here, a successful attack means that the generated adversarial sample can give a wrong prediction with high confidence. By a failed attack, it means that the adversarial sample is not able to misclassify the actual prediction. The attack recipes applied to the targeted models are described in Section II *C* of this article.

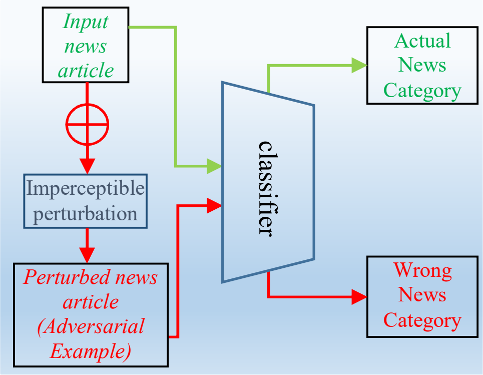

Input news article
Actual News Category
classifier
Imperceptible perturbation
Perturbed news article (Adversarial Example)
Wrong News Category

Figure 4 Framework for Adversarial Attack in News Categorization

IV. Experimental Approach

This section describes the dataset and the models used in the paper with a proper parameterization for our experimental analysis. These pre-trained models were then attacked using various *state-of-the-art* adversarial attack algorithms.

A. Dataset Description

The data used in this paper is the *AG news classification dataset* from the open-source hugging face library**.** Over a million news articles can be found in the AG corpus. A subset of AG's corpus of news items comprises the titles and descriptions of articles from the four most important categories (*World, Sports, Sci/Tech & Business*). There are 1,900 test samples and 30,000 training samples in each class of AG News. The models trained on this dataset with their descriptions are shown in Table 2.

B. Pre-trained Models

The description of the models trained along with their confidence scores are provided in the Table 2 below:

Table 2 Models Description

Model	Parameterization	Acc. Scores
Word-CNN	We employed three window sizes for the Word-CNN model: 3, 4, and 5, with 100 filters for each window size and a dropout of 0.3. Use a base of the 200-dimensional GLoVE embeddings.	**91.00%**
Word-LSTM	For the Word-LSTM, we used a 1-layer bidirectional LSTM with 150 hidden units and a dropout of 0.3. We employed 200-dimensional Glove word embeddings that had been previously trained on 6B tokens from Giga-words and Wikipedia.	**91.40%**
BERT	With a batch size of 16, a learning rate of 3e-05, and a maximum sequence length of 128 the model was tuned for five epochs. The model was trained using a cross-entropy loss function. The model's greatest performance on this task, as determined by the eval set accuracy, was 0.9514473684210526, discovered after three epochs.	**95.14%**

C. Attacking Target Models

Table 3 Attack Results on Word-CNN

Attack Recipe	Success (S)%	Failed (F)%	Skipped %	ASR= $(\frac{S}{S+F})$%	Average Perturbed Word%
Kuleshov [11]	93	1	6	98.94	15.70
TextFooler[3]	93	1	6	98.94	15.70
DWB[9]	92	2	6	97.87	18.48
Textbugger[15]	82	12	6	87.23	60.29
Pwws[14]	82	12	6	87.23	14.96
IGA[10]	69	25	6	73.40	17.86
PSO[13]	69	25	6	73.40	15.90
BAE[7]	24	70	6	25.53	06.39
A2T[6]	22	72	6	23.40	09.25
pruthi[12]	14	80	6	14.89	02.82
Checklist[8]	8	86	6	08.51	06.44

Table 4 Attack Results on Word-LSTM

Attack Recipe	Success (S)%	Failed (F)%	Skipped %	ASR= $(\frac{S}{S+F})$%	Average Perturbed Word%
TextFooler [3]	88	4	8	95.65	16.84
Kuleshov[11]	85	4	11	95.51	10.02
DWB[9]	78	14	8	84.78	18.50
PSO[13]	70	22	8	76.09	18.06
Pwws[14]	70	22	8	76.09	16.30
Textbugger [15]	68	24	8	73.91	53.41
IGA[10]	69	25	6	73.40	15.90
BAE[7]	25	67	8	27.17	05.60
A2T[6]	20	72	8	21.74	08.16
pruthi[12]	12	80	8	13.04	02.95
Checklist[8]	2	90	8	02.17	08.73

Table 5 Attack Results on BERT

Attack Recipe	Success (S)%	Failed (F)%	Skipped %	ASR= $\left(\frac{S}{S+F}\right)$%	Average Perturbed Word%
Kuleshov[11]	96	3	1	96.90	07.16
PSO[13]	95	4	1	95.95	20.06
TextFooler [3]	91	8	1	91.92	15.22
Pwws[14]	84	15	1	84.85	11.28
DWB[9]	83	16	1	83.84	15.31
Textbugger [15]	79	20	1	79.80	19.32
IGA[10]	77	22	1	77.78	12.28
BAE[7]	33	66	1	33.33	07.53
pruthi[12]	30	69	1	30.33	03.46
A2T[6]	29	70	1	29.29	08.95
Checklist[8]	2	97	1	02.02	36.59

V. Results & discussion

The attack success rate of all the attacks is evaluated on all the targeted models to find the most and least vulnerable model. We formulate the average attack success rate on each model using the formulae given below:

$$\boldsymbol{S_r} = \frac{\sum_{i=1}^{a} \frac{S_i}{S_i+F_i}}{a} \qquad (3)$$

$$\boldsymbol{S_r} = Attack\ Success\ rate\boldsymbol{;}\ \boldsymbol{a} = attack;\ \boldsymbol{S_i} = successful\ attack;\ \boldsymbol{F_i} = Failed\ attack$$

(The Attack Success Rate is $\boldsymbol{S_r}$, the no. of successful attacks is $\boldsymbol{S_i}$, the no. of unsuccessful attacks is $\boldsymbol{F_i}$, and the no. of attack recipes is $\boldsymbol{a}$. Since the skipped statements depend on model training and not Attacks, they were not included in the calculation. The statements the model initially incorrectly anticipated during its training can be seen by looking at the skipped values. We were interested in the success rates and the effectiveness of Attacks in misclassifying outputs.)

Table 6 ASR on different Models

Model	Success Rate%
Word-CNN	62.66
Word-LSTM	58.14
Bert	64.18
Total Average	61.66

As shown in Table 6, On evaluation, it has been found that the model *Bert* is most vulnerable to adversarial attacks besides its best accuracy among all models. It also shows a trade-off between robustness and accuracy. The word-LSTM model is the least vulnerable as compared to the other two. Top ranking Attacks from which models are more vulnerable are shown in Table 7, and it has been noticed that the 2 top ranking attacks are word-level attacks.

Table 7 Mean Success rate of particular attack type on all the Models

Attack Recipe	Success %	Level	Average Perturbed Word%
Kuleshov[11]	97.11	word	10.96
Textfooler [3]	95.50	word	15.92
DWB[9]	88.83	character	17.43
Pwws[14]	82.72	word	14.18
Textbugger [15]	80.31	word & character	44.34
BAE[7]	28.67	word & character	06.50
Pruthi[12]	19.42	character	03.07
A2T[6]	24.81	word	08.78
IGA[10]	74.86	word	15.34
PSO[13]	81.79	word	18.00
Checklist[8]	04.23	word	17.08

Word-level attacks are the most effective types. It appears that word-level attacks are more likely to affect the models than character-level or mixed attacks (character & word).

VI. Conclusion

This research set out to address the question, how susceptible to adversarial attacks is automatic news subject classification? This was put to the test by seeing if the majority of adversarial attack algorithms could identify the accurate forecast of news categories incorrectly. The findings unequivocally demonstrate that text classification of news articles may be circumvented by simply changing the words and characters for machine learning models while maintaining semantic similarity for human observers. We achieved an average attack success rate of 61.66% on all three models. The model BERT is considered to be a powerful transformer model that achieved a maximum confidence score of 95.14%, but it is the most vulnerable having an ASR of 64.18% as compared to Word-CNN and Word-LSTM.

The accuracy scores of Word-CNN & Word-LSTM are 91% and 91.4%, respectively, with ASR of 62.66 for Word-CNN and 58.14 for Word-LSTM, clearly showing that the Word-LSTM model is least vulnerable to adversarial manipulations. Hence, on our final evaluation, it has been found that the Word-LSTM model should be chosen if one is more concerned about adversarial manipulations, as it is more robust as compared to the other two models. Also, the model which achieved maximum accuracy is the most vulnerable, i.e., BERT. This shows a contradictory result between accuracy and robustness. Overall, it has been proved it is possible to obfuscate the automatic news classification models with adversarial modifications. Hence, there is a dire need for adversarial robust generalizations instead of standard generalizations for the betterment of society. This paper motivates the readers to also look for models which are more robust to adversarial attacks instead of only going for better confidence scores.

REFERENCES

[1] A. Peng, K. Deng, J. Zhang, S. Luo, H. Zeng, and W. Yu, “Gradient-Based Adversarial Image Forensics,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, vol. 12533 LNCS, doi: 10.1007/978-3-030-63833-7_35.

[2] P. Sunagar, A. Kanavalli, S. S. Nayak, S. R. Mahan, S. Prasad, and S. Prasad, “News Topic Classification Using Machine Learning Techniques,” in *Lecture Notes in Electrical Engineering*, 2021, vol. 733 LNEE, doi: 10.1007/978-981-33-4909-4_35.

[3] D. Jin, Z. Jin, J. T. Zhou, and P. Szolovits, “Is BERT Really Robust? A Strong Baseline for Natural Language Attack on Text Classification and Entailment,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, Jul. 2019, pp. 8018–8025, [Online]. Available: http://arxiv.org/abs/1907.11932.

[4] J. Y. Yoo, J. X. Morris, E. Lifland, and Y. Qi, “Searching for a Search Method: Benchmarking Search Algorithms for Generating NLP Adversarial Examples,” in *Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, 2020, pp. 323–332, [Online]. Available: https://github.com/QData/TextAttack.

[5] K. Papineni, S. Roukos, T. Ward, and W. Zhu, “BLEU : a Method for Automatic Evaluation of Machine Translation,” *Comput. Linguist.*, no. July, 2002.

[6] J. Y. Yoo and Y. Qi, “Towards Improving Adversarial Training of NLP Models,” 2021, doi: 10.18653/v1/2021.findings-emnlp.81.

[7] S. Garg and G. Ramakrishnan, “BAE: BERT-based Adversarial Examples for Text Classi□cation,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 6174–6181.

[8] M. T. Ribeiro, T. Wu, C. Guestrin, and S. Singh, “Beyond Accuracy: Behavioral Testing of NLP models with CheckList,” *ACL 2020 - 58th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (Long Pap.*, pp. 4902–4912, 2020.

[9] J. Gao, J. Lanchantin, M. Lou Soffa, and Y. Qi, “Black-box generation of adversarial text sequences to evade deep learning classifiers,” in *Proceedings - 2018 IEEE Symposium on Security and Privacy Workshops, SPW 2018*, 2018, pp. 1–21, doi: 10.1109/SPW.2018.00016.

[10] X. Wang, H. Jin, Y. Yang, and K. He, “Natural Language Adversarial Defense through Synonym Encoding,” 2021.

[11] V. Kuleshov, S. Thakoor, T. Lau, and S. Ermon, “Adversarial Examples for Natural Language Classification Problems,” in *ICLR 2018 : International Conference on Learning Representations*, 2018, vol. 142, no. 3.

[12] D. Pruthi, B. Dhingra, and Z. C. Lipton, “Combating adversarial misspellings with robust word recognition,” 2020, doi: 10.18653/v1/p19-1561.

[13] Y. Zang *et al.*, “Word-level Textual Adversarial Attacking as Combinatorial Optimization,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 6067–6080, doi: 10.18653/v1/2020.acl-main.540.

[14] S. Ren, Y. Deng, K. He, and W. Che, “Generating natural language adversarial examples through probability weighted word saliency,” 2020, doi: 10.18653/v1/p19-1103.

[15] J. Li, S. Ji, T. Du, B. Li, and T. Wang, “TextBugger: Generating Adversarial Text Against Real-world Applications,” in *26th Annual Network and Distributed System Security Symposium*, 2019, pp. 1–15, doi: 10.14722/ndss.2019.23138.

Forecasting of Satellite Based Carbon-Monoxide Time-Series Data Using a Deep Learning Approach

Abhishek Verma
Biometric Research Laboratory,
Department of Information Technology,
Delhi Technological University
Bawana Road, Delhi-110042, India
Abhishekcms08@gmail.com

Virendar Ranga
Department of Information Technology,
Delhi Technological University
Bawana Road, Delhi-110042, India
drvirender.ranga@gmail.com

Dinesh Kumar Vishwakarma
Biometric Research Laboratory,
Department of Information Technology,
Delhi Technological University
Bawana Road, Delhi-110042, India
dvishwakarma@gmail.com

2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT) | 978-1-6654-9414-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICITIIT57246.2023.10068609

Abstract— **In last few decades one of the major problems is air pollution which has raised the eyebrows of everyone. Despite all the efforts, it still lies in the category of dangerous. In air pollution there is one of the most hazardous gases named carbon monoxide which is a matter of concern & produced mostly whenever a material burns with a lack of oxygen. This paper presents the forecasting of carbon monoxide with the help of a satellite-based sentinel 5p dataset using earth engine. Further, with the help of the deep learning approach 'LSTM', we forecast a time series base result. We have trained and tested the data using a deep-learning model. We have evaluated the potential results by overlapping the original and predicated values and calculating Root-mean-square (RMS) error to validate our approach. The results show that the method of LSTM is very efficient and accurate.**

Keywords— Carbon-Monoxide, Earth Engine, Sentinel 5p, RMS, LSTM.

I. Introduction

Air pollution is one of the most threatening problems for health and environmental domain experts [1]. *Air quality* is an issue that could be more concise to a particular area. However, it is affected globally, and one of the most dangerous gases with the most adverse effect is Carbon-monoxide. The incomplete burning of carbonaceous fuels, including gasoline, natural gas, petroleum, coal, and wood, results in the production of carbon monoxide (CO), a colorless, odorless, tasteless, and deadly atmospheric pollutant. The primary anthropogenic source of carbon dioxide in India is vehicle exhaust. High CO concentrations, typical of polluted surroundings, impair hemoglobin's ability to carry oxygen (O2), which can negatively affect one's health. These impacts include headaches, an increased risk of chest pain for people with heart conditions, and delays in response time. Given its success with time series data, the recurrent neural network (RNN) model known as long short-term memory (LSTM) is utilized for this purpose. This research adds something unique by examining LSTM RNN, a deep-learning method for airborne missions [2]. Quality prediction looks only at machine learning algorithms—satellite-based air quality data. Copernicus Sentinel 5 (European union) collected and made available data as part of the Open Data Initiative. Time series data helps in future furcating data by the black box method of deep learning.

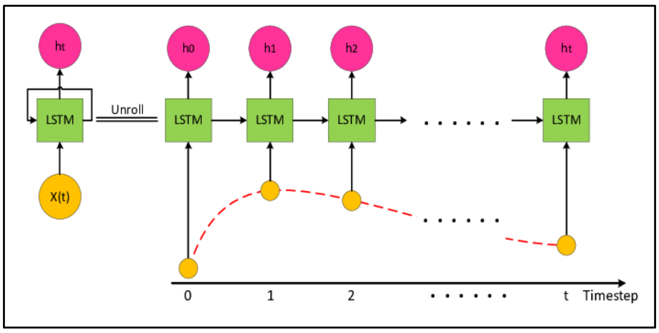


Figure 1 LSTM input-based model

In this work, we present how forecasting carbon monoxide is done with the help of the deep learning approach LSTM. This paper will aim to study the LSTM RNN models' performance for air quality forecasting. Figure 2 shows how carbon monoxide is visible in the study area's spatial view.

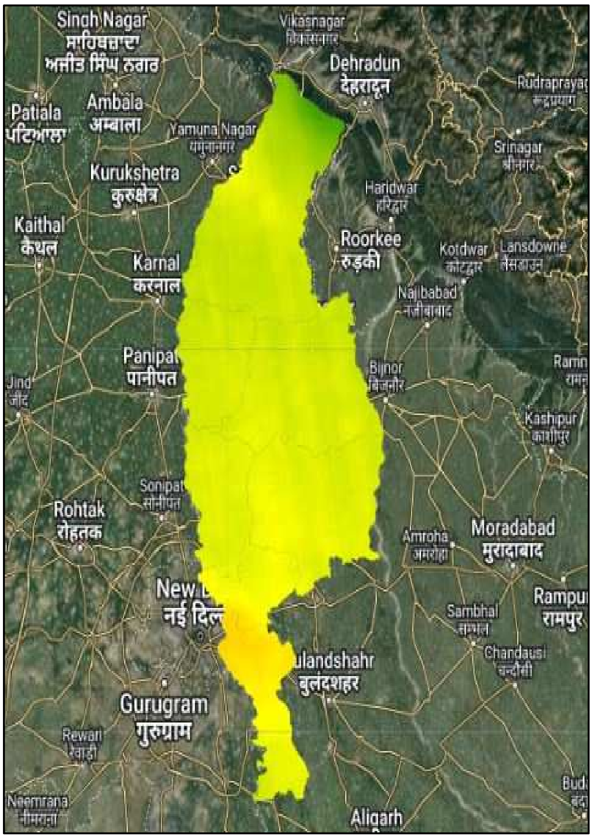


Figure 2 Carbon-monoxide spatial view of study area

978-1-6654-9414-4/23/$31.00 ©2023 IEEE

II. Related Work

A. *Concentraing air quality forecasting using LSTM and Wireless sensor networks*

Collecting real-time air quality and air quality forecasts is crucial to taking preventative and corrective measures because many metro cities around the globe have recently experienced significant air pollution and related health risks. Very significant. In order to monitor and gather instantaneous data on air pollution concentrations from various spots in the area and use this data to predict air pollutant concentrations, this study suggests a scalable architecture. We gathered information about air quality from two sources. At primary is a wireless sensor network with sensor nodes dispersed around Bengaluru, a city in southern India, that gathers pollution concentrations and sends them to servers The Ministry of Environment, Forests, and Climate Change's Open Data Initiative is the second source of real-time air quality data, which it has gathered and made accessible. The average hourly concentrations of several air contaminants are provided by both sources. Due to its expertise with time series data, a Long Short- Term Memory (LSTM) Recurrent Neural Network (RNN) model was chosen to complete the air quality prediction assignment. This white paper examines model performance in two areas that exhibit distinct variations in air quality over time. Model performance suffers as these oscillations grow, necessitating adaptive modelling.

B. *Predicting Air Quality with Deep Learning LSTM: Towards Comprehensive Models*

The problem of long-term forecasting of air quality in the Madrid region. A recurrent artificial neural network for short-term memory. The air quality in this study was
Concentrations of many air pollutants that have been shown to be harmful to health.CO, NO2, O3, PM10, SO2 and two genera (Plantago and Poaceae). These concentrations have been sampled at many locations within the city of Madrid. The problem of long-term forecasting of air quality in the Madrid region A recurrent artificial neural network for short-term memory. The air quality in this study was Concentrations of many air pollutants that have been shown to be harmful to health. CO, NO2, O3, PM10, SO2 and two genera (Plantago and Poaceae). These concentrations have been sampled at many locations within the city of Madrid. Excluding that Training a set of models, one per site, contaminants, multiple comprehensive deep networks Compare configurations to identify suitable configurations for extracting relevant information Predict daily air quality from a set of time series. Results supported by Statistical evidence suggests that a single comprehensive model may be a better option several individual models. Such a comprehensive model represents a successful tool that can: For example, it provides useful forecasts that can be used in a managed environment. Clinical facilities optimizing resources in anticipation of increased patient numbers due to exposure to poor air quality.
So, after reading this paper what we have found that the approaches are not applicable on satellite data and long-term data .so well will perform the experiment on the satellite data.

III. Proposed Approach

We have done this experiment into two phases firstly we have used google earth engine[3 [2]] to extract the latest data which is available ,so since the sentinel 5p has been an active satellite from June 2019 so the data has been taken from the earliest available date to latest available data that is 29 October 2022 then further extracting the data from with required pre-processing on earth engine to generate a timeseries data then after that we will use deep learning approach LSTM to forecast the data. Further in LSTM we trained 80 % data and then evaluated and checked our result on 20 % of the data. We have to then compare both predicted values and actual values to see whether predicted values overlap actual value

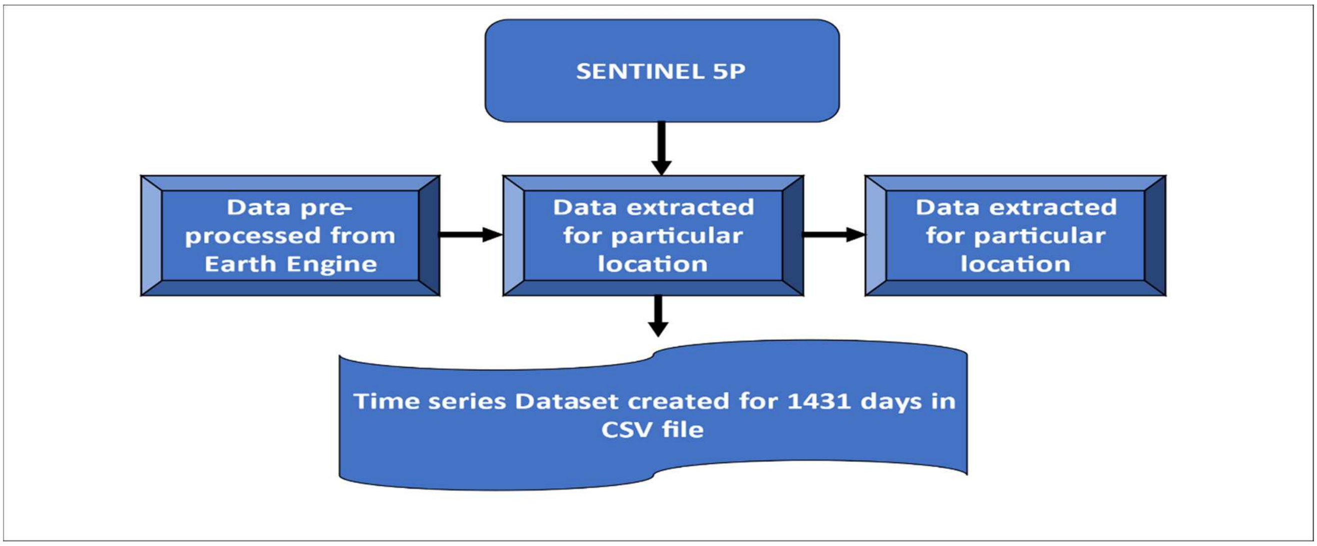

SENTINEL 5P
Data pre-processed from Earth Engine
Data extracted for particular location
Data extracted for particular location
Time series Dataset created for 1431 days in CSV file

Figure 3 Methodology for obtain satellite-based sentinel 5 CO data

978-1-6654-9414-4/23/$31.00 ©2023 IEEE

IV. Experimental Approach

This section describes the dataset and the models used in the paper with a proper parameterization for our experimental analysis.

A. Dataset Description

The data used in this paper is the is Copernicus sentinel 5p dataset which is openly available for public use, by the preprocessing of the data set Carbon monoxide and the data is extracted in form of time series. the data which is provided it is the pollution level of tropospheric level which can be further gives the estimation of hotspot and major regions of pollution.

B. Model Used A Long Short term memory

The Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) architecture is used in the prediction model. Building blocks for LSTM RNN layers are LSTM units. The block state, which saves the data that the LSTM unit contains, is a crucial component of the LSTM unit. Gates, which have the ability to add or delete information from the cell state, are the structures that govern LSTM units. Additionally, LSTM units have a hidden state that stores prior knowledge about the observed sequence. There are 3 gates in the LSTM unit. “Entry Gate," "Forgotten Gate," and "Exit Gate." The following characteristics of these gates:

1. Based on the forget gate weight, hidden state, and current input, the forget gate chooses which portion of the cell state to delete.
2. The weights of the hidden state, the input gate and the current input are used to decide how much additional information to add to the cell's state.
3. The output gate analyses the current input, the hidden state, and the weights of the output gate to determine how the state of the cell affects the current output. When deciding what should be added to the cell state, what should be subtracted from the cell state, and how much the cell state influences the output of the LSTM unit, forgetting, input, and output gate weights are crucial.

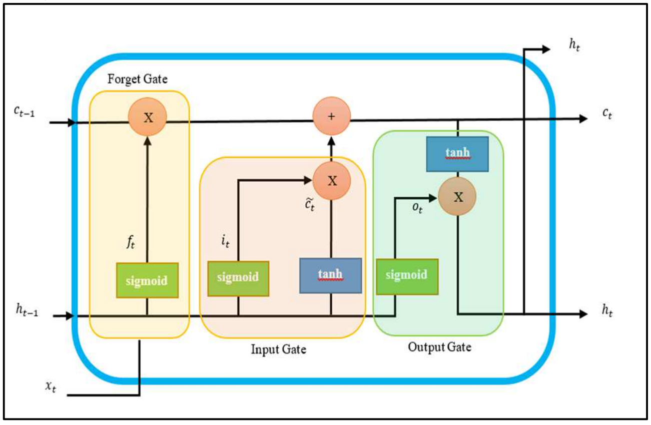


Figure 4 LSTM working architecture

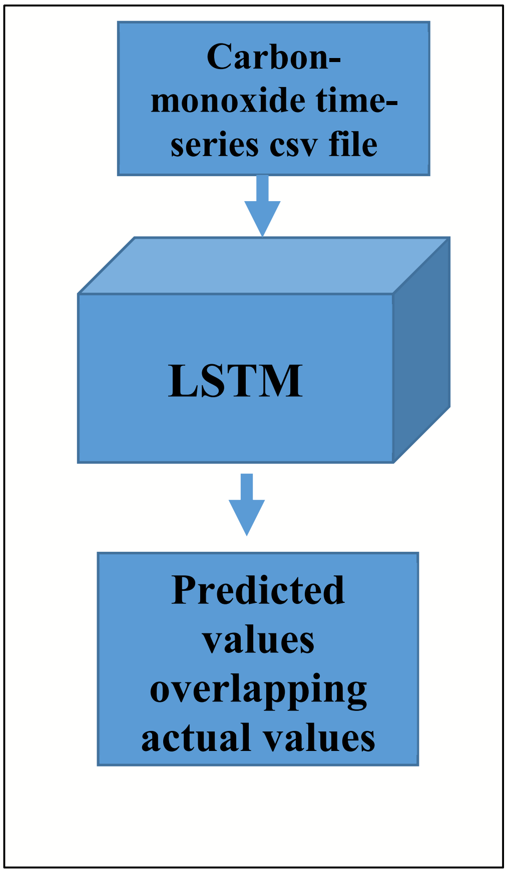
C. Training Process

The model is trained using the data produced after interpolation. The model was trained on 80% of this data, while the remaining 20% was utilized to test the model [5]. The training procedure reduces the mean squared error—which is the loss—between the anticipated data and the actual data. For training, Adam optimizer is utilized. In this optimization approach, the network weights are iteratively modified based on the training data. Each network parameter's learning rate is kept separate and adjusted when new information is learned. The network was trained across 5 epochs with early stopping, which halts the training procedure if the model starts to over-fit, that is, if train loss starts to decline while test loss starts to rise.

D. Prediction process

So, we predict values after training is completed using already trained 600 values from the trained data and observed that it almost overlaps with the actual value.

V. Result



Carbon-monoxide time-series csv file
LSTM
Predicted values overlapping actual values

Figure 5 LSTM Forecasting processing

So as in Figure 6, we can see that the orange line is an overlapping blue line, so the result represents that the predicted values are lying on actual values very princely. The experiment is a success. We trained 1100 values which have given LSTM units as 100 and 5 epochs[4], resulting in almost accurate and overlapping predicted values over actual values as shown in figure 6, and the result RMS is 0.0036.

978-1-6654-9414-4/23/$31.00 ©2023 IEEE

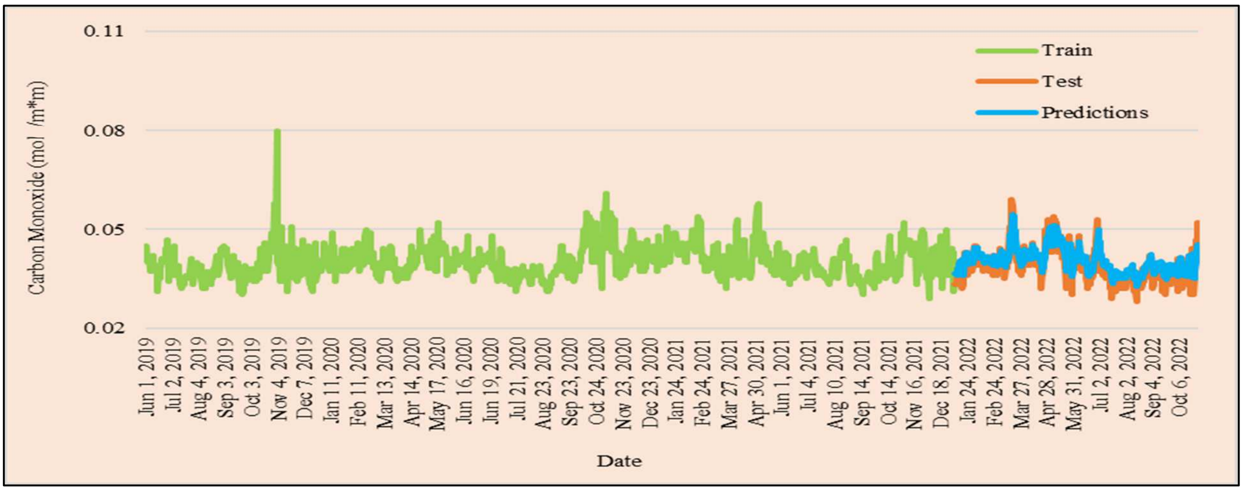


Figure 6 Training, Testing and Predictions using LSTM

VI. References

[1] G. Petnehazi, "Recurrent neural networks for time series forecasting," p. 1901.00069, 2019..

[2] R. K. S. J. K. B. R. S. a. J. S. ". K. Greff, "LSTM: A search space odyssey".," *IEEE Transactions on Neural.*

[3] G. G. Kaplan and Z. Yigit Avdan, *"Space-borne air pollution observation from Sentinel-5p Tropomi: relationship between pollutants, geographical and demographic data", International Journal of Engineering and Geosciences, ,* Vols. vol. 5, no. 3, , pp. pp. 130-137, , Oct. 2022.

[4] D. C. C. Y. T. &. Z. X. Zhu, " "A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization."".

[5] Y. B. M. M. V. S. C. a. B. Zhang, "A. Real-time air quality forecasting, part I: History, techniques, and current," *part I: History, techniques, and current.*

2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT) | 978-1-6654-5357-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCT56969.2023.10075941

Identification and Screening of Novel ACE Inhibitors using Computational Approach

Murali Mohan Mishra
Dept. of Biotechnology
Molecular Neuroscience and Functional Genomics Laboratory,
Delhi Technological University
Delhi – 110042, India
muralimohanmishra_2k21ibt13@dtu.ac.in

Pravir Kumar*
Dept. of Biotechnology
Molecular Neuroscience and Functional Genomics Laboratory,
Delhi Technological University
Delhi – 110042, India
pravirkumar@dtu.ac.in

***Abstract*-Use of acetylcholinesterase (AChE) inhibitor in treating the neurological disorders has long been studied due to its potential to cross the endothelial tight junctions, longer bioavailability, and better ability to penetrate skin. Alzheimer's disease is found to have closely related with the decline in the level of neurotransmitters which leads to deterioration of the cholinergic neurons of the neocortex and the hippocampus of the rat's brain. Impairment in the transmission of cholinergic nerve signals results in the formation of senile plaque and neurofibrillary tangles (NFT). As a result, one of the main goals for the development of therapeutic approaches for Alzheimer's disease has been to improve the cholinergic activities of the brain. The discovery of one of the most efficient acetylcholinesterase inhibitors called Donepezil was proved to be a much better approach as compared to other drugs such as physostigmine and Tacrine. In the present study we have focused on the role of 5,6-dimethoxy-2-(piperidin-4-ylmethyl)-2,3-dihydroinden-1-one as an important acetylcholinesterase in the treatment of Alzheimer's disease. We have performed molecular docking to see the interaction of ACE target protein and the inhibitory ligands and further validated the pharmacokinetic properties of the drug via ADME analysis of the drug.**

Keyword- acetylcholinesterase, cholinergic neurons, neocortex, neurofibrillary tangles (NFT), Donepezil, 5,6-dimethoxy-2-(piperidin-4-ylmethyl)-2,3-dihydroinden-1-one.

I. INTRODUCTION

Alzheimer's disease is a fast progressive neurodegenerative disease featured by various lunatic and physiological disorders. It is found to have associated with decline in various neurotransmitter such as acetylcholine (Ach), serotonin and norepinephrine (NE)[1]. The cholinergic neurons in the hippocampus and cortex degenerate as a result of the empirical decline, which furthers the decrease of cholinergic transmission. [2]. Disturbance in the cholinergic transmission process also results in the formation of amyloid plaques and the neurofibrillary tangles (NFT)[3]. Acetylcholinesterase is found to have an important role in the progression of amyloid plaques.

Mainly two hypotheses are developed when dealing with Alzheimer's disease including – "Cholinergic hypothesis" and the "Amyloid hypothesis". The idea of Cholinergic hypothesis is focused on the importance of impaired cholinergic functionality in the hippocampus and neocortex region of brain responsible for memory, learning, emotional and behavioural responses[4]. On the other hand, Amyloid hypothesis is based on the fact that acetylcholinesterase produces secondary non-cholinergic roles including increase in the accumulation of β-amyloid peptides in the form plaques or the neurofibrillary tangles in the patients of Alzheimer's disease[5].

Atrophy of brain tissues is considered to be a most important physiological consequence of the Alzheimer's disease. This atrophy is characterized by the reduction in the amount of neurotransmitter such as acetylcholine which are accountable for the transmission of electrical impulsive signal from nerve cell to another due to the hydrolysis by acetylcholinesterase enzyme[6]. An effort to enhance the brain's cholinergic activity has been a major therapeutic target for the Alzheimer's disease treatment[7]. Inhibitors of cholinesterase increases the transmission of cholinergic signals by inhibiting the acetylcholine and butyrylcholine hydrolysing enzymes.

Butrylcholinesterase is an enzyme nearly associated with acetylcholinesterase and have prominent role in the regulation of cholinergic neurotransmission by the hydrolysis of acetylcholine. It has been reported that butrylcholinesterase activity get increased by 40 to 90% in the hippocampus and temporal cortex area of the brain during development of Alzheimer's disease[4]. The increased activity of butrylcholinesterase has a predominant role in the accumulation of amyloid-β peptides during the initial phase of plaque formation. From the above findings it has become clear that both the enzymes i.e., acetylcholinesterase and butrylcholinesterase can be recorded as an important target for managing the Alzheimer's disease by enhancing the acetylcholinesterase availability in the severally effected regions of brain and decreasing the deposition of amyloid-β peptides. One of the major limitations of butrylcholinesterase is the restricted localization of the enzyme in the outlying tissues such as plasma and in lower amount in the brain. Further, acetylcholinesterase is favoured over butrylcholinesterase due to fewer side-effects because of peripheral inhibition of cholinesterase enzyme[8].

Use of a wide range of compounds having cholinesterase inhibiting characteristics was engaged to increase the cholinergic activity. The first acetylcholinesterase inhibitor being discovered was Tacrine, but due to its adverse side effects such as hepatotoxicity, peripheral adverse effects, and the multiple dosing of the drug its usage was discontinued[9]. Later, a second class of acetylcholinesterase inhibitor called Physostigmine was discovered but due to its limitations such as frequent dosing regimen and severe side effects it was also discontinued[10].

To cater the above challenges of Tacrine and physostigmine a chain of indanone derivatives were screened and synthesized. N-Benzylpiperazine, one of the major groups, was discovered to produce significant effects in rats. The most enhanced inhibiting activity was observed in the case of cyclic amide derivatives. Among various derivatives of indanone, Donepezil hydrochloride also called Aricept® was found to have better inhibitory effect and balanced strength[11]. Since Donepezil was selective for acetylcholinesterase over the butrylcholinesterase, the peripheral side effects of Donepezil were expected to be lower as compared to other drugs. Donepezil was reported with a half-life of 70 to 80 hours whereas for the other two drugs it ranged between 0.3 to 12 hours[12]. This longer half-life of the Donepezil allowed convenient dosing of one dose per day.

II. MATERIALS AND METHODS

A. *Drug target exploration*

The possible targets of the drug Donepezil were searched from drug Bank Database (https://go.drugbank.com/) . Out of all the targets one with the inhibitory results was chosen for the further studies. It is an online, comprehensive and an open access database providing information on the drugs and their respective targets. It is a novel bioinformatics/cheminformatics tools that recombines the chemical details of the drug molecules with the comprehensive protein target of the drug molecules. Drug bank database is a manually created database that is maintained by University of Alberta and The Metabolomics Innovation Center located in the Alberta, Canada. The frequency of data release follows every two-year release with updates and corrections on the monthly basis. The drug ban database contains more than 4100 drug entities among which more than 800 entities are FDA approved small sized molecules and more than 3200 drugs are based on experimental validation.

A. *Search for similar molecules*

All the drug molecules similar to Donepezil was searched from the PubChem Database (https://pubchem.ncbi.nlm.nih.gov/) . In all we encountered 1000 similar structures. On further sorting of the drug molecules by applying various filters such as molecular weight count and heavy atom count within the respective minimum range we found eight major drug molecules possessing structural similarity with Donepezil.

B. *Target structure retrieval*

The main target of the selected drug molecule i.e., Acetylcholinesterase (ACE) was searched on the Uniprot database (https://www.uniprot.org). The 3D structure of the acetylcholinesterase was saved in the .pdb format.

C. *Target and Ligand preparation*

For the target preparation the water and the ligand molecules of the ACE receptor was removed using Discovery studio and the structure was saved in .pdb format. For the ligand's preparation, structure of all the eight drug molecules was save in .mol2 format using discovery studio.

D. *Swiss Dock*

Swiss dock is web based open access server used for the docking of protein and ligands. The target protein structure in .pdb format and the ligands structure in .mol2 format were uploaded on the server (http://www.swissdock.ch/docking) and the docking was conducted.

E. *Protein-ligand complex analysis*

After the termination of docking process, all the structures of target-ligands interaction were saved in .chimera format. All the interactions were analyzed using UCSF Chimera Software (Version 1.6).

F. *ADME Analysis*

For all the eight selected drug molecules ADME analysis was done using online open access tool called SWISS ADME (http://www.swissadme.ch/index.php). The major parameters for the analysis were- GI absorption, blood-brain-barrier permeability, Lipinski's Rule, Violations, water solubility, lipophilicity, and the bioavailability.

III. RESULTS

A. *PPI interaction of ACE receptor and AD responsible genes*

The PPI interaction of ACE and other genes responsible in AD showed few important interactions. Using STRING dataset stat analysis, the p-value for PPI enrichment was determined to be 1.0e-16. For the given proteins, an average local clustering coefficient of 0.633 was found. Edges represent protein-protein associations. Known Interactions were show by pink and light-blue lines, Predicted Interactions were shown by green red and navy-blue lines and other interactions were shown by yellow and black lines (shown in figure 1). Total number of nodes were 39and total number of edges were 190. The results showed that there are enormous interactions between the ACE and other genes responsible for the Alzheimer's Disease.

A. *Drug target analysis*

A total of eight potential targets were identified for the Donepezil from the Drug Bank Database. Among these targets ACE was found to have major impact on the amyloid-β accumulations with an inhibitory action.

B. *Interaction between ACE and Inhibitory ligands*

The docking results showed a favorable interaction between ACE and the ligands. The full fitness energy and the respective ΔG of the docked protein-ligand complex showed that the ACE and the ligand inhibitor has an important role in the amyloid-β clearance. The most negative ΔG corresponded to the most stable interaction of the target-ligand complex.

TABLE I. BINDING ENERGY OF POTENTIAL DRUG MOLECULES AND TARGET ACE PROTEIN

S. No	Compound CID	Full Fitness Energy (kcal/mol)	Evaluated ΔG (kcal/mol)
1	10446897	-2468.29	-7.53
1	15626614	-2480.66	-7.34
2	2275555	-2465.53	-7.65
3	22311400	-2460.81	-7.4
4	21808365	-2469.23	-7.47
5	59360150	-2469.66	-7.39
6	161784086	-2461.36	-7.54
7	59863314	-2487.99	-9.04

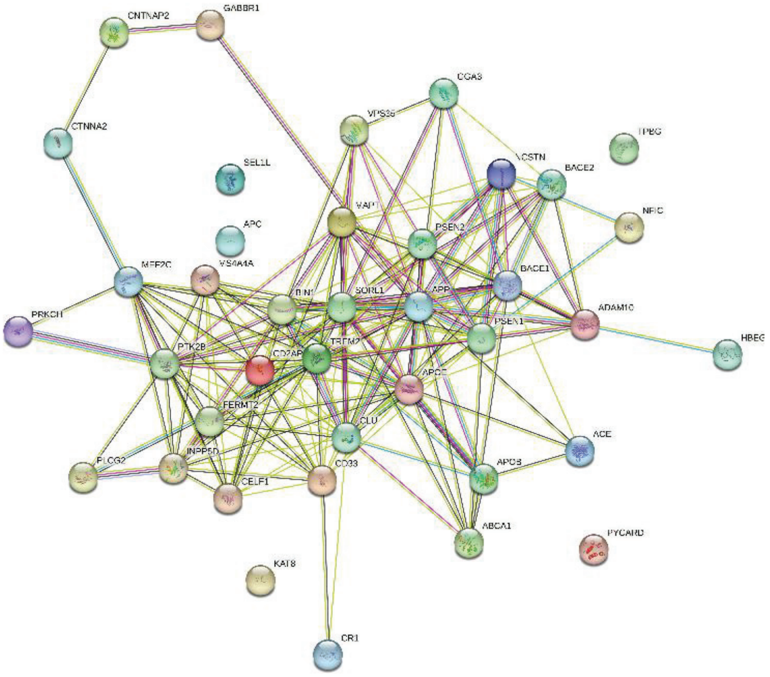


Fig. 1. PPI Interaction of ACE receptor and other AD related protein

The most negative ΔG was found to be for 5,6-dimethoxy-2-(piperidin-1-ium-4-ylmethyl)-2,3-dihydroinden-1-one with the compound CID- 59863314 with cluster 7and element 2 that is -9.04Kkcal/mole and the full fitness energy of - 2487.99kcal/mole (Table2). Using UCSF Chimera, structural analysis of the complex was done as shown in Table-2.

TABLE II. FULL FITNESS AND ΔG OF THE DRUG CANDIDATES WITH MOST STABLE INTERACTIONS

Cluster	Element	Full Fitness (kcal/mol)	Estimated ΔG (kcal/mol)
7	2	-2487.99	-9.04
7	0	-2488.10	-8.97
7	1	-2488.01	-8.96
7	5	-2486.76	-8.86
7	3	-2487.27	-8.83

C. ADME analysis of 5,6-dimethoxy-2-(piperidin-1-ium-4-ylmethyl)

The ADME analysis for the drug molecule showed positive pharmacokinetics results. The GI absorption values was recorded to be significantly higher for the drug along with superior capability to penetrate the endothelial tight junctions of brain. It followed the Lipinski's rule, and no violation were observed for the drug. Further the lipophilicity of the drug was recorded to be 2.32 (Log Po/w (XLOGP3)). The bioavailability score was 0.55. The skin permeability (Kp) of the rug was also recorded to be quite significant with the value of Log-6.42 cm/s. The BOILED EGG image of the drug is shown in the figure 2.

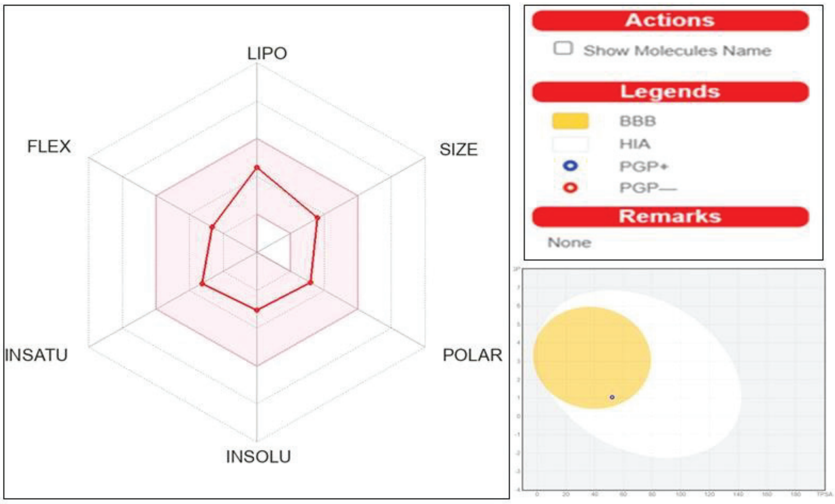


Fig. 2. BOILED EGG image for ADME analysis of 5,6-dimethoxy-2-(piperidin-1-ium-4-ylmethyl)

IV. CONCLUSION

Alzheimer's disease is marked by decline in the level of various neurotransmitters which leads to degeneration of Cholinergic neuron in the hippocampus and the cortex region of the rats resulting in the decline in cholinergic transmission. Disturbance in the cholinergic transmission leads to formation of NFT and senile amyloid plaques. Further, an atrophy of the brain tissues is also responsible for the decreased level of neurotransmitter. Acetylcholinesterase and butrylcholinesterase both regulate the cholinergic transmission, but ACE inhibitor is favored over BCHE inhibitor due to some of the major limitations such as limited localization of the enzyme in the outlying tissues such as plasma and in least amount in the brain[8]. The PPI of ACE receptor protein with other potential AD related genes showed that a strong network exists between these genes that are responsible for the onset and the development of Alzheimer's Disease. Hence, inhibition of these enzymes has become a potential target for the treatment modalities of Alzheimer's disease.

Recently discovered drug called Donepezil was preferred over previously discovered drugs called Tacrine and Physostigmine. We found a molecule called 5,6-dimethoxy-2-(piperidin-1-ium-4-ylmethyl) having more than 90% similarity with the conventional Donepezil structure. Further, on performing the molecular docking this protein with the ACE inhibitor protein target we found that the interactions should a favourable binding thus suggesting that 5,6-dimethoxy-2-(piperidin-1-ium-4-ylmethyl) can be used an important therapeutic agent in the treatment of Alzheimer's disease. ADME results further confirmed the better pharmacokinetic properties of the drug molecule.

REFERENCES

[1] E. L. Barner and S. L. Gray, "Donepezil use in Alzheimer disease," *Ann. Pharmacother.*, vol. 32, no. 1, pp. 70–77, 1998, doi: 10.1345/APH.17150.

[2] N. R. Cutler and J. J. Sramek, "The Role of Bridging Studies in the Development of Cholinesterase Inhibitors for Alzheimer's Disease," *CNS Drugs 1998 105*, vol. 10, no. 5, pp. 355–364, Aug. 2012, doi: 10.2165/00023210-199810050-00005.

[3] R. S. Doody, "Clinical profile of donepezil in the treatment of Alzheimer's disease," *Gerontology*, vol. 45 Suppl 1, no. SUPPL. 1, pp. 23–32, 1999, doi: 10.1159/000052761.

[4] P. Anand and B. Singh, "A review on cholinesterase inhibitors for Alzheimer's disease," *Arch. Pharmacal Res. 2013 364*, vol. 36, no. 4, pp. 375–399, Feb. 2013, doi: 10.1007/S12272-013-0036-3.

[5] A. Castro and A. Martinez, "Peripheral and dual binding site acetylcholinesterase inhibitors: implications in treatment of Alzheimer's disease," *Mini Rev. Med. Chem.*, vol. 1, no. 3, pp. 267–272, Mar. 2001, doi: 10.2174/1389557013406864.

[6] C. J. Ladner and J. M. Lee, "Pharmacological drug treatment of Alzheimer disease: the cholinergic hypothesis revisited," *J. Neuropathol. Exp. Neurol.*, vol. 57, no. 8, pp. 719–731, 1998, doi: 10.1097/00005072-199808000-00001.

[7] H. Brodaty, "Realistic expectations for the management of Alzheimer's disease," *Eur. Neuropsychopharmacol.*, vol. 9 Suppl 2, no. SUPPL. 2, Apr. 1999, doi: 10.1016/S0924-977X(98)00044-3.

[8] P. Anand and B. Singh, "Synthesis and evaluation of novel 4-[(3H,3aH,6aH)-3-phenyl)-4,6-dioxo-2-phenyldihydro-2H-pyrrolo[3,4-d]isoxazol-5(3H,6H,6aH)-yl]benzoic acid derivatives as potent acetylcholinesterase inhibitors and anti-amnestic agents," *Bioorg. Med. Chem.*, vol. 20, no. 1, pp. 521–530, Jan. 2012, doi: 10.1016/J.BMC.2011.05.027.

[9] M. L. Crismon, "Tacrine: first drug approved for Alzheimer's disease," *Ann. Pharmacother.*, vol. 28, no. 6, pp. 744–751, 1994, doi: 10.1177/106002809402800612.

[10] R. C. Mohs, B. M. Davis, C. A. Johns, A. A. Mathé, B. S. Greenwald, and T. B. Horvath, "Oral physostigmine treatment of patients with Alzheimer's disease," *Am. J. Psychiatry*, vol. 142, no. 1, pp. 28–33, 1985, doi: 10.1176/AJP.142.1.28.

[11] M. G. Cardozo, A. J. Hopfinger, Y. Iimura, H. Sugimoto, and Y. Yamanishi, "QSAR analyses of the substituted indanone and benzylpiperidine rings of a series of indanone-benzylpiperidine inhibitors of acetylcholinesterase," *J. Med. Chem.*, vol. 35, no. 3, pp. 584–589, Feb. 1992, doi: 10.1021/JM00081A022.

[12] A. Nordberg and A. L. Svensson, "Cholinesterase inhibitors in the treatment of Alzheimer's disease: a comparison of tolerability and pharmacology," *Drug Saf.*, vol. 19, no. 6, pp. 465–480, 1998, doi: 10.2165/00002018-199819060-00004.

Journal of Low Temperature Physics
https://doi.org/10.1007/s10909-023-02948-4

Impact of Impurity on the Mean Energy, Heat Capacity, Free Energy, Entropy and Magnetocaloric Effect of $Ga_{1-\chi}Al_{\chi}As$ Quantum Wire

Sakshi Arora[1] · Yash Gupta[1] · Pranay Khosla[1] · Priyanka[1] · Rinku Sharma[1]

Received: 17 January 2023 / Accepted: 7 February 2023

Abstract
In this work, the thermodynamic properties of a $Ga_{1-\chi}Al_{\chi}As$ quantum wire under a parabolic confinement potential, influence of Rashba SOI and presence of Al impurity are studied. External electric and magnetic fields have also been considered. We first formulate the Hamiltonian of the system and then find the eigenenergies, which are used to calculate the partition function. The partition function is the basis of formulating the thermodynamic properties under consideration, viz. mean energy, heat capacity, free energy, entropy and magnetocaloric effect, which are plotted against temperature and impurity. The results show that the mean energy rises with temperature, a peak structure is observed in the heat capacity and the magnetocaloric effect, the free energy steadily decreases with temperature, and the entropy first increases, and then converges to a constant value. The mean energy, heat capacity and free energy increase with impurity, whereas the magnetocaloric effect decreases. The behaviour of all the properties with respect to impurity reverses when the value of impurity becomes greater than ~ 0.6.

Keywords Quantum Wire · Thermodynamic properties · Rashba SOI · Impurity · Magnetocaloric effect

1 Introduction

The study of nanostructures has steadily been gaining popularity and demand due to its uses in various fields such as medical diagnosis and therapy, plasmonics, photonics and photovoltaics, and food sciences [1–4]. For example, nanostructure-based medical procedures promise an increased sensitivity and speed with reduced cost and labour as compared to the current diagnostic techniques. Moreover, photonic

✉ Rinku Sharma
rinkusharma@dtu.ac.in

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

Published online: 17 March 2023

and nanophotonic structures have the potential to be developed as efficient photothermal conversion methods by manipulating electromagnetic waves on nanoscales. Additionally, the thermal effects of plasmonic nanostructures can be used for energy conversion, optical trapping and thermal management [5–7]. Quantum wires have seen applications in fields such as construction of topological insulators, optical gain in laser fields, and gas sensors [8–10]. The methods for the fabrication of nanostructures have seen major recent developments, such as chemical precipitation, thermal conversion of precursors, and laser ablation in liquid (LAL) [11, 12].

With their increasing demand, extensive research on various properties of Quantum wires is being conducted. Bouazra et al. [13] have utilised the finite difference method to study the optical properties and their dependence on geometrical parameters in InAs/InAlAs quantum wires. Al et al. [14] have studied the effects of electric and magnetic fields on the optical absorption coefficients and refractive index changes in quantum wells. Liu et al. [15] have studied the photoelectric properties of Ag_2S quantum dots. The thermodynamic properties of quantum wires are of such high importance as the physical and chemical characteristics of nanomaterials are completely different from those of bulk materials because of the grand difference in their surface energies [16]. The thermodynamic properties of a GaAs quantum dot with an effective parabolic potential have been investigated by Khordad et al. [17]. Liu et al. [18] have studied the influence of electric field on the thermodynamic properties of the particle confined in a quantum well. For more information, the reader can refer to [19–22].

An important factor which influences the thermodynamic, optical and other properties of semiconductor nanostructures is the electron spin, which can be controlled by lifting the spin degeneracy by using Spin Orbit Coupling [23]. There are two types of SOI in nanostructures—Rashba (arising due to structural inversion symmetry) and Dresselhaus (arising due to bulk inversion asymmetry) [24]. Khoshbakht et al. [25] have studied the magnetic and thermodynamic properties of nanowires in the presence of Rashba SOI. Najafi et al. [26] have studied the thermodynamics of mono-layer quantum wires with Rashba SOI. Donfack and Fotue [27] have studied the thermodynamic properties of a quantum pseudodot under the influence of SOI. Other important factors affecting the properties of nanostructures are confinement potential [17], electric field [18], magnetic field [19], temperature and pressure [28], and laser field [29].

The presence of impurity in the sample greatly alters its thermodynamic and optical properties, as it affects their energy spectrum and physical properties [30]. Khordad et al. [30] have studied the effect of impurity on the entropy of a double cone-like quantum dot. Sedehi et al. [31] have studied the impact of a hydrogenic donor impurity on a hexagonal quantum dot. Heyn and Duque [32] have made a theoretical study of the optical and electronic properties of cylindrical quantum dots in the presence of an arbitrarily located donor impurity. Hosseinpur [33] has studied the impact of a Gaussian impurity on the energy dispersion spectrum of electrons and holes in a doped quantum wire.

In this work, we study the thermodynamic properties viz. mean energy, heat capacity, free energy, entropy and magnetocaloric effect of a quantum wire under the influence of external electric field, magnetic field, and Rashba SOI. We also study the effect of the presence of varying impurity levels in the system. In the "Theoretical Framework" section, we define the theoretical framework of our work, where we solve the Schrödinger equation to calculate the eigenenergies and eigenvalues for the system. We then use these to calculate the values of the thermodynamic properties as functions of temperature for varying levels of impurity. In the "Results" section the effect of impurity on thermodynamic properties is observed using 2D and 3D plots and the observed trends are discussed.

2 Theoretical Framework

This study is formulated for a $\mathrm{Ga}_{1\text{-}\chi}\mathrm{Al}_{\chi}\mathrm{As}$ quantum wire with impurity, by considering a two-dimensional electron gas in the x–y plane. This is done by confining the electron motion in x-direction using a parabolic lateral confinement potential, resulting in a quantum wire in the y-direction. When an external magnetic field $\vec{B} = B\hat{k}$ is applied to the quantum wire in the z-direction, the Hamiltonian of the system is given by [23, 34]:

$$H = \frac{1}{2m_{\mathrm{eff}}(\chi)}(\vec{p} + e\vec{A})^2 + \frac{1}{2}m_{\mathrm{eff}}(\chi)\omega_0^2x^2 + \frac{1}{2}g\mu_B\vec{\sigma}\cdot\vec{B} + H_{\mathrm{R}} \tag{1}$$

where $\vec{p}$ is the momentum, e is the electronic charge, $\vec{A}$ is the vector potential due to magnetic field given by $\vec{A} = Bx\hat{j}$, ω_0 is the oscillator strength, g is the Lande's g factor, $\mu_B = \frac{e\hbar}{2m_{\mathrm{e}}}$ is the Bohr magneton, m_{e} is the rest mass of the electron, and $\vec{\sigma}$ is the Pauli spin matrix vector. The second term of Eq. (1) represents the parabolic potential term.

$m_{eff}(\chi)$ is the effective mass of the charge carrier in the presence of Al impurity (χ), which is given by [35]:

$$m_{\mathrm{eff}}(\chi, P, T) = m_e[1 + \frac{\Pi^2(\chi)}{3}\left(\frac{2}{E_{\mathrm{g}}(\chi, P, T)} + \frac{1}{E_{\mathrm{g}}(\chi, P, T) + \Delta_0(\chi)}\right) + \delta(\chi)]^{-1} \tag{2}$$

where $\Pi(\chi)$ is the inter-band matrix element $\left[\Pi^2(\chi) = (28900 - 6290\chi)\mathrm{meV}\right]$, $\Delta_0(\chi) = (341 - 66\chi)\mathrm{meV}$ is the valence band spin–orbit splitting, and $\delta(\chi) = -3.935 + 0.488\chi + 4.938\chi^2$ is used to consider the remote-band effects.

$E_{\mathrm{g}}(\chi, P, T)$ is the energy gap function in the conduction band given by [36]:

$$E_{\mathrm{g}}(\chi, P, T) = p + q\chi + r\chi^2 + sP - \frac{\beta T^2}{\gamma + T} \tag{3}$$

where $p = 1519.4\mathrm{meV}$, $q = 1360\mathrm{meV}$, $r = 220\mathrm{meV}$, $s = 10.7\,\mathrm{meV/kbar}$, $\beta = 0.5405\ \mathrm{meV/K}$ and $\gamma = 204K$. The pressure and temperature values are taken as $P = 15\mathrm{kbar}$ and $T = 298K$.

In Eq. (1), H_{R} is the Rashba term given by:

$$H_{\mathrm{R}} = \frac{\alpha}{\hbar}(\vec{\sigma} \times \left(\vec{p} + e\vec{A}\right))_z \tag{4}$$

where α is the Rashba SOI factor, which can be varied with the gate voltage [34].

Upon applying an external electric field $\vec{E} = E\hat{i}$ in the x-direction and expanding the vector terms, the combined Hamiltonian of the system from Eq. (1) becomes:

$$\begin{aligned} H = {} & \frac{1}{2m_{\mathrm{eff}}(\chi)}\left(p_x^2 + \left(p_y + eBx\right)^2\right) + \frac{1}{2}m_{\mathrm{eff}}(\chi)\omega_0^2x^2 + eEx + \frac{1}{2}g\mu_B\sigma_zB \\ & + \frac{\alpha}{\hbar}\left(\sigma_x\left(p_y + eBx\right) - \sigma_yp_x\right) \end{aligned} \tag{5}$$

The Hamiltonian does not change with translation along the length of the wire, i.e. it is translationally invariant. Hence the energy eigenstates of H can be written in terms of plane wave solution, where the wave function $\Psi(x, y)$ is given by [34]:

$$\Psi(x, y) = \phi(x)\exp\left(ik_yy\right) \tag{6}$$

where k_{y} is the wavenumber in y-direction. According to the de Broglie hypothesis, p_y can be represented by $\hbar k_y$ in the Hamiltonian, which transforms the Hamiltonian into $H = H_0 + H_{\mathrm{R}}$ such that

$$\begin{aligned} H_0 = {} & \frac{p_x^2}{2m_{\mathrm{eff}}(\chi)} + \frac{1}{2}m_{\mathrm{eff}}(\chi)\omega^2(x - x_0)^2 - \frac{e^2E^2}{2m_{\mathrm{eff}}(\chi)\omega^2} \\ & + \frac{\omega_0^2\hbar^2k_{\mathrm{y}}^2}{2m_{\mathrm{eff}}(\chi)\omega^2} - \frac{e^2EB\hbar k_{\mathrm{y}}}{m_{\mathrm{eff}}^2(\chi)\omega^2} + \frac{1}{2}g\mu_B\sigma_zB \end{aligned} \tag{7}$$

and

$$H_{\mathrm{R}} = \alpha\left(\sigma_x\left(k_{\mathrm{y}} + \frac{eBx}{\hbar}\right) - i\sigma_y\frac{\mathrm{d}}{\mathrm{d}x}\right) \tag{8}$$

where $x_0 = -\left(\frac{eE}{m_{\mathrm{eff}}(\chi)\omega^2} + \frac{eB\hbar k_{\mathrm{y}}}{m_{\mathrm{eff}}^2(\chi)\omega^2}\right)$ is the guiding centre coordinate for the harmonic oscillator, $\omega = \sqrt{\omega_0^2 + \omega_{\mathrm{c}}^2}$ is the effective cyclotron frequency and $\omega_{\mathrm{c}} = \frac{eB}{m_{\mathrm{eff}}(\chi)}$ is the cyclotron frequency.

The energy eigenvalues and eigenvectors for H_0 can be found out by:

$$H_0\Psi_{n\sigma}(x) = E_{n\sigma}\Psi_{n\sigma}(x) \tag{9}$$

such that

$$\Psi_{n\sigma}(x) = \frac{1}{\left(2^n\sqrt{\pi}c_ln!\right)^{1/2}}H_n\left(\frac{x - x_0}{c_l}\right)\exp\left(-\frac{1}{2}\left(\frac{x - x_0}{c_l}\right)^2\right)X_\sigma \tag{10}$$

where $c_{\mathrm{l}} = \sqrt{\frac{\hbar}{m_{\mathrm{eff}}(\chi)\omega}}$ is the characteristic length of the harmonic oscillator, n is the range of all integers and represents the energy levels, and $\sigma = \pm$ represents up and down spin. $H_n(x)$ and X_σ are the hermite polynomial and the spinor functions for spin up ($X_+ = (10)^T$) and for spin down, ($X_- = (01)^T$) respectively.

The eigenenergies found by solving Eq. (9) are:

$$E_{n\sigma} = \hbar\omega\left(n + \frac{1}{2}\right) - \frac{e^2E^2}{2m_{\text{eff}}(\chi)\omega^2} + \frac{\omega_0^2\hbar^2k_y^2}{2m_{\text{eff}}(\chi)\omega^2} - \frac{e^2EB\hbar k_y}{m_{\text{eff}}^2(\chi)\omega^2} + \frac{1}{2}g\mu_B\sigma_zB \quad (11)$$

$\phi(x)$ can be rewritten in terms of $\Psi_{n\sigma}(x)$ as $\phi(x) = \sum_{n\sigma} a_{n\sigma}\Psi_{n\sigma}(x)$, Eq. (11) can thus be transformed into [34]:

$$\sum_{n\sigma} a_{n\sigma}\left(E_{n\sigma} - E\right)\Psi_{n\sigma}(x) + \sum_{n\sigma} a_{n\sigma}H_R\Psi_{n\sigma}(x) = 0 \quad (12)$$

The orthogonality condition for the wave function gives,

$$\left(E_{n\sigma} - E\right)a_{n\sigma} + \sum_{n'\sigma'} a_{n'\sigma'}\left\langle\Psi_{n\sigma}\middle|H_R\middle|\Psi_{n'\sigma'}\right\rangle = 0 \quad (13)$$

where the eigenenergies corresponding to H_{R} are given by:

$$\begin{aligned}\left\langle n\sigma\middle|H_R\middle|n'\sigma'\right\rangle = \alpha&\left[\left(1 - \frac{\omega_c^2}{\omega^2}\right)k_y - \frac{\omega_c eE}{\hbar\omega^2}\right]\delta_{n,n'}\delta_{\sigma,-\sigma'} \\ &+ \frac{\alpha}{c_l}\left[\left(\frac{\omega_c}{\omega} + \sigma\right)\sqrt{\frac{n+1}{2}}\delta_{n,n'-1} + \left(\frac{\omega_c}{\omega} - \sigma\right)\sqrt{\frac{n}{2}}\delta_{n,n'+1}\right]\delta_{\sigma,-\sigma'}\end{aligned} \quad (14)$$

where δ is the Dirac Delta function.

Equations (11, 13, 14) are used to obtain the eigenvalues of the system. Figure 1 shows the eigenenergies and normalised wave functions obtained for the 12 lowest energy levels, along with the potential energy of the whole system.

The partition function Z is a mathematical formulation which is the basis for calculating the thermodynamics of a system. It is given by [23]:

$$Z = \sum_n e^{-\beta E_n} \quad (15)$$

where $\beta = \frac{1}{k_BT}$, k_B is the Boltzmann constant, and E_n are the eigenenergies found at different energy levels.

The thermodynamic properties of the quantum wire can be calculated using the partition function Z through the following formulae [21, 37]:

1) Mean energy: $U = -\frac{\partial \log Z}{\partial\beta}$
2) Heat capacity: $C = \frac{\partial U}{\partial T}$
3) Free energy: $F = -k_BTlogZ$
4) Entropy: $S = k_B\text{log}Z - k_B\beta\frac{\partial \text{log}Z}{\partial\beta}$
5) Magnetocaloric effect: $\Delta S = S(B \neq 0, T) - S(B = 0, T)$

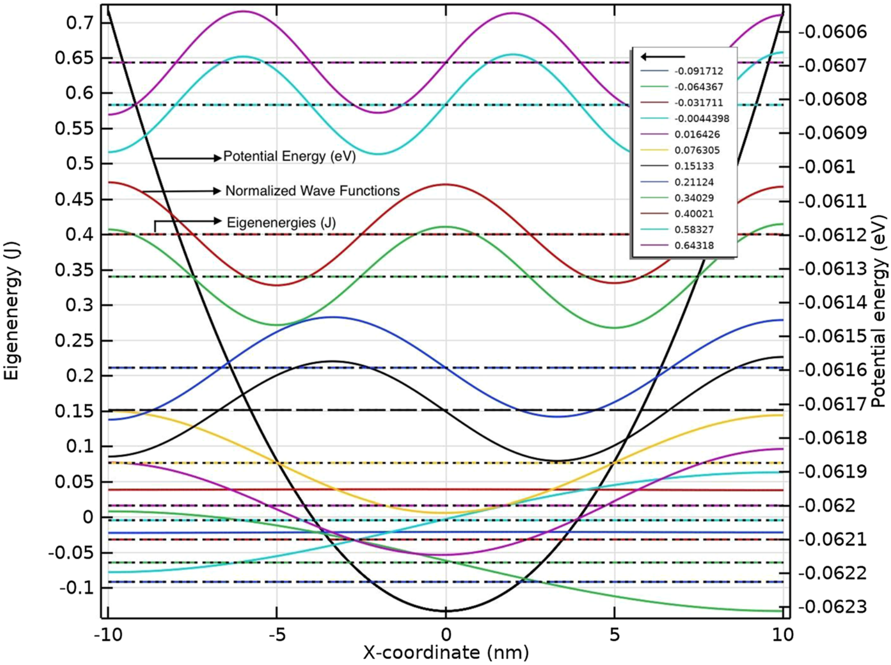


Fig. 1 Potential energy, eigenenergies, and normalised wave functions of the 12 lowest energy levels of the Quantum Wire under electric and magnetic fields, Rashba SOI, and impurity

3 Results

In this study, thermodynamic properties, viz. Mean Energy, Heat Capacity, Free Energy, Entropy and Magnetocaloric Effect of a quantum wire have been analysed as a function of temperature for the 6 lowest energy levels of the system ($n = 0\,\text{to}\,5$). An electric field with field strength $F = 0.6 \times 10^6\,\text{V/m}$ and a magnetic field with field strength $B = 1T$ have been applied. The value of Lande's g factor is considered to be $g = -0.44$. The influence of Rashba Spin Orbital Interaction (SOI) has been taken into consideration with the Rashba coefficient $\alpha = 2.5 \times 10^{-11}\,\text{eVm}$. These results have been analysed for varying levels of impurity present in the system with its range as $\chi = 0\,\text{to}\,2$.

3.1 Mean Energy

At low temperature, the quantum particle is in a low-energy state, hence the mean energy is also low. As the temperature rises, the particle gains kinetic energy, thus increasing its mean energy [18]. This trend was observed in Fig. 2a, b.

The behaviour of the mean energy with a change in the impurity was explained by Figs. 3 and 4. Figure 3 shows the relation between effective mass and impurity. The

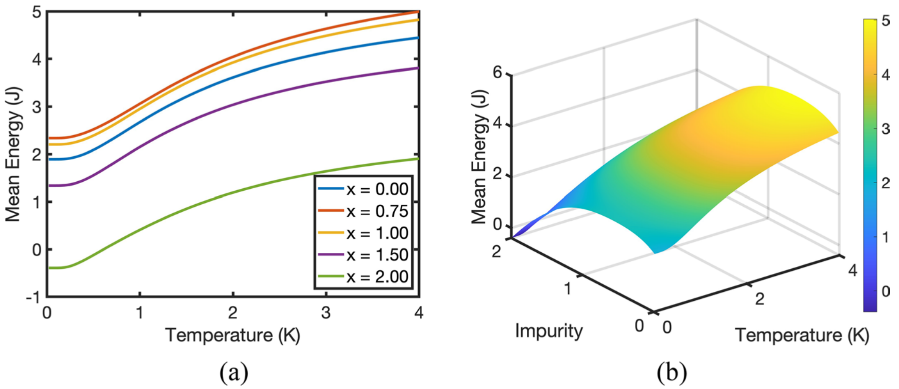


Fig. 2 Mean energy as a function of temperature for varying levels of impurity

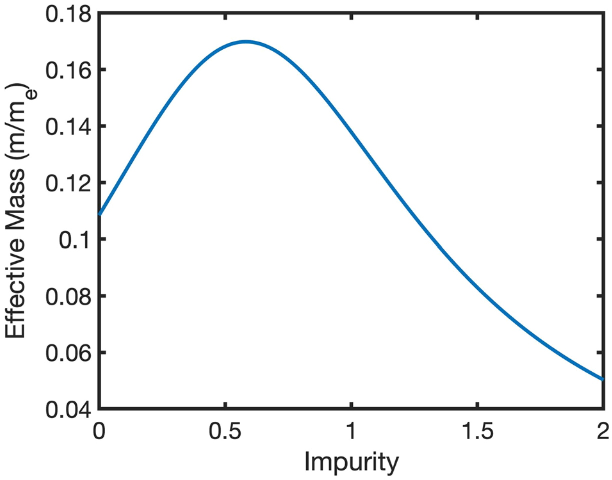


Fig. 3 Effective mass as a function of impurity

effective mass increases with increasing impurity level, reaching its peak value at an impurity level of ~ 0.6, after which it starts decreasing. Figure 4 shows the relation between mean energy and effective mass. It is evident that the mean energy strictly monotonically increases with increasing effective mass. Both these inferences justify the behaviour of mean energy, i.e. it first increases with an increase in the impurity level, reaching its peak value at ~ 0.6 impurity level, and then decreases with increasing impurity. Additionally, the mean energy at a specific impurity level increases or decreases by a constant value for all temperatures, i.e. the plot only shifts upwards or downwards, without any apparent change in the shape of the plot.

This trend was observed with more clarity in Fig. 2b, which is a 3D plot between temperature, impurity and the mean energy. The mean energy increases with increasing temperature, and first increases then decreases with increasing impurity.

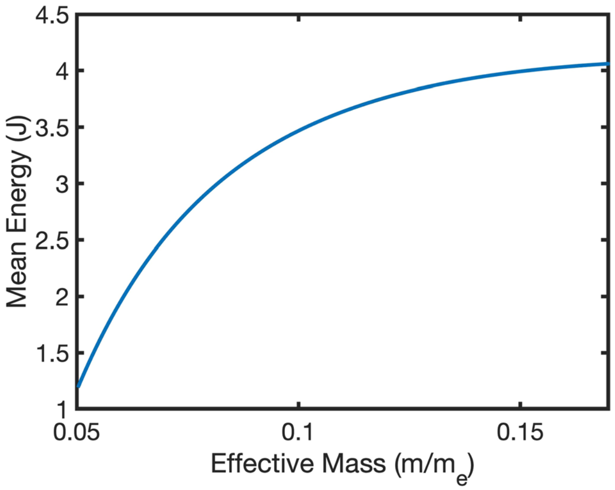


Fig. 4 Mean energy as a function of effective mass

3.2 Heat Capacity

The behaviour of the heat capacity can be explained with the help of the internal energy of the system. The heat capacity is directly proportional to the gradient of the internal energy with respect to the temperature. At very low temperatures, only the lowest energy level of a quantum system is populated, implying a low internal energy; since there is negligible change in the internal energy, its gradient (and effectively the heat capacity of the system) is a very small value. As the temperature rises, higher energy levels also start getting populated, and the internal energy of the system rises rapidly, indicating a high value of the gradient and effectively the heat capacity. As the temperature rises even more, all the possible energy levels get populated evenly, which implies that no more energy can be absorbed by the system. At this point the internal energy starts getting saturated, and the gradient (and heat capacity) decreases steadily, eventually acquiring a very small constant value. This behaviour was observed in Fig. 5a–c; in Fig. 5a, the heat capacity first increases rapidly, reaches a maximum, and then steadily decreases, demonstrating a "peak" structure. It was observed more clearly in Fig. 5b, which is a 3D plot of the heat capacity with temperature and impurity.

The variation of heat capacity with impurity can be explained by Figs. 3 and 6. As in the case of the mean energy, the heat capacity also increases with increasing effective mass. The effective mass increases with increasing impurity until impurity reaches a value of ~ 0.6. As a result, the plot of heat capacity shifts to the right until impurity reaches ~ 0.6, after which the plot starts moving back towards the left, and eventually the peak of the plot also starts decreasing, since the effective mass is steadily decreasing. Figure 5c shows the top view of the 3D plot between heat capacity, temperature and impurity; it verified the shifting of the peak (represented by the yellow region) first towards right and then left.

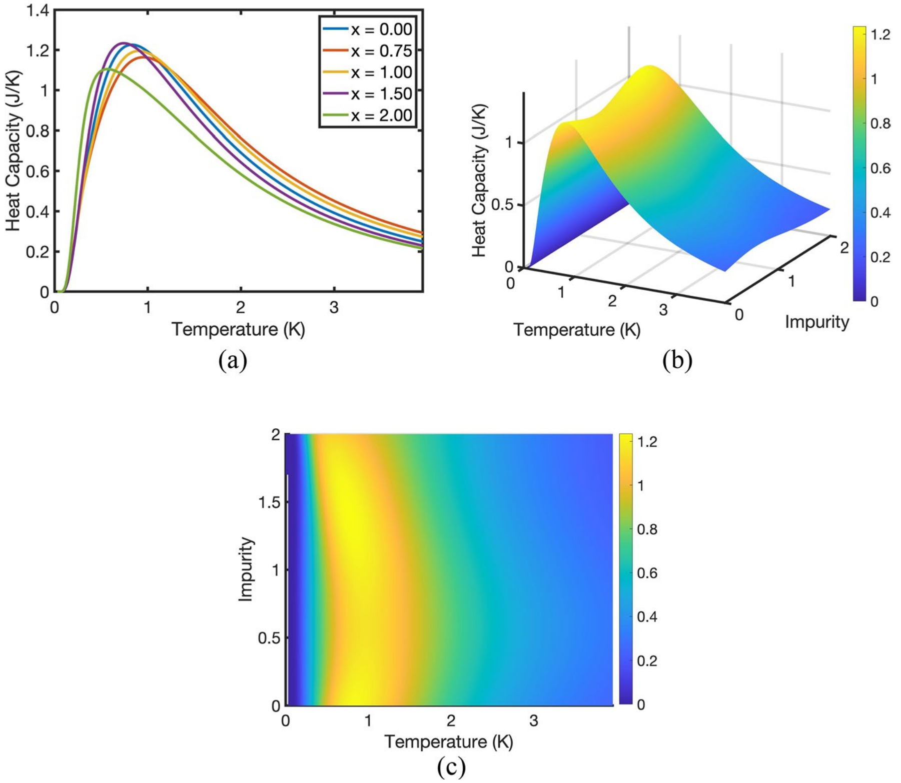


Fig. 5 Heat capacity as a function of temperature for varying levels of impurity

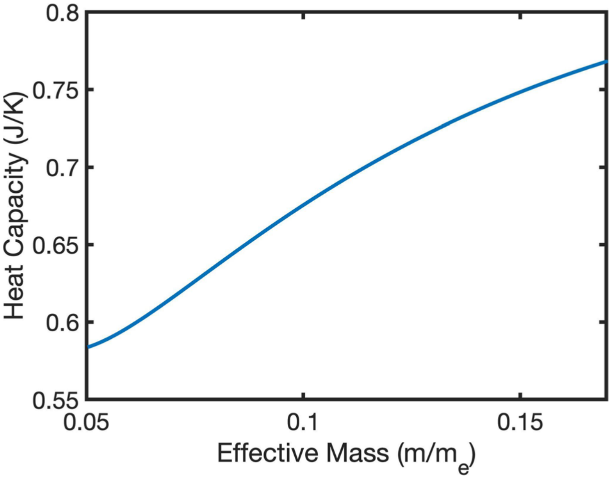


Fig. 6 Heat capacity as a function of effective mass

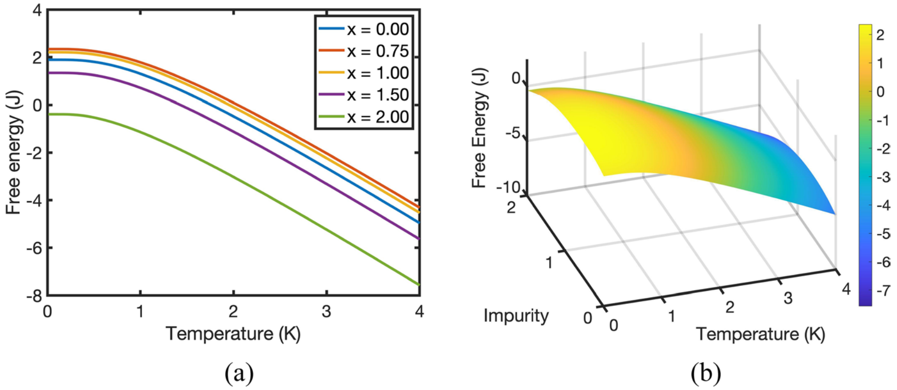


Fig. 7 Free Energy as a function of temperature for varying levels of impurity

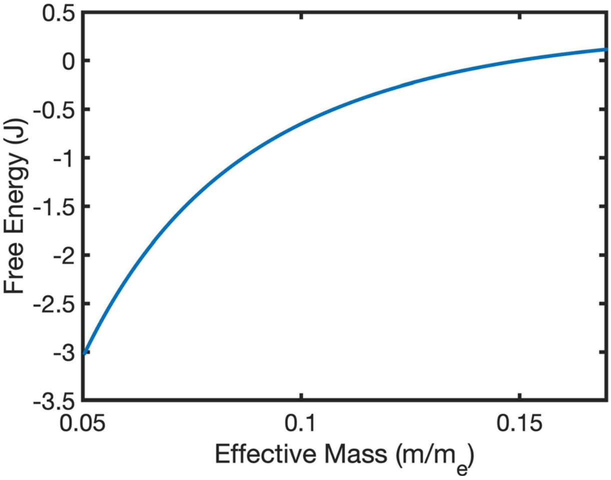


Fig. 8 Free Energy as a function of effective mass

3.3 Free Energy

It was observed in Fig. 7a, b that at very low temperatures, the free energy has a small positive, constant value, which becomes negative and then decreases monotonically as the temperature increases [18, 25].

It can be seen in Figs. 3 and 8 that the free energy monotonically increases with increasing effective mass, and the effective mass increases with increasing impurity, reaches a peak at ~ 0.6, and then decreases. This verifies the observation that the free energy first increases with increasing impurity (until impurity reaches a value of ~ 0.6), and then decreases. Figure 7b verifies this variation of free energy with impurity at the 3D level.

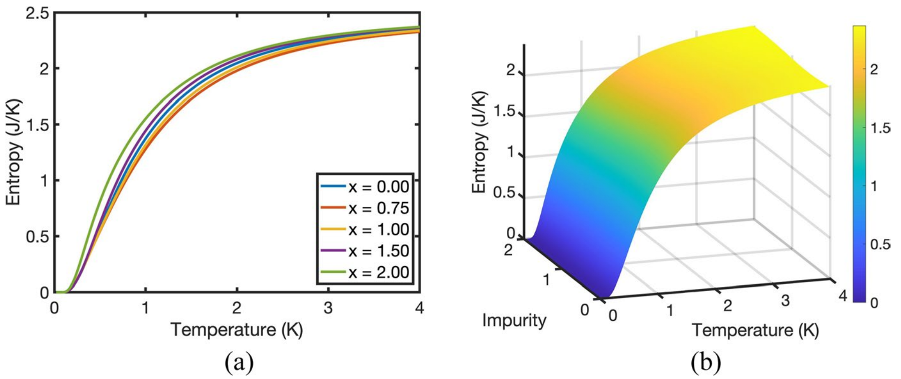
3.4 Entropy

At very low temperatures, the quantum particle populates only the lowest energy level, which is the most stable, thus explaining the small value of entropy. As the temperature rises, the kinetic energy of the particle increases, and higher energy levels also start getting populated. The probability of the particle being at different energy levels increases, which means the particle does not localise. This increases the randomness, and effectively the entropy of the system [18]. The reason behind the convergence of the entropy could be that, as the temperature rises further, all the possible energy levels get populated evenly, thus stabilising the entropy of the system. This trend was observed in Fig. 9a, b, which show that the entropy of the system increases with increasing temperature, and eventually stabilises, reaching an almost constant value. As impurity is added to the system, the quantum particle gets bound to the impurity particles due to Coulombic force, and the randomness decreases, decreasing the entropy [31]. Hence the entropy decreases very slightly till the impurity value is ~ 0.6, after which its behaviour reverses, similar to the other thermodynamic properties.

3.5 Magnetocaloric Effect

Figure 10 is a plot between *S*, the entropy at a specified magnetic field ($B = 1$ T) and S_0, the entropy at zero magnetic field, at a fixed impurity of 2. The magnetocaloric effect quantifies the difference between these two values of entropy (S and S_0) at any particular temperature and impurity level. In Fig. 11a, we see that the magnetocaloric effect first increases with temperature, reaches a peak, and then decreases, eventually converging to a constant value which tends to zero.

As demonstrated in Fig. 11b, c, which are 3D plots of the magnetocaloric effect v/s temperature and impurity, it was also observed that the magnetocaloric effect first decreases with impurity till the impurity reaches a value of ~ 0.6 and then increases. Figure 11c is a side view of the same, which clearly shows the minimum lying at a value of ~ 0.6 of impurity.



(a) (b)

Fig. 9 Entropy as a function of temperature for varying levels of Impurity

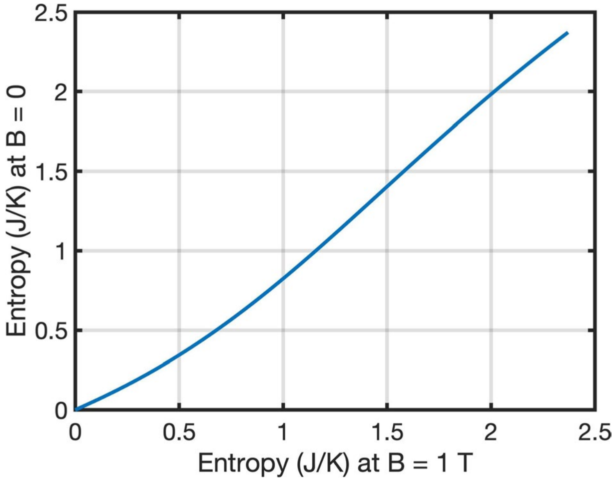


Fig. 10 Entropy at $B = 0$ as a function of entropy at $B = 1$ T

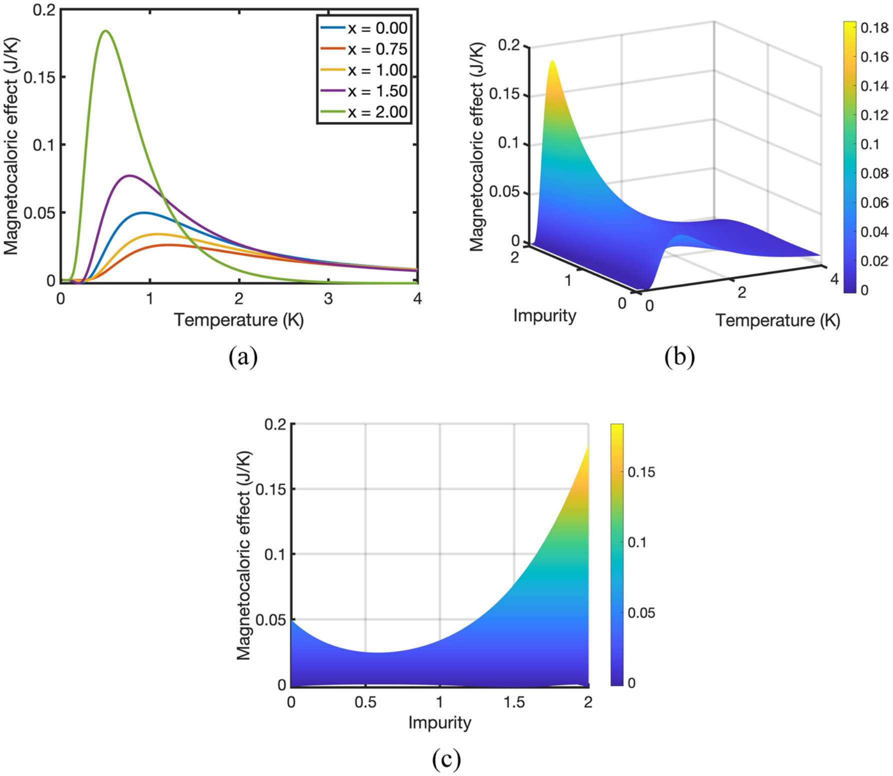


Fig. 11 Magnetocaloric Effect as a function of temperature for varying levels of Impurity

4 Conclusion

We have studied the thermodynamic properties, viz. mean energy, heat capacity, free energy, entropy, and magnetocaloric effect of a quantum wire under the influence of external magnetic and electric fields, Rashba SOI and presence of impurity. The Hamiltonian, eigenenergies, partition function and finally the thermodynamic properties were calculated for varying levels of temperature and impurity. The results show that the mean energy and entropy increase, free energy decreases, and heat capacity and magnetocaloric effect show a peak structure with temperature. The mean energy, heat capacity and free energy first increase and the magnetocaloric effect first decreases with impurity, which is reversed as the impurity reaches a value ~ 0.6. The entropy does not vary much with impurity. Since the thermodynamic properties show considerable amounts of variations when impurity is added, this work gives important insights which can potentially play a very important role in fabricating real quantum wires.

Acknowledgements Sakshi Arora, Yash Gupta, Pranay Khosla, and Rinku Sharma are grateful for the research facilities provided by Delhi Technological University. Priyanka also thanks the University Grants Commission for their financial assistance.

Author contributions SA, YG, and PK contributed to Methodology, Software, Validation, Formal analysis, Data curation, Writing—original draft. P and RS contributed to Resources, Investigation, Supervision, Writing—review & editing, Project administration, Conceptualization, Formal analysis, and Visualization.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. J. Martin, J. Plain, Fabrication of aluminium nanostructures for plasmonics. J. Phys. D: Appl. Phys. **48**(18), 184002 (2014). https://doi.org/10.1088/0022-3727/48/18/184002
2. F. Priolo, T. Gregorkiewicz, M. Galli, T.F. Krauss, Silicon nanostructures for photonics and photovoltaics. Nat. Nanotechnol. **9**(1), 19–32 (2014). https://doi.org/10.1038/nnano.2013.271
3. K. Pathakoti, M. Manubolu, H.-M. Hwang, Nanostructures: current uses and future applications in food science. J. Food Drug Anal. **25**(2), 245–253 (2017). https://doi.org/10.1016/j.jfda.2017.02.004
4. E. Sagi, Y. Oreg, Non-Abelian topological insulators from an array of quantum wires. Phys. Rev. B. (2014). https://doi.org/10.1103/physrevb.90.201102
5. V.T. Tran, H.-Q. Nguyen, Y.-M. Kim, G. Ok, J. Lee, Photonic–plasmonic nanostructures for solar energy utilization and emerging biosensors. Nanomaterials **10**(11), 2248 (2020). https://doi.org/10.3390/nano10112248
6. Md.M. Bellah, S.M. Christensen, S.M. Iqbal, Nanostructures for medical diagnostics. J. Nanomater. **2012**, 1–21 (2012). https://doi.org/10.1155/2012/486301
7. J. Wu and Y. Wang (2017) Optical Absorption and thermal effects of plasmonic nanostructures In: *Nanoplasmonics-Fundamentals and Applications,* http://dx.doi.org/https://doi.org/10.5772/67505. Accessed 03 Feb 2023
8. S. Saravanan, A.J. Peter, C.W. Lee, Laser field induced optical gain in a group III-V quantum wire. Eur Phys J D (2016). https://doi.org/10.1140/epjd/e2016-70191-8

9. Z. Song et al., Sensitive Room-temperature H_2S gas sensors employing sno_2 quantum wire/reduced graphene oxide nanocomposites. Chem. Mater. **28**(4), 1205–1212 (2016). https://doi.org/10.1021/acs.chemmater.5b04850
10. Q. Zhang et al., CuO nanostructures: synthesis, characterization, growth mechanisms, fundamental properties, and applications. Prog. Mater Sci. **60**, 208–337 (2014). https://doi.org/10.1016/j.pmatsci.2013.09.003
11. J. Xiao, P. Liu, C.X. Wang, G.W. Yang, External field-assisted laser ablation in liquid: an efficient strategy for nanocrystal synthesis and nanostructure assembly. Prog. Mater Sci. **87**, 140–220 (2017). https://doi.org/10.1016/j.pmatsci.2017.02.004
12. M. Elsabahy, G.S. Heo, S.-M. Lim, G. Sun, K.L. Wooley, Polymeric Nanostructures for Imaging and Therapy. Chem. Rev. **115**(19), 10967–11011 (2015). https://doi.org/10.1021/acs.chemrev.5b00135
13. A. Bouazra, S.A.-B. Nasrallah, M. Said, Theory of electronic and optical properties for different shapes of InAs/In0.52Al0.48As quantum wires. Phys. E **75**, 272–279 (2016). https://doi.org/10.1016/j.physe.2015.09.039
14. E.B. Al, F. Ungan, U. Yesilgul, E. Kasapoglu, H. Sari, I. Sökmen, Effects of applied electric and magnetic fields on the nonlinear optical properties of asymmetric GaAs/Ga1-χAlχAs double inverse parabolic quantum well. Opt. Mater. **47**, 1–6 (2015). https://doi.org/10.1016/j.optmat.2015.06.048
15. B. Liu et al., Photoelectrical properties of Ag2S quantum dot-modified TiO2nanorod arrays and their application for photovoltaic devices. Dalton Trans. **42**(6), 2232–2237 (2013). https://doi.org/10.1039/c2dt32031b
16. K. V. Chandekar, Mohd. Shkir, and S. AlFaify, “A structural, elastic, mechanical, spectroscopic, thermodynamic, and magnetic properties of polymer coated CoFe2O4 nanostructures for various applications,” *Journal of Molecular Structure*, vol. 1205, p. 127681, Apr. 2020, doi: https://doi.org/10.1016/j.molstruc.2020.127681.
17. R. Khordad, B. Mirhosseini, M.M. Mirhosseini, thermodynamic properties of a GaAs quantum dot with an effective-parabolic potential: theory and simulation. J. Low Temp. Phys. **197**(1–2), 95–110 (2019). https://doi.org/10.1007/s10909-019-02218-2
18. X. Liu et al., Influence of the spatially inhomogeneous electric field on the thermodynamic property of the particle confined in a quantum well. Phys. Scr. **97**(10), 105308 (2022). https://doi.org/10.1088/1402-4896/ac90f9
19. B. Donfack and A. J. Fotuea, Thermodynamic properties of asymmetric semiconductor quantum wire under the magnetic field, Research Square Platform LLC, Sep. 2022. Accessed: Jan. 10, 2023. [Online]. Available: http://dx.doi.org/https://doi.org/10.21203/rs.3.rs-2011010/v1
20. R. Khordad, Thermodynamical properties of triangular quantum wires: entropy, specific heat, and internal energy. Continuum Mech. Thermodyn. **28**(4), 947–956 (2015). https://doi.org/10.1007/s00161-015-0429-2
21. H.R. Rastegar Sedehi, R. Khordad, Magnetocaloric effect, magnetic susceptibility and specific heat of tuned quantum dot/ring systems. Phys. E: Low-dimens. Syst. Nanostructures **134**, 114886 (2021). https://doi.org/10.1016/j.physe.2021.114886
22. M. Servatkhah, R. Khordad, A. Firoozi, H.R. Rastegar Sedehi, A. Mohammadi, Low temperature behavior of entropy and specific heat of a three dimensional quantum wire: Shannon and Tsallis entropies. Eur. Phys. J. B (2020). https://doi.org/10.1140/epjb/e2020-10034-5
23. R. Khordad, H.R. Rastegar Sedehi, Low temperature behavior of thermodynamic properties of 1D quantum wire under the Rashba spin–orbit interaction and magnetic field. Solid State Commun. **269**, 118–124 (2018). https://doi.org/10.1016/j.ssc.2017.10.018
24. A.J. Fotue, T.V. Diffo, E. Baloitcha, F.C. Fobasso Mbognou, G.T. Tedondje, M.N. Hounkonnou, Spin–orbit interaction on the thermodynamics of three-dimensional impurity magnetopolaron under strong parabolic potential. Eur. Phys. J. Plus (2020). https://doi.org/10.1140/epjp/s13360-020-00441-5
25. Y. Khoshbakht, R. Khordad, H.R. Rastegar Sedehi, Magnetic and thermodynamic properties of a nanowire with Rashba spin–orbit interaction. J. Low Temp. Phys. **202**(1–2), 59–70 (2020). https://doi.org/10.1007/s10909-020-02522-2
26. D. Najafi, B. Vaseghi, G. Rezaei, R. Khordad, Thermodynamics of mono-layer quantum wires with spin-orbit interaction”. Eur. Phys. J. Plus (2018). https://doi.org/10.1140/epjp/i2018-12102-3
27. B. Donfack, A.J. Fotue, Effects of spin orbit interaction (SOI) on the thermodynamic properties of a quantum pseudodot. J. Low Temp. Phys. **204**(5–6), 206–222 (2021). https://doi.org/10.1007/s10909-021-02604-9

28. R. Khordad, A.R. Firoozi, H.R.R. Sedehi, Simultaneous effects of temperature and pressure on the entropy and the specific heat of a three-dimensional quantum wire: tsallis formalism. J Low Temp. Phys. **202**(1–2), 185–195 (2020). https://doi.org/10.1007/s10909-020-02536-w
29. B. Donfack, F.C.F. Mbognou, G.T. Tedondje, T.M. Cedric, A.J. Fotue, Cumulative Effects of Laser and Spin-Orbit Interaction (SOI) on the Thermal Properties of Quantum Pseudo-dot. J. Low Temp. Phys. **206**(1–2), 63–79 (2021). https://doi.org/10.1007/s10909-021-02623-6
30. R. Khordad, H.R. Rastegar Sedehi, H. Bahramiyan, Simultaneous effects of impurity and electric field on entropy behavior in double cone-like quantum dot. Commun. Theor. Phys. **69**(1), 95 (2018). https://doi.org/10.1088/0253-6102/69/1/95
31. H.R. Rastegar Sedehi, R. Khordad, H. Bahramiyan, Optical properties and diamagnetic susceptibility of a hexagonal quantum dot: impurity effect. Opt. Quantum Electron. (2021). https://doi.org/10.1007/s11082-021-02927-7
32. C. Heyn, C.A. Duque, Donor impurity related optical and electronic properties of cylindrical GaAs-AlxGa1−x As quantum dots under tilted electric and magnetic fields. Sci. Rep. (2020). https://doi.org/10.1038/s41598-020-65862-9
33. P. Hosseinpour, Effect of Gaussian impurity parameters on the valence and conduction subbands and thermodynamic quantities in a doped quantum wire. Solid State Commun. **322**, 114061 (2020). https://doi.org/10.1016/j.ssc.2020.114061
34. M. Kumar, S. Lahon, P.K. Jha, M. Mohan, Energy dispersion and electron g-factor of quantum wire in external electric and magnetic fields with Rashba spin orbit interaction. Superlattices Microstruct. **57**, 11–18 (2013). https://doi.org/10.1016/j.spmi.2013.01.007
35. E. ReyesGómez, N. Raigoza, L.E. Oliveira, Effects of hydrostatic pressure and aluminum concentration on the conduction-electrongfactor in GaAs (Ga, Al)as quantum wells under in-plane magnetic fields. Phys. Rev. B (2008). https://doi.org/10.1103/physrevb.77.115308
36. H.M. Baghramyan, M.G. Barseghyan, A.A. Kirakosyan, R.L. Restrepo, C.A. Duque, Linear and nonlinear optical absorption coefficients in GaAs/Ga1−xAlxAs concentric double quantum rings: effects of hydrostatic pressure and aluminum concentration. J. Lumin. **134**, 594–599 (2013). https://doi.org/10.1016/j.jlumin.2012.07.024
37. R. Khordad, H.R.R. Sedehi, Thermodynamic properties of a double ring-shaped quantum dot at low and high temperatures. J. Low Temp. Phys. **190**(3–4), 200–212 (2017). https://doi.org/10.1007/s10909-017-1831-x

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

Introducing Switched Adaptive Control for Self-Reconfigurable Mobile Cleaning Robots

Madan Mohan Rayguru, Spandan Roy, Lim Yi, Mohan Rajesh Elara, and Simone Baldi, *Senior Member, IEEE*

***Abstract*— Reconfigurable robots provide an attractive option for cleaning tasks, thanks to their better area coverage and adaptability to changing environment. However, the ability to change morphology creates drastic changes in the reconfigurable robot dynamics, and existing control design techniques do not take this into account. Neglecting configuration changes can lead to performance degradation and, in the worst scenarios, instability. This paper proposes to embed the changes arising from reconfiguration in the control design, via a switched uncertain Euler-Lagrangian model. Accordingly, a novel switched adaptive design is proposed for trajectory tracking. Closed-loop stability is assured using the multiple Lyapunov function framework, and the design is implemented and validated on a self-reconfigurable pavement cleaning mobile robot (PANTHERA).**

***Note to Practitioners*—Self-reconfigurable mobile cleaning robots, which can change their configurations as per the application requirements, are now predominantly used for cleaning and maintenance operations because of their better area coverage, less manpower requirement and consistent performance. However, the state-of-the-art control strategies for conventional robots cannot always ensure stability and performance under the simultaneous effects of configuration changes and uncertainties. The switched Euler-Lagrange model formulated in this work can capture the configuration changes of the robot and the proposed switched adaptive controller can tackle uncertainties of each configurations of the robot. The simulation and experimental results clearly show the potential issues of the state-of-the-art methods and the remarkable benefits of the proposed approach.**

***Index Terms*— Reconfigurable robots, switched Euler-Lagrange dynamics, robust adaptive control.**

I. Introduction

THE field of robotics with its engineering applications is rapidly expanding. Currently, the robots utilized in most applications can be broadly categorized into rigid, flexible and soft [1]. Flexible and soft robots can be advantageous in handling objects which are delicate and safety critical [2]. On the other hand, rigid robots are often utilized to handle objects of fixed shape and withstand higher contact forces. However, the fixed shape and size of the rigid robots sometimes lead to complications and less efficient actions when the tasks involve objects with different physical structures [3]. For this reason, the last decade has seen a growing interest in realizing rigid robots which can change their shape according to the environmental situations and the nature of the assignments. This class of robots are called as reconfigurable robots, which can change their configurations either manually (e.g. remotely operated reconfigurable robots) or autonomously (e.g. self reconfigurable robots) [3], [4], [5], [6].

Different types of reconfigurable mobile robots are now used increasingly in cleaning and service tasks, as they can deform to a required shape according to the need [7], [8]. The indoor and outdoor cleaning works are particularly challenging, as the cleaning surface/area may comprise of obstacles with different sizes, shapes and curvatures. In these scenarios, suitably designed self-reconfigurable mobile robots can provide efficient, safe and time critical completion of the assigned tasks [7], [9]. However, self reconfigurable mobile robots call for complex path planning and control strategies, so as to deal with the dynamic environment (moving obstacles, varying external forces) and shape changes. Some representative strategies and results will be reviewed hereafter, so as to highlight some challenges in the field.

A. Related Works and Open Problems

Different control approaches for reconfigurable robots have been discussed in the literature [10], [11], [12]. The conventional proportional-integral-derivative (PID) control was utilized for speed control in the pavement sweeping robots and tetris inspired floor cleaning robots [7]. A robust backstepping

Manuscript received 11 November 2022; revised 19 January 2023; accepted 11 March 2023. This article was recommended for publication by Associate Editor V. Stojanovic and Editor P. Rocco upon evaluation of the reviewers' comments. This work was supported in part by the National Robotics Programme under its Robotics Enabling Capabilities and Technologies under Project 192 25 00051; in part by the National Robotics Programme under its Robot Domain Specific under Project 192 22 00058; in part by the Agency for Science, Technology and Research; in part by the National Key Research and Development Program of China under Grant 2022YFE0198700; and in part by the Natural Science Foundation of China under Grant 62150610499 and Grant 62073074. *(Madan Mohan Rayguru and Spandan Roy contributed equally to this work.) (Corresponding authors: Simone Baldi; Madan Mohan Rayguru.)*

Madan Mohan Rayguru is with the Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India (e-mail: mmrayguru@dtu.ac.in).

Spandan Roy is with the Robotics Research Center, International Institute of Information Technology Hyderabad, Hyderabad 500032, India (e-mail: spandan.roy@iiit.ac.in).

Lim Yi and Mohan Rajesh Elara are with the Engineering Product Development Pillar, Singapore University of Technology and Design, Singapore 487372 (e-mail: yi_lim@sutd.edu.sg; rajeshelara@sutd.edu.sg).

Simone Baldi is with the School of Mathematics, Southeast University, Nanjing 210096, China (e-mail: s.baldi@tudelft.nl).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TASE.2023.3260099.

Digital Object Identifier 10.1109/TASE.2023.3260099

based controller was synthesized for path tracking of a reconfigurable mobile robot with three modules [8]. Liu et. al. [13] proposed a robust distributed control method for controlling a spring assisted reconfigurable robot. Instantaneous rotation based trajectory tracking control was proposed in [4] for a differential drive based self-reconfigurable robot h-Tetro. Fuzzy logic and numerical optimization based controller designs for reconfigurable mobile robots were presented in [14], [15]. The authors of [16] and [17] proposed a path tracking procedure based on resolved motion rate control. The works [18], [19] developed assembling techniques for different sub-modules into some specific robot configurations, required for manufacturing applications. The papers [20] and [21] proposed optimal and robust approaches to realize different configurations by automated docking and un-docking of multiple maritime vessels. The authors of [22] and [23] have proposed machine learning inspired path planning and area coverage techniques for tiling based reconfigurable mobile robots.

Despite the progress in the field, crucial limitations of the current state-of-the-art controllers for reconfigurable mobile robots still exist, which mainly stem from the following issues:

- The robot parameters may vary significantly after configuration changes. If not captured in the mathematical model and the control design, the configuration changes have the potential to deteriorate the closed-loop performance, or even destabilize the system [24], [25], [26]. The state-of-the-art controllers proposed for mobile robots [27], [28], [29], [30] (and references therein) do not address this issue. In fact, the simulation and experimental studies show that multiple reconfigurations (i.e., a practical scenario representing repeated narrow and wide paths) of the robot lead to significant performance loss for non-switched controllers [28], [30].
- The fault detection and compensation based controllers [29], [31], [32] are proved to provide satisfactory performances in many practical applications, but they do not account explicitly for multiple configurations. Meanwhile adaptive optimal control and neural network based approaches [33], [34], [35] are potentially suitable for robots with unmodelled dynamics, but their adaptation mechanisms rely on complex function approximations with great number of gains, which also do not account explicitly for multiple configurations. Indeed, faults or unmodelled dynamics require a different approach as compared to multiple and repeated parametric variations after configuration changes.
- Kinematic based controller designs for mobile robots [4], [7], [8], [36] do not consider other important sources of uncertainty in cleaning tasks, such as the influence of friction variation in different surfaces, gravitational forces due to inclinations in the path and external forces affecting the robot.

B. Contribution of This Study

The above discussion points out that a new strategy capturing the shape change in reconfigurable robots is compelling. As a matter of fact, a controller which can assure closed-loop stability in the presence of discontinuities (due to configuration change) and system uncertainties, while guaranteeing a good tracking performance for reconfigurable mobile robots is still lacking. The major contributions of the paper are as follows:

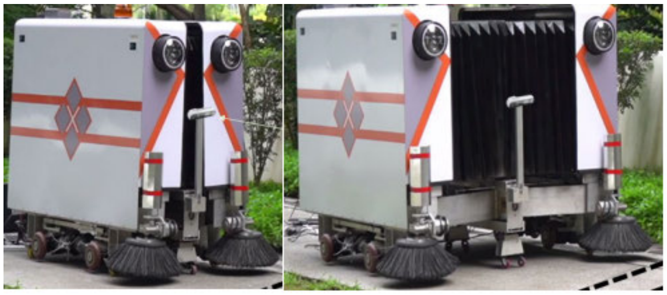


Fig. 1. PANTHERA platform in compressed and expanded configuration.

- The configuration changes in the robot during operation are captured using a switched Euler-Lagrange model.
- We propose an adaptive control design that neither needs the explicit knowledge of system dynamics before and after configuration changes, nor it requires exact upper bounds of the uncertainties like friction and external disturbances. Moreover, the number of gain parameters to adapt is comparatively less in number than adaptive optimal and neural network based approaches.
- Apart for the specific robot considered in this study, the switched dynamics formulation and the proposed adaptive control solution are discussed for a quite general stability framework called multiple Lyapunov function stability. This extends the applicability of the proposed framework to other reconfigurable mobile robots.
- The proposed design is validated on a reconfigurable pavement sweeping robot named PANTHERA (Fig. 1).

The rest of the paper is organized as follows: Sect. II discusses the architecture of the PANTHERA self-reconfigurable mobile robot, its switched dynamics modelling and its generalization to other similar platforms; the proposed adaptive switched controller is designed and analysed in Sect. III; Sect. IV presents the comparative experimental results, while Sect. V concludes the paper.

The following notations are used throughout the paper: $\lambda_{\min}(\bullet)$ and $||\bullet||$ represent minimum eigenvalue and Euclidean norm of $(\bullet)$ respectively.

II. Switched Modeling of Reconfigurable Robots

This section is divided into two parts: the first part briefly describes the architecture of PANTHERA reconfigurable robot; in the second part, the switched dynamics model is derived. It is noteworthy here that, although the work evolves around the PANTHERA platform, the subsequent modelling and control design are applicable to any similar (nonholonomic differential drive) reconfigurable mobile robots.

A. Architecture of PANTHERA

The PANTHERA reconfigurable mobile robot (cf. Fig. 1) is built to automate the cleaning of walkways [37]. It can

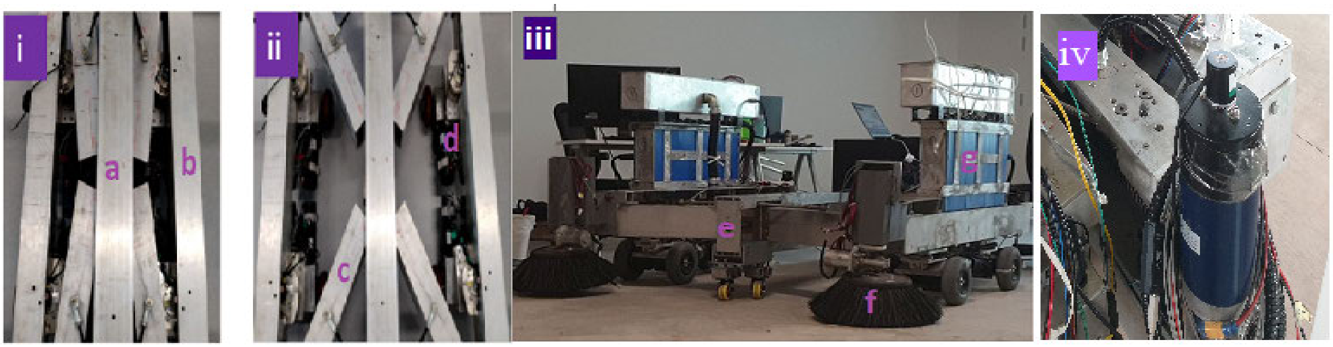


Fig. 2. PANTHERA base (i-ii) and supporting components (iii): The side frames change their positions (from expanded state (ii) to contracted state (i) and vice versa) by the linear motion of lead screw carriage: a) Central Beam; b) Side Beam; c) Guiding Links; d) Wheel Motor; e) Lead-Screw; f) Sweeping Brush; g) Battery. The associated hinged joints and the guiding links transform the linear movement into expansion and contraction (iv) Lead-Screw motor.

TABLE I

SYSTEM PARAMETERS FOR PANTHERA

Parameter	Magnitude
Robot Width (b)	Compressed: 0.5m Expanded: 1.1m
Robot Length ($2l$)	1.2m
Wheel Radius (r_w)	0.1m
Platform Mass (m)	200kg
Wheel Inertia (I_w)	$0.8kgm^2$
Platform Inertia (I_s)	Compressed: $28.3kgm^2$ Expanded: $44.2kgm^2$

change its width to a compressed state or an expanded state depending on the needs, such as navigating pavements with varying width, or avoiding obstacles. The shape change is facilitated by a reconfigurable aluminium base made of a central hollow fixed beam, two side beams and four guiding links (cf. Fig. 2 (i-ii)). The central and side beams are connected by four guiding links via hinge joints and revolute joints respectively. The configuration change is achieved through a lead screw attached with the central beam and a set of left and right handed screw threads supporting the side beams. The two sides of the screw threads are connected with a motor and a flange bearing respectively.

The lead screw is attached with a link having double sided beam through a revolute joint. The other side of the link connects the guiding link through a spherical joint. The spherical nature of the joint helps in relative rotation between two beams, and also avoids misalignment. A geared DC motor (24 V, 110 rpm) is used to rotate the lead screw (cf. Fig. 2 (iv)). The various system parameters of PANTHERA are provided in Table I.

B. Switched System Modeling

Even though each wheel is actuated by a separate set of motors, the PANTHERA platform is driven by a reverse differential steering principle: that is, according to Fig. 3, left side wheels W_{lf} and W_{lb} are given one control input while, right side wheels W_{rf} and W_{rb} are given another control input.

The pose of the platform (cf. Fig. 3) can be described by the robot center of mass (COM) position (x_c, y_c) and heading angle φ. The variable $\alpha = \tan^{-1}(b/2l)$ denotes the angle between the COM and the wheels, whereas $L = \sqrt{(b^2/4 + l^2)}$ is the diagonal distance between COM to the center of the wheels. For a given configuration (i.e., when fixed (L, α)), the relative position of the wheels with respect to the robot pose (x_c, y_c, φ) are derived as:

$$\begin{aligned}
W_{lf}(x_c, y_c, \varphi) &= x_c + L\cos(\alpha + \varphi),\ y_c + L\sin(\alpha + \varphi)\\
W_{rf}(x_c, y_c, \varphi) &= x_c + L\cos(-\alpha + \varphi),\ y_c + L\sin(-\alpha + \varphi)\\
W_{lb}(x_c, y_c, \varphi) &= x_c + L\cos(\pi - \alpha + \varphi),\ y_c\\
&\quad + L\sin(\pi - \alpha + \varphi)\\
W_{rb}(x_c, y_c, \varphi) &= x_c + L\cos(\pi + \alpha + \varphi),\ y_c\\
&\quad + L\sin(\pi + \alpha + \varphi).
\end{aligned}$$

Following the differential drive formulation as in [38] and [39], the dynamics of PANTHERA under a fixed-shape can be expressed as:

$$M_e(q)\ddot{q} + C_e(q, \dot{q})\dot{q} + F_e(\dot{q}) + d_e = \tau_e, \tag{1}$$

$$\dot{q}_R = \underbrace{\begin{bmatrix} \frac{r_w}{b}\left(\frac{b}{2}\cos(\varphi) - l\sin(\varphi)\right) & \frac{r_w}{b}\left(\frac{b}{2}\cos(\varphi) + l\sin(\varphi)\right) \\ \frac{r_w}{b}\left(\frac{b}{2}\sin(\varphi) + l\cos(\varphi)\right) & \frac{r_w}{b}\left(\frac{b}{2}\sin(\varphi) - l\cos(\varphi)\right) \\ r_w/b & -r_w/b \\ 1 & 0 \\ 0 & 1 \end{bmatrix}}_{S_e} \dot{q}, \tag{2}$$

$$\text{where } q = [\theta_r, \theta_l]^T, q_R = [x_c, y_c, \varphi, \theta_r, \theta_l]^T, \tag{3}$$

$$M_e = S_e^T M S_e = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}, C_e = S_e^T(M\dot{S}_e + CS_e),$$

$$C = \begin{bmatrix} md\dot{\varphi}^2\cos(\varphi) \\ md\dot{\varphi}^2\sin(\varphi) \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{matrix} m_{11} = m_{22} = \{m(\frac{b^2}{4} + l^2) - I_s\}\frac{4{r_w}^2}{b^2} \\ m_{12} = m_{21} = I_w + \{I_s + m(\frac{b^2}{4} - l^2)\}\frac{4{r_w}^2}{b^2} \end{matrix}$$

$$M = \begin{bmatrix} m & ml\sin(\varphi) & 0 & 0 & 0 \\ 0 & m & -ml\cos(\varphi) & 0 & 0 \\ ml\sin(\varphi) & -ml\cos(\varphi) & I_s & 0 & 0 \\ 0 & 0 & 0 & I_w & 0 \\ 0 & 0 & 0 & 0 & I_w \end{bmatrix} \tag{4}$$

where $\tau_e = [\tau_r, \tau_l]^T$; $M_e(q) \in \mathbb{R}^{n \times n}$ is the mass/inertia matrix; $C_e(q, \dot{q}) \in \mathbb{R}^{n \times n}$ denotes the Coriolis, centripetal

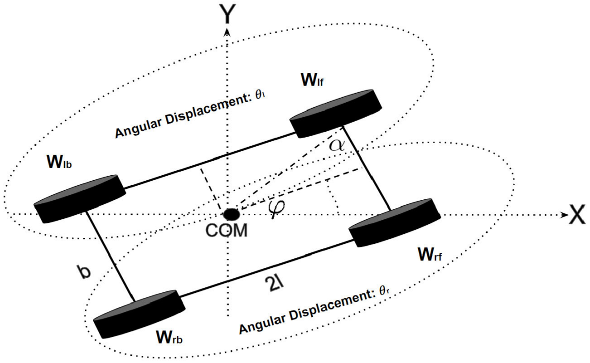

Y
X
W$_{lf}$
W$_{lb}$
W$_{rf}$
W$_{rb}$
Angular Displacement: θ_l
Angular Displacement: θ_r
α
φ
COM
b
2l

Fig. 3. Schematic of the base of PANTHERA.

terms; $F_e(\dot{q}) \in \mathbb{R}^n$ represents the vector of unknown damping and friction forces; $d_e \in \mathbb{R}^n$ denotes unknown but bounded external disturbances (e.g. mechanical vibrations arising from uneven terrain or an inclination change in the robot path). The variable dependency in M_e, C_e in (1) holds owing to the relation between φ and (θ_r, θ_l) via (2).

1) System Dynamics Under Reconfiguration: When the robotic platform reconfigures to the compressed state, the distance between wheels (b) and COM changes with it. Therefore, the associated variables m_{ij}, C_e and the inertia I_s[1] in (1)-(3) gets modified. It is reasonable to represent a self-reconfigurable mobile platform as a switched EL model as:

$$M_\sigma(q)\ddot{q} + C_\sigma(q, \dot{q})\dot{q} + F_\sigma(\dot{q}) + d_\sigma = \boldsymbol{\tau}_\sigma, \tag{5}$$

$$\dot{q}_R = \underbrace{\begin{bmatrix} \frac{r_w}{b_\sigma}\left(\frac{b_\sigma}{2}\cos(\varphi) - l_\sigma \sin(\varphi)\right) & \frac{r_w}{b_\sigma}\left(\frac{b_\sigma}{2}\cos(\varphi) + l_\sigma \sin(\varphi)\right) \\ \frac{r_w}{b_\sigma}\left(\frac{b_\sigma}{2}\sin(\varphi) + l_\sigma \cos(\varphi)\right) & \frac{r_w}{b_\sigma}\left(\frac{b_\sigma}{2}\sin(\varphi) - l_\sigma \cos(\varphi)\right) \\ r_w/b_\sigma & -r_w/b_\sigma \\ 1 & 0 \\ 0 & 1 \end{bmatrix}}_{S_\sigma} \dot{q}, \tag{6}$$

where $\sigma(t) = \{1, 2\}$, is a piecewise constant function of time, typically called the switching signal, where mode $\sigma = 1$ denotes the fully compressed configuration and mode $\sigma = 2$ denotes the expanded configuration.

During operations (i) only a finite number of reconfigurations can occur in a finite time and (ii) duration of $\sigma = 1$ and $\sigma = 2$ may be different. Therefore, the switching signal $\sigma(t)$ can be suitably classified under the average dwell time (ADT) category as defined below:

Definition 1: (ADT [40]) For a switching signal $\sigma(\cdot)$ and each $t_2 \geq t_1 \geq 0$, let $N_\sigma(t_1, t_2)$ denotes total number of switching inside the time interval $[t_1, t_2)$. Then, $\sigma(\cdot)$ has an average dwell time ϑ if for a given scalar $N_0 > 0$

$$N_\sigma(t_1, t_2) \leq N_0 + (t_2 - t_1)/\vartheta, \quad \forall t_2 \geq t_1 \geq 0.$$

2) Important Dynamical Properties: Following the celebrated EL mechanics [38], the following properties hold for PANTHERA for each σ.

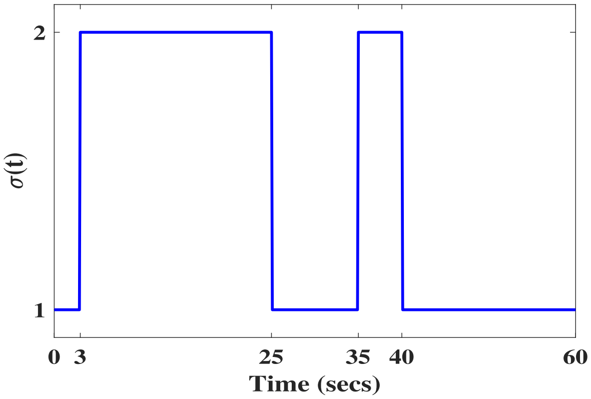


Fig. 4. Switching signal.

- Property 1: M_σ is symmetric and uniformly positive definite with respect to q. So, $\exists \underline{m}_\sigma, \overline{m}_\sigma \in \mathbb{R}^+$ such that

$$0 < \underline{m}_\sigma I \leq M_\sigma(q) \leq \overline{m}_\sigma I \tag{7}$$

- Property 2: The matrices $(\dot{M}_\sigma(q) - 2C_\sigma(q, \dot{q}))$ are skew-symmetric, i.e., $z^T(\dot{M}_\sigma(q) - 2C_\sigma(q, \dot{q}))z = 0$ for any $z \neq 0$.
- Property 3: $\exists \overline{c}_\sigma, \overline{f}_\sigma, \overline{d}_\sigma \in \mathbb{R}^+$ such that $||C_\sigma(q, \dot{q})|| \leq \overline{c}_\sigma ||\dot{q}||$, $||F_\sigma(\dot{q})|| \leq \overline{f}_\sigma ||\dot{q}||$ and $||d_\sigma(t)|| \leq \overline{d}_\sigma$.

III. Control Design: Challenges and Proposed Solution for PANTHERA

A. Motivational Scenario

To demonstrate the control challenge under switched dynamics, a simulation study is carried out in comparison with non-switched controller such as the sliding mode controller (SMC) [41] and the recently proposed disturbance observer based (DOB) controller design [29], [42]. These robust control methods rely on a priori knowledge of uncertainties and, therefore, become more convincing to motivate and to highlight the potential issues of non-switched controllers applied to a switched dynamics even when uncertainty bounds are known (state-of-the-art adaptive control strategies are compared with the proposed one later in Sects. IV and V).

In the simulation, the PANTHERA is commanded to follow a straight line path by setting same desired velocity to the wheels as $\dot{\theta}_r^d = \dot{\theta}_l^d = 4$. The system is initially at fully compressed mode ($\sigma = 1$) and changes to the fully expanded mode ($\sigma = 2$) after 3 sec interval. This is followed by several switchings between $\sigma = 1$ and $\sigma = 2$ as presented in Fig. 4. The system parameters for these two switched scenarios are taken from Table I. For different switched conditions, friction forces are selected as $F_1 = -10\tanh(q + \dot{q}) + 1.5\dot{q}$, $F_2 = 1.5\tanh(q + \dot{q}) + 1.5\dot{q}$, while the external disturbances are considered as $d_1 = 0.2\sin(7t)$ and $d_2 = -0.4\cos(6t)$.

The sliding variable $s = [s_1, s_2]^T$ for SMC is selected as

$$s = \dot{e} + 10e$$

[1] Note that the inertia values ($\frac{1}{12}m(b^2 + 4l^2)$ along z-axis) are computed by approximating the platform as a rectangular slab, excluding sensors, computing modules, sweeping brushes etc.

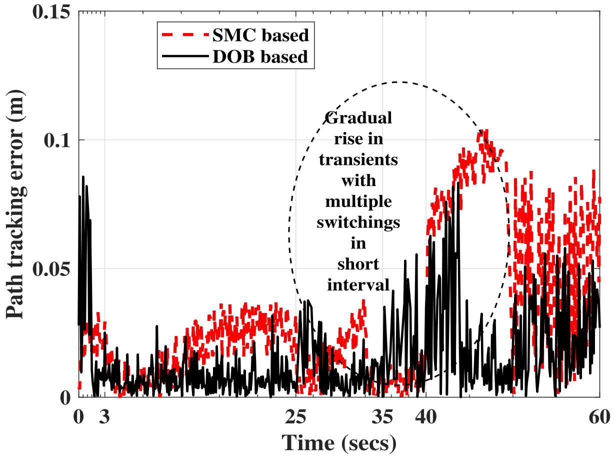


Fig. 5. Effect of configuration change for a straight desired path, under non-switched controllers.

with the control law $\tau = \tau_{eq} - 100\tau_{sw}$, where

$$\tau_{eq} = \begin{bmatrix} 70 & 0 \\ 0 & 100 \end{bmatrix} (\ddot{q}^d + \dot{e} + 2e)$$

$$\tau_{sw} = \begin{cases} \dfrac{s}{||s||} & \text{if } ||s|| \geq 0.1 \\ \dfrac{s}{0.1} & \text{if } ||s|| < 0.1. \end{cases}$$

The design parameters for DOB are taken from [30]. The gains of both SMC and DOB are tuned for the fully compressed configuration i.e., when $(\sigma = 1)$.

The path tracking errors (Euclidean distance between and x and y position error) for both the controllers are shown in Fig. 5. It can be observed that whenever the system switches from $\sigma = 1$ to $\sigma = 2$ or vice versa, both control laws lead to some undesired transients, which hampers the tracking performance. Owing to its disturbance estimation property DOB performs better after the first switching during $t = 3 - 25$ secs; however, as the uncertainties from switched dynamics keep on propagating after multiple switchings, DOB also lost considerable performance.

These results highlight that, under configuration changes, gain settings of a non-switched robust controller like SMC tuned for one mode (configuration) may not work for another mode, leading to significant drop in controller performance. Meanwhile, a DOB-based non-switched controller may work well when the average dwell time is large enough to estimate and compensate the disturbances, but may not be able to perform satisfactorily after multiple configuration changes (switchings). This situation becomes even more challenging in practical circumstances under imprecise parametric knowledge (e.g., changes in inertial values addition of sensors, computing modules, sweeping brushes etc.), unmodelled dynamics (e.g., friction) and unknown external disturbances, which calls for some switched adaptive mechanism as proposed later.

B. Problem Formulation

To summarize, the main control challenges can be listed as: (i) knowing the exact COM parameters of PANTHERA (or many other robots) is very difficult, which causes uncertainty: for example, its shape is not an ideal rectangle due to the presence of cleaning brushes and hollow cover which is a combination of steel and rubber (cf. Fig. 1); (ii) knowing the exact mass and inertia parameters is also difficult, due to non-uniform weight distribution along the platform and distance between wheels during reconfiguration; (iii) friction force parameters are always difficult to model due their dependence on road conditions; (iv) all the aforementioned parameters may deeply change after reconfiguration, while the reconfiguration can even generate transients (e.g. as in Fig. 4) that can build up in destabilizing way for the robot. These challenges lead to uncertainty in M_σ, C_σ and F_σ, under a time-varying σ (e.g. as in Definition 1); these uncertainty and changes should be tackled by the control law.

In view of the above discussions, we summarize the system uncertainties in the following assumption:

Assumption 1: For all σ, the scalars $\underline{m}_\sigma$ and $\overline{m}_\sigma$ of (5) are available; but, the terms C_σ, F_σ, d_σ and their upper bounds, i.e., $\overline{c}_\sigma$, $\overline{g}_\sigma$, $\overline{f}_\sigma$, $\overline{d}_\sigma$ are unknown.

It is well established (cf. [38], [39]) that the goal of following a desired path by the robot can be suitably converted as tracking a desired trajectory q^d with bounded desired velocity $\dot{q}^d$ and bounded desired acceleration $\ddot{q}^d$: for example, a desired circular (or straight line) path can be followed by giving a differential (same) velocity to the wheels; or a lawn mower type path can be followed by designing a sinusoidal differential desired velocity.

The control objective is outlined as: *Under Assumption 1 and Properties 1-3, derive a switched adaptive control law τ_σ such that the trajectories $q(t)$ of the reconfigurable platform (5) follow the desired trajectory $q^d(t)$.*

The following subsection describes the solution of the control problem.

C. Switched Controller Design and Analysis

Let us define the tracking error as $e \triangleq q - q^d$, and a composite error variable r as

$$r \triangleq \dot{e} + \Phi e, \tag{8}$$

where the matrix $\Phi \in \mathbb{R}^{2\times 2}$ is positive definite. From (5), we get the dynamics

$$\begin{aligned} M_\sigma \dot{r} &= M_\sigma(\ddot{q} - \ddot{q}^d + \Phi\dot{e}) \\ &= \tau_\sigma - C_\sigma r + \varphi_\sigma, \end{aligned} \tag{9}$$

where $\varphi_\sigma \triangleq -(C_\sigma \dot{q} + F_\sigma + d_\sigma + M_\sigma \ddot{q}^d - M_\sigma \Phi \dot{e} - C_\sigma r)$ is the cumulative uncertainty.

Let us define an augmented error variable $\xi = [e^T \ \dot{e}^T]^T$. It follows from definition of ξ that, $||\xi|| \geq ||e||$, $||\xi|| \geq ||\dot{e}||$. Exploiting the structural properties of the system and the assumptions, one can derive [24]

$$||\varphi_\sigma|| \leq \eta^*_{0\sigma} + \eta^*_{1\sigma}||\xi|| + \eta^*_{2\sigma}||\xi||^2 \triangleq Y^T_\sigma(||\xi||)\eta^*_\sigma, \forall \sigma \in \Omega \tag{10}$$

where $\eta^*_{i\sigma} \in \mathbb{R}^+$ $i = 0, 1, 2$ are *unknown* but finite scalars, $Y_\sigma \triangleq [1 \ ||\xi|| \ ||\xi||^2]^T$ and $\eta^*_\sigma = [\eta^*_{0\sigma} \ \eta^*_{1\sigma} \ \eta^*_{2\sigma}]^T$.

Using (10), a switched adaptive control law τ_σ is proposed as

$$\tau_\sigma = -\Lambda_\sigma r - e - \Delta\tau_\sigma,\ \ \Delta\tau_\sigma = \begin{cases} \zeta_\sigma \dfrac{r}{||r||} & \text{if } ||r|| \geq \varpi \\ \zeta_\sigma \dfrac{r}{\varpi} & \text{if } ||r|| < \varpi, \end{cases} \tag{11}$$

$$\zeta_\sigma = \hat{\eta}_{0\sigma} + \hat{\eta}_{1\sigma}||\xi|| + \hat{\eta}_{2\sigma}||\xi||^2 + \gamma_\sigma \triangleq Y_\sigma^T(||\xi||)\hat{\eta}_\sigma + \gamma_\sigma, \tag{12}$$

where $\Lambda_\sigma > 0$ is a user-defined gain matrix; $\Delta\tau_\sigma$ handles the uncertainties via the variable ζ_σ; the positive constant $\varpi > 0$ is used for boundary layer control; $\hat{\eta}_\sigma \triangleq [\hat{\eta}_{0\sigma}\ \ \hat{\eta}_{1\sigma}\ \ \hat{\eta}_{2\sigma}]^T$ represents the estimate of the vector η_σ^* and γ_σ is an auxiliary gain used for closed-loop stabilization (cf. Remark 1 at the end of stability analysis).

Let $\overline{\sigma}$ denotes the inactive subsystem in the interval $t \in [t_l\ t_{l+1})$ (i.e., $\overline{\sigma} = 2$ (resp. 2 when $\sigma = 1$ (resp. 1)). The controller gains $\hat{\eta}_{i\sigma}, \gamma_\sigma$ are computed using the adaptive laws expressed as:

$$\dot{\hat{\eta}}_{i\sigma} = ||r||||\xi||^i - \alpha_{i\sigma}\hat{\eta}_{i\sigma}, \quad \dot{\gamma}_\sigma = 0 \tag{13a}$$

$$\dot{\hat{\eta}}_{i\overline{\sigma}} = 0, \quad \dot{\gamma}_{\overline{\sigma}} = -\left(\beta_{\overline{\sigma}} + \frac{1}{2}\sum_{i=0}^{2}\hat{\eta}_{\overline{\sigma}}^2\right)\gamma_{\overline{\sigma}} + \beta_{\overline{\sigma}}\nu_{\overline{\sigma}}, \tag{13b}$$

$$\text{with } \hat{\eta}_{i\sigma}(t_0) > 0,\ \ \gamma_{\overline{\sigma}}(t_0) > \nu_{\overline{\sigma}}, \tag{13c}$$

where $\alpha_{i\sigma}, \beta_{\overline{\sigma}}, \nu_{\overline{\sigma}} \in \mathbb{R}^+$, $i = 0, 1, 2$, are design constants and t_0 is the initial time.

Let us define $\bar{\varrho}_M \triangleq \max\{\overline{m}_\sigma\}$ and $\underline{\varrho}_m \triangleq \min\{\underline{m}_\sigma\}$ for $\sigma = 1, 2$. Following the definition of ADT, the reconfiguration changes are defined via the switching law

$$\vartheta > \ln\mu/\kappa, \tag{14}$$

where $\mu \triangleq \bar{\varrho}_M/\underline{\varrho}_m$; $0 < \kappa < \iota$ and

$$\iota = \frac{\min_{i,\sigma}\{\lambda_{\min}(\Lambda_\sigma), \lambda_{\min}(\Phi), (\alpha_{i\sigma}/2)\}}{\max_\sigma\{\overline{m}_\sigma, 1, (1/2)\}}, \quad i = 0, 1, 2,\ \ \sigma = 1, 2.$$

The following main stability theorem holds.

Theorem 1: Let the Properties 1-3 and Assumption 1 hold, and let the wheel input torques be selected as (11)-(13c). Then, under any reconfiguration scenario satisfying the ADT (14), the closed-loop trajectories of the reconfigurable mobile robot (5) remain Uniformly Ultimately Bounded (UUB).

Proof: See Appendix.

IV. Simulations, Experiments and Discussion

In this section, we verify the effectiveness of the proposed controller compared to the state-of-the-art using the PANTHERA reconfigurable robot via simulations and experiments.

A. Verification Scenario and Control Parameter Selection

To test the effectiveness of the proposed controller, a common testing scenario for both simulation and experiments along with same control design parameters is designed. The PANTHERA is tasked to clean a pathway by moving in a circular path of 2m radius: the platform starts with compressed mode ($\sigma = 1$), and it changes its configuration to the expanded mode ($\sigma = 2$) at $t = 3$ sec, and then again changes its configuration to $\sigma = 1$ at $t = 25$ sec. The configuration change follows the switching signal presented in Fig. 4.

The desired circular path is generated by setting two different desired wheel positions

$$\theta_r^d = 4t \text{ rad}, \quad \theta_l^{\mathrm{d}} = 3.5t \text{ rad},$$

which are fed to the wheels on the respective sides in PANTHERA [29], [39]. The kinematics (6) is solved using (η_r^d, η_l^d) to obtain the desired platform position.

Based on the nominal inertia parameters as in Table I and on the mass of various associated hardwares (supplied by the manufacturer), the bounds are computed as

$$18I \leq M_1(q) \leq 75I, 18I \leq M_2(q) \leq 60I.$$

The design matrices are chosen as

$$\Lambda_1 = 30I, \Lambda_2 = 40I, \Phi = 10I$$

leading to $\mu = 4.16, \iota = 0.133$. These selections along with the choice of $\kappa = 0.1$ results into $\ln\mu/\kappa = 12.8$ sec. From the switching signal as in Fig. 4, the ADT can be calculated as $\vartheta = 15$ sec (4 switchings in 60 sec of operation), which satisfies the ADT condition (14). Other control design parameters are chosen as $\alpha_{i\sigma} = 1.5, \beta_\sigma = 2, \nu_\sigma = 0.1$ $\varpi = 0.1$ for $i = 0, 1, 2,\ \sigma = 1, 2$ with the initial values $\hat{\eta}_{i\sigma}(0) = \gamma_{i\sigma}(0) = 0.5$.

To properly judge the effectiveness of the proposed adaptive switching controller, its performance is compared with the adaptive sliding mode controller (ASMC) [43]. For ASMC, the sliding variable s is selected as

$$s = \dot{e} + 10e$$

with the control law

$$\tau = \tau_{eq} - K_{sw}\tau_{sw},\ \ \tau_{eq} = \begin{bmatrix} 70 & 0 \\ 0 & 100 \end{bmatrix}(\ddot{q}_d + \dot{e} + 2e),$$

$$\tau_{sw} = \begin{cases} \dfrac{s}{||s||} & \text{if } ||s|| \geq 0.1 \\ \dfrac{s}{0.1} & \text{if } ||s|| < 0.1 \end{cases},\ \ K_{sw} = \begin{bmatrix} K_{sw_1} & 0 \\ 0 & K_{sw_2} \end{bmatrix}$$

and the gains K_{sw_i} $(i = 1, 2)$ are adapted as

$$\dot{K}_{sw_i} = \begin{cases} 10||s_i||\text{sign}(s_i - (0.04\ K_{sw_i})) & \text{if } ||K_{sw_i}|| \geq 0.1 \\ 0.1 & \text{if } ||K_{sw_i}|| < 0.1 \end{cases}$$

with $K_{sw_i}(0) = 0.5$ and $\text{sign}(\cdot)$ represents signum function.

For designing the DOB [42] based robust controller, the high gain disturbance observer gain is selected as $\frac{1}{\epsilon} = 100$, and the high gain controller gain is chosen as $\frac{1}{\mu} = 50$. The coefficients of Hurwitz polynomial required for controller design are taken 20, 21 for both side of wheels (for details refer to [29]).

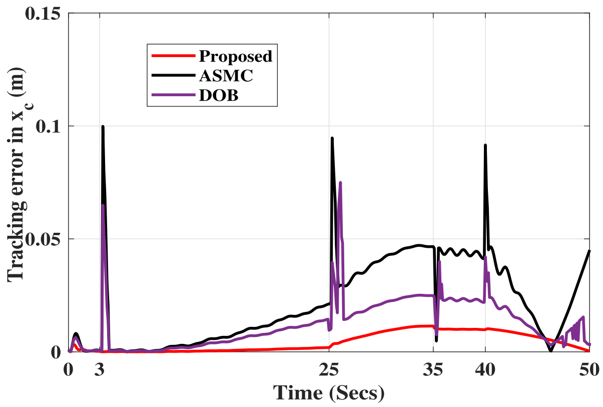


Fig. 6. Tracking error comparison in x_c position.

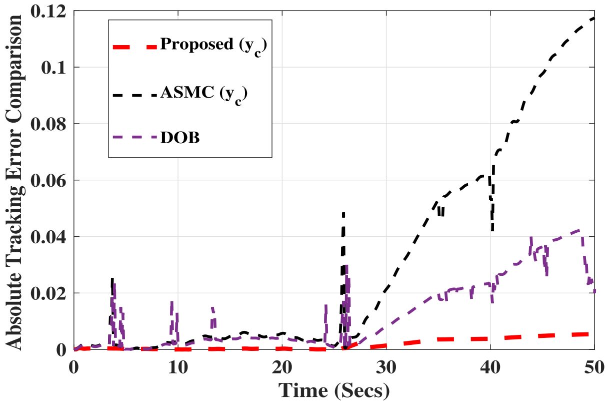


Fig. 7. Tracking error comparison in y_c position.

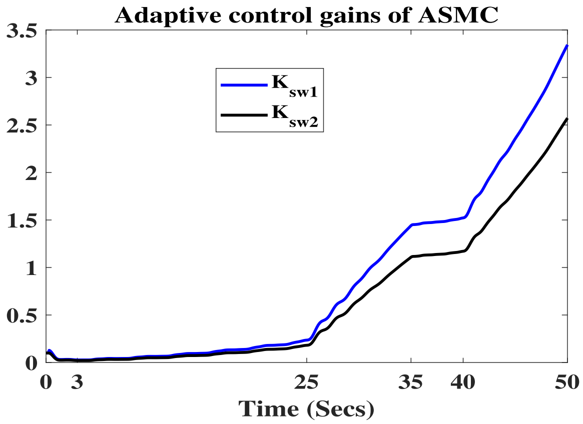


Fig. 8. Adaptive gains of ASMC.

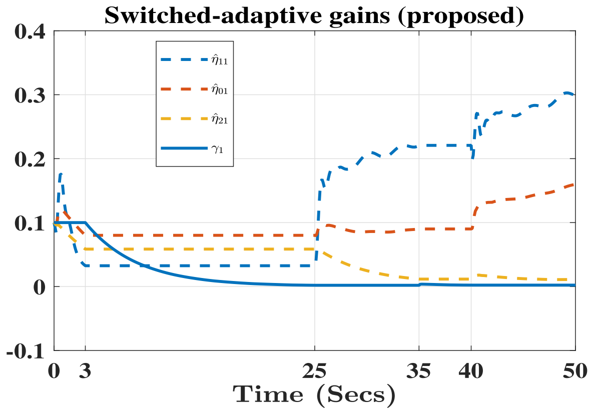


Fig. 9. Adaptive switched control gains of the proposed controller for $\sigma = 1$.

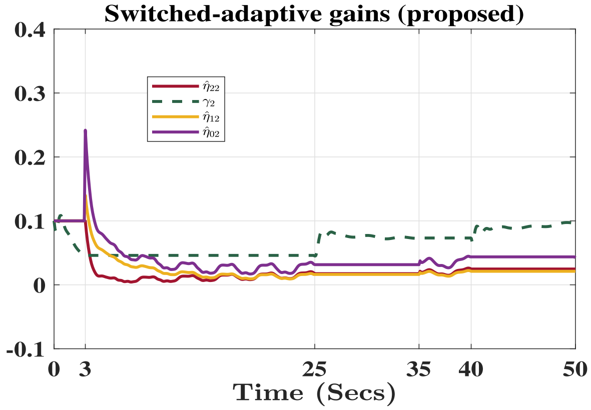


Fig. 10. Adaptive switched control gains of the proposed controller for $\sigma = 2$.

B. Simulation Results and Analysis

For simulation, the combination of friction forces, state-dependent unmodelled dynamics and unknown external disturbances for various configurations are selected as

$$F_1 + d_1 = 1.3\dot{q} + q\dot{q} + 0.7\sin(2t) + 1.3\text{sign}(\dot{q}),$$
$$F_2 + d_2 = 1.3\dot{q} + 1.5q\dot{q} - 0.2\cos(3t) - 0.5\text{sign}(\dot{q}).$$

Figures 6-7 reveal that ASMC suffers from unwanted spikes in the error plots at every switching instants. The reason can be attributed to the monotonically increasing nature of the adaptive gains of ASMC (Fig. 8), which is not suitable for handling sudden parametric changes due to switching; in fact, such high gain may lead to instability in longer time frame. The DOB design performs better than ASMC as the time-varying disturbances are compensated, but they lead to more transients compared to ASMC. The proposed control design, on the other hand, does not show any spikes in its response, thanks to its dedicated set of gains for each configuration as in Figs. 9-10: when gains η_{i1} for configuration $\sigma = 1$ are active, gains η_{i2} for configuration $\sigma = 2$ remain constant (i.e., not updated) for the corresponding time duration, and vice-versa. As a result, the proposed controller can successfully negotiate the uncertainties for each configuration without any unwanted transients. Note that the gains γ_1 and γ_2 are responsible for closed-loop stabilization for the inactive configurations, and hence, they are active when η_{i1} and η_{i2} are inactive, respectively (cf. Figs. 9-10).

C. Experiments and Discussion

1) PANTHERA Setup: The electronics assembly of PANTHERA is schematically presented in Fig. 11. The sensory measurements for task space position, and wheel rotations are gathered by a LiDAR (Velodyne Puck VLP-16) and optical encoders respectively. The LiDAR, mounted on the top of the platform, is connected to a computer (octa core CPU,

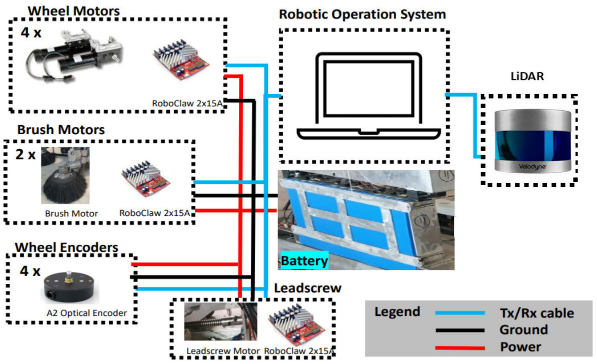


Fig. 11. PANTHERA electronics layout.

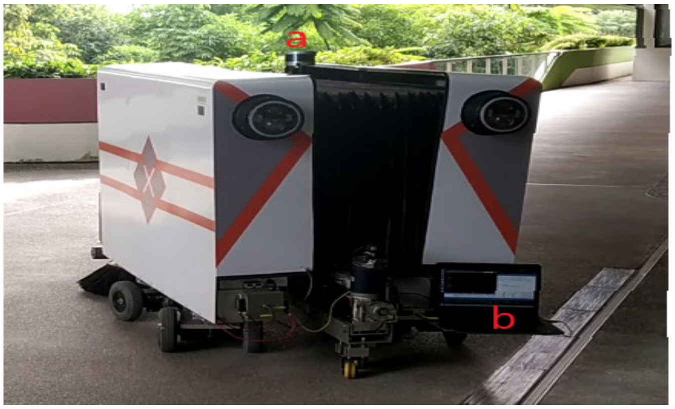


Fig. 12. PANTHERA experimental setup: a) LiDAR placement and b) workstation.

16GB RAM, dedicated GPU) as in Fig. 12. The workstation uses the Robot Operating System (ROS)-PYTHON setup for implementing control algorithm. The LiDAR odometry data is subscribed and published through the rosbag files in the ROS. The wheels on one side of the robot receive same torque input through the associated motor controllers.

2) Results and Analysis: Figure 13 shows the PANTHERA platform tracking a circular trajectory while going through configuration changes according to the switching signal Fig. 4. The tracking performance comparison between the proposed switched adaptive controller, ASMC and DOB are given in Figs. 14-15. The control inputs of the proposed switched controller is shown via Fig. 16. For better inference, the performance comparisons are further tabulated in Table II in terms of root-mean-squared (RMS) path error, including percentage tracking error reduction for the proposed controller over the other controllers. A few important observations follow from the last column of Table II: after switching back to $\sigma = 1$ from $\sigma = 2$ at $t = 25$ sec and $t = 40$ sec, VSMC and DOB loose more performances compared to the proposed controller than the previous mode of $\sigma = 2$ for $t \in [3\ 25)$ and $t \in [40\ 50)$, respectively. Non-switched adaptive control performed satisfactorily (with slight deterioration of 12% and 17%) owing to its adaptive gain nature (compensating disturbances for DOB case); however, it has to learn or adapt every time the configuration (or mode) changes. The DOB controller performs better than ASMC after configuration changes, but, still suffers from some unwanted transients due to its dependency on high gain parameters (unlike the proposed case). Whereas, for the proposed switched adaptive design, the adaptive gains $\hat{\eta}_i$'s are not updated when switched-out to another mode. As a result, the adaptive gains do not need to go through the learning transients again, leading to better RMS error when $\sigma = 1$ or $\sigma = 2$ is repeated again (cf. Table II).

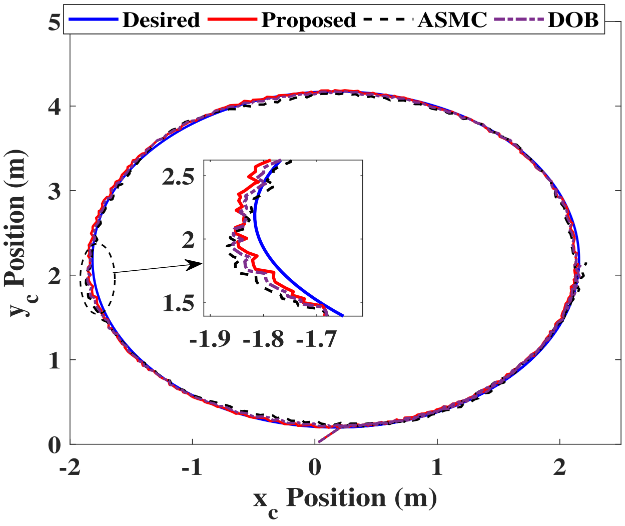


Fig. 13. Circular path tracking comparison.

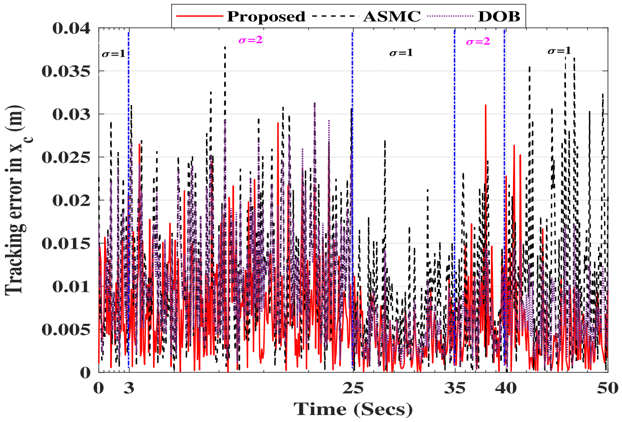


Fig. 14. Tracking error comparison (x_c) for different controllers.

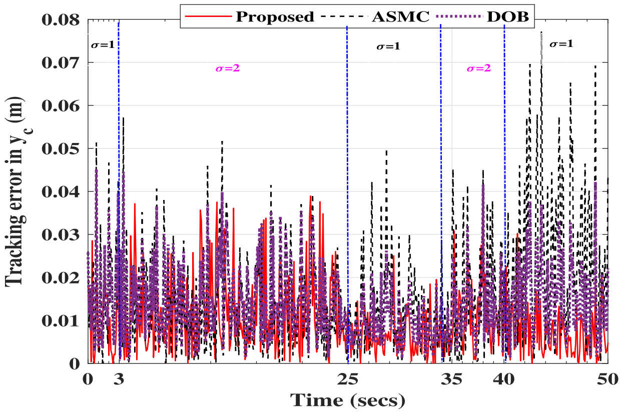


Fig. 15. Tracking error comparison (y_c) for different controllers.

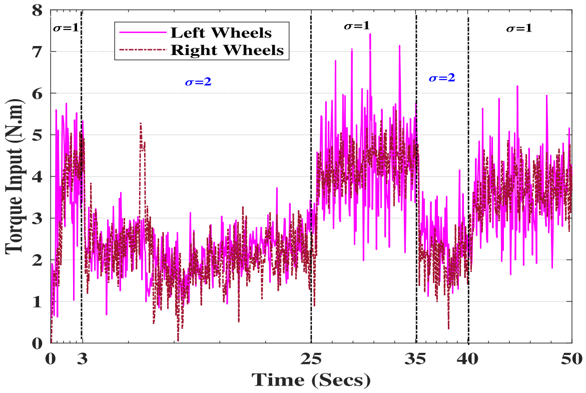


Fig. 16. Control inputs for the proposed approach in different configurations.

TABLE II
RMS Path Tracking Error (M) Comparison

Configuration	ASMC	DOB	Proposed	Tracking error reduction over (ASMC, DOB)
$\sigma = 1,\ t \in [0-3)$	0.018	0.017	0.015	(17%, 12%)
$\sigma = 2,\ t \in [3-25)$	0.019	0.017	0.016	(17%, 6%)
$\sigma = 1,\ t \in [25-35)$	0.011	0.009	0.006	(45%, 33%)
$\sigma = 2,\ t \in [35-40)$	0.012	0.012	0.010	(17%, 17%)
$\sigma = 1,\ t \in [40-50)$	0.015	0.010	0.008	(47%, 20%)

V. Conclusion

A new approach is proposed for modeling and control of self-reconfigurable mobile platforms. The configuration changes and the uncertainties are captured through a novel uncertain switched Euler-Lagrangian dynamics. The proposed switched controller does not need any structural knowledge except the upper and lower bounds of the mass-inertia matrices. It adaptively compensates for the uncertainties in the system dynamics, the discontinuities due to the configuration changes and the external disturbances. The performance of the controller is tested by implementing it on a pavement cleaning reconfigurable robot called PANTHERA. The results are promising, and motivate to extend the idea into other classes of reconfigurable platforms.

Appendix
Proof of Theorem 1

Applying the standard results of linear time-varying systems to (13a) one has [44]

$$\hat{\eta}_{i\sigma}(t) = \underbrace{\exp(-\alpha_{i\sigma} t)\hat{\eta}_{i\sigma}(0)}_{\geq 0} + \underbrace{\int_0^t \exp(-\alpha_{i\sigma}(t-\tau))(||r||||\xi||^i)\mathrm{d}\tau}_{\geq 0}$$
$$\Rightarrow\ \hat{\eta}_{i\sigma} \geq 0,\ \forall t \geq 0,\quad \forall \sigma \in \{1,2\}, \forall t \geq 0. \quad (15)$$

Further, (13b) and initial conditions for the gains (13c), one can deduce $\exists\ \underline{\gamma}, \bar{\gamma} \in R^+$ such that

$$0 < \underline{\gamma} \leq \gamma_{i\sigma}(t) \leq \bar{\gamma},\quad \forall \sigma \in \{1,2\}, \forall t \geq 0. \quad (16)$$

Exploiting (16), choose a candidate Lyapunov function:

$$V(t) = \frac{1}{2}r^T(t)M_{\sigma(t)}r(t) + \frac{1}{2}e(t)^T e(t) + \sum_{\sigma=1}^{2}\{\sum_{i=0}^{2}\frac{1}{2}(\hat{\eta}_{i\sigma}(t) - \eta^*_{i\sigma})^2 + \gamma_\sigma(t)/\underline{\gamma}\}. \quad (17)$$

The function $V(t)$ can be discontinuous as M_σ changes during reconfigurations. Therefore, the stability analysis consists of two stages: studying the closed-loop dynamics (i) at the switching instant and (ii) in-between two consecutive configuration changes. These studies are carried out subsequently:

Stability at the switching instant: Let $\sigma(t^-_{l+1})$ and $\sigma(t_{l+1})$ be the configurations before and after the switching (reconfiguration) at $t_{l+1},\ l \in \mathbb{N}^+$. Then, before and after the reconfiguration we have

$$V(t^-_{l+1}) = \frac{1}{2}r^T(t^-_{l+1})M_{\sigma(t^-_{l+1})}r(t^-_{l+1}) + \frac{1}{2}e^T(t^-_{l+1})e(t^-_{l+1}) + \frac{1}{2}\sum_{\sigma=1}^{2}\{\sum_{i=0}^{2}(\hat{\eta}_{i\sigma}(t^-_{l+1}) - \eta^*_{i\sigma})^2 + \gamma_\sigma(t^-_{l+1})/\underline{\gamma}\},$$
$$V(t_{l+1}) = \frac{1}{2}r^T(t_{l+1})M_{\sigma(t_{l+1})}r(t_{l+1}) + \frac{1}{2}e^T(t_{l+1})e(t_{l+1}) + \frac{1}{2}\sum_{\sigma=1}^{2}\{\sum_{i=0}^{2}(\hat{\eta}_{i\sigma}(t_{l+1}) - \eta^*_{i\sigma})^2 + \gamma_\sigma(t_{l+1})/\underline{\gamma}\}.$$

Owing to the continuity property of the system states and of the parameter update laws (13), we have $r(t^-_{l+1}) = r(t_{l+1})$, $e(t^-_{l+1}) = e(t_{l+1})$, $(\hat{\eta}_{i\sigma}(t^-_{l+1}) - \eta^*_{i\sigma}) = (\hat{\eta}_{i\sigma}(t_{l+1}) - \eta^*_{i\sigma})$ and $\gamma_\sigma(t^-_{l+1}) = \gamma_\sigma(t_{l+1})$. This leads to

$$V(t_{l+1}) - V(t^-_{l+1}) = \frac{1}{2}r^T(t_{l+1})(M_{\sigma(t_{l+1})} - M_{\sigma(t^-_{l+1})})r(t_{l+1}) \leq \frac{\bar{\varrho}_M - \underline{\varrho}_m}{\underline{\varrho}_m}V(t^-_{l+1}) \Rightarrow V(t_{l+1}) \leq \mu V(t^-_{l+1}), \quad (18)$$

with $\mu = \bar{\varrho}_M/\underline{\varrho}_m \geq 1$.

Stability in-between two configuration changes: The evolution of $V(t)$ in-between two consecutive configuration changes when $(t \in [t_l \quad t_{l+1}))$ is analyzed subsequently.

From the equations (9), (11) and the fact $\dot{e} = (r - \mathbf{\Phi}e)$, the differentiation of (17) with respect to time yields

$$\dot{V}(t) = r^T(t)(\tau_\sigma - C_\sigma r + \varphi_\sigma) + (1/2)r^T(t) \times \dot{M}_{\sigma(t^-_{l+1})}r(t) - \Phi e(t)) + e^T(t)(r(t) + \sum_{\sigma=1}^{2}\left\{\sum_{i=0}^{2}(\hat{\eta}_{i\sigma}(t) - \eta^*_{i\sigma})\dot{\hat{\eta}}_{i\sigma}(t)\right\} + \dot{\gamma}_{\bar{\sigma}}(t)/\underline{\gamma}$$
$$\leq r^T(t)(-\Lambda_{\sigma(t^-_{l+1})}r(t) - \Delta\tau_{\sigma(t^-_{l+1})} + \varphi_{\sigma(t^-_{l+1})}) + (1/2)r^T(t)(\dot{M}_{\sigma(t^-_{l+1})} - 2C_{\sigma(t^-_{l+1})})r(t) - e^T(t)\Phi e(t) + \sum_{\sigma=1}^{2}\left\{\sum_{i=0}^{2}(\hat{\eta}_{i\sigma}(t) - \eta^*_{i\sigma})\dot{\hat{\eta}}_{i\sigma}(t)\right\} + \dot{\gamma}_{\bar{\sigma}}(t)/\underline{\gamma}. \quad (19)$$

Application of Property 2 to the second term of (19) yields

$$\begin{aligned}\dot{V}(t) \leq\ & r^T(t)(-\Lambda_{\sigma(t_{l+1}^-)}r(t) - \Delta\tau_{\sigma(t_{l+1}^-)} + \varphi_{\sigma(t_{l+1}^-)})\\ & - e^T(t)\Phi e(t) + \sum_{\sigma=1}^{2}\left\{\sum_{i=0}^{2}(\hat{\eta}_{i\sigma}(t) - \eta_{i\sigma}^*)\dot{\hat{\eta}}_{i\sigma}(t)\right\}\\ & + \dot{\gamma}_{\overline{\sigma}}(t)/\underline{\gamma}. \end{aligned} \tag{20}$$

Two scenarios arise from the structure of $\Delta\boldsymbol{\tau}_\sigma$ in (11), (i) S1: $||r|| \geq \varpi$ and (ii) S2: $||r|| < \varpi$, and their analysis are carried out subsequently:

Scenario S1: The adaptation laws (13) are such that the control gains γ_σ do not change for the active configurations and $\hat{\eta}_{i\overline{\sigma}}$ don't change for inactive configurations. Furthermore, $\varphi_\sigma(.)$ is upper bounded by (10) and $\gamma_{i\sigma(t)}(t) > 0\ \forall t \geq t_0$.

Hence, (20) can be simplified as:

$$\begin{aligned}\dot{V}(t) \leq\ & -r^T(t)\Lambda_{\sigma(t_{l+1}^-)}r(t) - e^T(t)\Phi e(t) - Y_{\sigma(t_{l+1}^-)}^T(\hat{\eta}_{\sigma(t_{l+1}^-)}\\ & - \eta_{\sigma(t_{l+1}^-)}^*)||r(t)|| + \sum_{i=0,\sigma(t_{l+1}^-)}^{2}(\hat{\eta}_{i\sigma}(t) - \eta_{i\sigma}^*)\dot{\hat{\eta}}_{i\sigma}(t)\\ & + \dot{\gamma}_{\overline{\sigma}}(t)/\underline{\gamma}. \end{aligned} \tag{21}$$

Using the adaptive law (13a) we have

$$\begin{aligned}\sum_{i=0}^{2}(\hat{\eta}_{i\sigma} - \eta_{i\sigma}^*)\dot{\hat{\eta}}_{i\sigma} &= \sum_{i=0}^{2}||r||(\hat{\eta}_{i\sigma} - \eta_{i\sigma}^*)||\boldsymbol{\xi}||^i\\ &\quad + \alpha_{i\sigma}(\hat{\eta}_{i\sigma}\eta_{i\sigma}^* - \hat{\eta}_{i\sigma}^2)\\ &\leq Y_\sigma^T(\hat{\eta}_\sigma - \eta_\sigma^*)||r||\\ &\quad - \sum_{i=0}^{2}\frac{\alpha_{i\sigma}}{2}\{(\hat{\eta}_{i\sigma} - \eta_{i\sigma}^*)^2 + (\hat{\eta}_{i\sigma}^*)^2\}. \end{aligned} \tag{22}$$

where the last inequality arrives from

$$\begin{aligned}\hat{\eta}_{i\sigma}\eta_{i\sigma}^* - \hat{\eta}_{i\sigma}^2 &= \frac{1}{2}(\hat{\eta}_{i\sigma} - \eta_{i\sigma}^*)^2 - \frac{\hat{\eta}_{i\sigma}^2}{2} + \frac{(\hat{\eta}_{i\sigma}^*)^2}{2}\\ &\leq \frac{1}{2}\{(\hat{\eta}_{i\sigma} - \eta_{i\sigma}^*)^2 + (\hat{\eta}_{i\sigma}^*)^2\}. \end{aligned}$$

Similarly, using (13b) we have

$$\frac{\dot{\gamma}_{\overline{\sigma}}}{\underline{\gamma}} = -\frac{\left(\beta_{\overline{\sigma}} + (1/2)\sum_{i=0}^{2}\hat{\eta}_{i\overline{\sigma}}^2\right)\gamma_{\overline{\sigma}} + \beta_{\overline{\sigma}}\nu_{\overline{\sigma}}}{\underline{\gamma}}. \tag{23}$$

As $\gamma_{\overline{\sigma}} \geq \underline{\gamma}\ \ \forall t \geq t_0$ from (16), (23) yields

$$\frac{\dot{\gamma}_{\overline{\sigma}}}{\underline{\gamma}} \leq -\beta_{\overline{\sigma}}\frac{\gamma_{\overline{\sigma}}}{\underline{\gamma}} - (1/2)\sum_{i=0}^{2}\hat{\eta}_{i\overline{\sigma}}^2 + \frac{\beta_{\overline{\sigma}}\nu_{\overline{\sigma}}}{\underline{\gamma}}. \tag{24}$$

Substitution of (22) and (24) in the equations (21) gives

$$\begin{aligned}\dot{V}(t) \leq\ & -\lambda_{\min}(\Lambda_{\sigma(t_{l+1}^-)})||r(t)||^2 - \lambda_{\min}(\Phi)||e(t)\ ||^2\\ & - \sum_{i=0,\sigma(t_{l+1}^-)}^{2}\frac{\alpha_{i\sigma}}{2}\{(\hat{\eta}_{i\sigma} - \eta_{i\sigma}^*)^2 + (\hat{\eta}_{i\sigma}^*)^2\}\\ & - \left(\beta_{\overline{\sigma}}\frac{\gamma_{\overline{\sigma}}}{\underline{\gamma}} + (1/2)\sum_{i=0}^{2}\hat{\eta}_{i\overline{\sigma}}^2 - \frac{\beta_{\overline{\sigma}}\nu_{\overline{\sigma}}}{\underline{\gamma}}\right). \end{aligned} \tag{25}$$

The definition of candidate Lyapunov function (17) produces

$$V \leq \overline{m}_\sigma||r||^2 + ||e||^2 + \sum_{\sigma=1}^{2}\{\sum_{i=0}^{2}\frac{1}{2}(\hat{\eta}_{i\sigma} - \eta_{i\sigma}^*)^2 + \gamma_\sigma/\underline{\gamma}\}. \tag{26}$$

Using (26) and the definitions for ϱ, $\alpha_{i\sigma}$, $\beta_{i\overline{\sigma}}$, the equation (25) is simplified to

$$\dot{V}(t) \leq \quad -\iota V(t) + \frac{\gamma_\sigma(t)}{\underline{\gamma}} + \sum_{\sigma=1}^{2}\sum_{i=0}^{2}\frac{(\hat{\eta}_{i\sigma}^*)^2}{2} + \frac{\beta_{\overline{\sigma}}\nu_{\overline{\sigma}}}{\underline{\gamma}}, \tag{27}$$

where ι is defined in (14). Let us define $0 < \kappa < \iota$ and using (16), (27) can be simplified to

$$\dot{V}(t) \leq -\kappa V(t) - (\iota - \kappa)V(t) + \delta, \tag{28}$$

where $\delta \triangleq \max_{\forall\sigma,\overline{\sigma}}\{\frac{\bar{\gamma}}{\underline{\gamma}} + \sum_{i=0}^{2}\frac{(\hat{\eta}_{i\sigma}^*)^2}{2} + \frac{\beta_{\overline{\sigma}}\nu_{\overline{\sigma}}}{\underline{\gamma}}\}$.

Scenario S2: For this scenario, $||r|| < \varpi$. So, we get

$$\begin{aligned}\dot{V}(t) \leq\ & -r^T(t)\Lambda_{\sigma(t_{l+1}^-)}r(t) - e^T(t)\Phi e(t) - \zeta_{\sigma(t_{l+1}^-)}(||r(t)||^2/\varpi)\\ & + Y_{\sigma(t_{l+1}^-)}^T\eta_{\sigma(t_{l+1}^-)}^*||r(t)||\\ & + \sum_{i=0,\sigma(t_{l+1}^-)}^{2}(\hat{\eta}_{i\sigma}(t) - \eta_{i\sigma}^*)\dot{\hat{\eta}}_{i\sigma}(t) + \dot{\gamma}_{\overline{\sigma}}(t)/\underline{\gamma}\\ \leq\ & -r^T(t)\Lambda_{\sigma(t_{l+1}^-)}r(t) - e^T(t)\Phi e(t)\\ & + Y_{\sigma(t_{l+1}^-)}^T\eta_{\sigma(t_{l+1}^-)}^*||r(t)||\\ & + \sum_{i=0,\sigma(t_{l+1}^-)}^{2}(\hat{\eta}_{i\sigma}(t) - \eta_{i\sigma}^*)\dot{\hat{\eta}}_{i\sigma}(t) + \dot{\gamma}_{\overline{\sigma}}(t)/\underline{\gamma}. \end{aligned} \tag{29}$$

Following the similar lines of arguments as pursued in Scenario $S1$, we get

$$\dot{V}(t) \leq -\kappa V(t) - (\iota - \kappa)V(t) + Y_{\sigma(t_{l+1}^-)}^T\hat{\eta}_{\sigma(t_{l+1}^-)}||r(t)|| + \delta. \tag{30}$$

From (8) it can be verified that $||r|| < \varpi \Rightarrow ||\boldsymbol{\xi}|| \in \mathcal{L}_\infty$. Consequently, the adaptation law (13a) implies $||r||, ||\boldsymbol{\xi}|| \in \mathcal{L}_\infty \Rightarrow \hat{\eta}_{i\sigma}(t)\ \in \mathcal{L}_\infty$. So, there exists a $\delta_1 \in \mathbb{R}^+$ such that

$$Y_{\sigma(t_{l+1}^-)}^T\hat{\boldsymbol{\eta}}_{\sigma(t_{l+1}^-)} \leq \delta_1\ \ \forall\sigma \in \{1, 2\}$$

while $||r|| < \varpi$. Using this in (30) yields

$$\dot{V}(t) \leq -\kappa V(t) - (\iota - \kappa)V(t) + \delta + \varpi\delta_1. \tag{31}$$

From the analysis carried out for both the scenarios, it is concluded that $V(t) \geq \mathcal{B} \Rightarrow \dot{V}(t) \leq -\kappa V(t)$, where

$$\mathcal{B} \triangleq \frac{\delta + \varpi\delta_1}{(\iota - \kappa)}. \tag{32}$$

Unlike the conventional analysis for non-switched dynamical systems, further analysis is required to investigate the behavior of $V(t)$ once it enters the bound $\mathcal{B}$. Let T_1 be the instant when $V(t)$ enters the bound $\mathcal{B}$ and $\bar{N}(t)$ be the number of all switching intervals for $t \in [t_0\ \ t_0 + T_1)$.

Using (18), (28) and the definition of $N_\sigma(t_0, t)$, we have

$$\begin{aligned}V(t) &\leq \exp\big(-\kappa(t - t_{\bar{N}(t)-1})\big)V(t_{\bar{N}(t)-1})\\ &\leq \mu \exp\big(-\kappa(t - t_{\bar{N}(t)-1})\big)V(t^{-}_{\bar{N}(t)-1})\\ &\leq \exp\big(-\kappa(t - t_{\bar{N}(t)-1})\big)\\ &\quad \cdot \mu^2 \exp\big(-\kappa(t_{\bar{N}(t)-1} - t_{\bar{N}(t)-2})\big)V(t^{-}_{\bar{N}(t)-2})\\ &\quad \vdots\\ &\leq \mu \exp\big(-\kappa(t - t_{\bar{N}(t)-1})\big)\mu \exp\big(-\kappa(t_{\bar{N}(t)-1} - t_{\bar{N}(t)-2})\big)\\ &\quad \cdots \mu \exp(-\kappa(t_1 - t_0))V(t_0)\\ &= c(\exp(-\kappa + (\ln \mu/\vartheta)))V(t_0), \end{aligned} \tag{33}$$

where $c \triangleq \exp(N_0 \ln \mu)$ is a scalar. Invoking the average dwell time condition $\vartheta > \ln \mu/\kappa$ in (33), it can be derived that $V(t) < cV(t_0)$ for $t \in [t_0 \quad t_0 + T_1)$. As $V(t_0 + T_1) < \mathcal{B}$, we get $V(t_{\bar{N}(t)+1}) < \mu\mathcal{B}$ (from (18)), where $t_{\bar{N}(t)+1}$ denotes the subsequent switching instant following $t_0 + T_1$. Now following the standard recursive analysis as in [40], [45], and [46], one can prove that $V(t) < c\mu\mathcal{B}$ for $t \in [t_0 + T_1 \quad \infty)$. Hence, the control law and ADT switching law (14) ensure $V(t)$ cannot go beyond $c\mu\mathcal{B}$ in any time interval once it enters the bound $[0, \mathcal{B}]$, implying global uniform ultimate boundedness for the closed-loop dynamics. Moreover,

$$V(t) \leq \max\{cV(t_0), c\mu\mathcal{B}\}, \quad \forall t \geq t_0. \tag{34}$$

Using (34) and the fact $V(t) \geq (1/2)||e(t)||^2$,

$$||e||^2 \leq 2 \max\{cV(t_0), c\mu\mathcal{B}\}, \quad \forall t \geq t_0., \tag{35}$$

leading to an ultimate bound b on tracking error e as

$$b = \sqrt{\frac{2\bar{\varrho}_M^{(N_0+1)}(\delta + \varpi\delta_1)}{\underline{\varrho}_m^{(N_0+1)}(\iota - \kappa)}}. \tag{36}$$

Remark 1: It is important to note that the term '$-(1/2)\sum_{i=0}^{2} \hat{\eta}^2_{i\overline{\sigma}}$' contributed by the adaptive law of $\gamma_{\overline{\sigma}}$ in (25) cancels the corresponding term stemming from (27) to arrive (28) and (30), and achieve closed-loop stability.

References

[1] D. Rus and M. T. Tolley, "Design, fabrication and control of soft robots," *Nature*, vol. 521, no. 7553, pp. 467–475, May 2015.

[2] C. Majidi, "Soft robotics: A perspective—Current trends and prospects for the future," *Soft Robot.*, vol. 1, no. 1, pp. 5–11, Mar. 2014.

[3] R. Thakker, A. Kamat, S. Bharambe, S. Chiddarwar, and K. M. Bhurchandi, "ReBiS—Reconfigurable bipedal snake robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2014, pp. 309–314.

[4] Y. Shi, M. R. Elara, A. V. Le, V. Prabakaran, and K. L. Wood, "Path tracking control of self-reconfigurable robot hTetro with four differential drive units," *IEEE Robot. Autom. Lett.*, vol. 5, no. 3, pp. 3998–4005, Jul. 2020.

[5] M. Yim, "Modular self-reconfigurable robot systems: Challenges and opportunities for the future," *IEEE Robot. Autom. Mag.*, vol. 10, pp. 2–11, 2007.

[6] J. Seo, J. Paik, and M. Yim, "Modular reconfigurable robotics," *Annu. Rev. Control, Robot., Auto. Syst.*, vol. 2, pp. 63–88, May 2019.

[7] T. T. Tun, L. Huang, R. E. Mohan, and S. G. H. Matthew, "Four-wheel steering and driving mechanism for a reconfigurable floor cleaning robot," *Autom. Construct.*, vol. 106, Oct. 2019, Art. no. 102796.

[8] R. Parween, M. V. Heredia, M. M. Rayguru, R. E. Abdulkader, and M. R. Elara, "Autonomous self-reconfigurable floor cleaning robot," *IEEE Access*, vol. 8, pp. 114433–114442, 2020.

[9] N. Tan, A. A. Hayat, M. R. Elara, and K. L. Wood, "A framework for taxonomy and evaluation of self-reconfigurable robotic systems," *IEEE Access*, vol. 8, pp. 13969–13986, 2020.

[10] G. Gungor, B. Fidan, and W. W. Melek, "Decentralized model reference adaptive control design for modular and reconfigurable robots," in *Advances in Motion Sensing and Control for Robotic Applications: Selected Papers From the Symposium on Mechatronics, Robotics, and Control (SMRC'18)-CSME International Congress 2018, May 27-30, 2018 Toronto, Canada*. Springer, 2019, pp. 109–125.

[11] M. M. Rayguru, M. R. Elara, A. A. Hayat, B. Ramalingam, and S. Roy, "Modeling and control of PANTHERA self-reconfigurable pavement sweeping robot under actuator constraints," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2021, pp. 2742–2748.

[12] G. Liu, S. Abdul, and A. A. Goldenberg, "Distributed control of modular and reconfigurable robot with torque sensing," *Robotica*, vol. 26, no. 1, pp. 75–84, Jan. 2008.

[13] G. Liu, Y. Liu, and A. A. Goldenberg, "Design, analysis, and control of a spring-assisted modular and reconfigurable robot," *IEEE/ASME Trans. Mechatronics*, vol. 16, no. 4, pp. 695–706, Aug. 2011.

[14] T. Jianguo, L. Xiong, Y. Fei, and D. Zongquan, "A wheel-arm reconfigurable mobile robot design and its reconfigurable configuration," in *Proc. ASME/IFToMM Int. Conf. Reconfigurable Mech. Robots*, Jun. 2009, pp. 550–557.

[15] M. Tarokh, "A unified kinematics modeling, optimization and control of universal robots: From serial and parallel manipulators to walking, rolling and hybrid robots," *Auto. Robots*, vol. 44, pp. 1233–1248, Jul. 2020.

[16] A. Alamdari and V. Krovi, "Active reconfiguration for performance enhancement in articulated wheeled vehicles," in *Proc. Dyn. Syst. Control Conf.*, vol. 46193. New York, NY, USA: American Society of Mechanical Engineers, 2014, Art. no. V002T27A004.

[17] A. Alamdari, X. Zhou, and V. N. Krovi, "Kinematic modeling, analysis and control of highly reconfigurable articulated wheeled vehicles," in *Proc. Int. Design Eng. Tech. Conf. Comput. Inf. Eng. Conf.*, vol. 55935. New York, NY, USA: American Society of Mechanical Engineers, 2013, Art. no. V06AT07A070.

[18] Y. Zhang, Y. Koga, and D. Balkcom, "Interlocking block assembly with robots," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 3, pp. 902–916, Jul. 2021.

[19] B. Huang, Y. Yang, Y.-Y. Tsai, and G.-Z. Yang, "A reconfigurable multirobot cooperation workcell for personalized manufacturing," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 3, pp. 2581–2590, Jul. 2022.

[20] J. Paulos et al., "Automated self-assembly of large maritime structures by a team of robotic boats," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 3, pp. 958–968, Jul. 2015.

[21] J. Ye, S. Roy, M. Godjevac, and S. Baldi, "A switching control perspective on the offshore construction scenario of heavy-lift vessels," *IEEE Trans. Control Syst. Technol.*, vol. 29, no. 1, pp. 470–477, Jan. 2021.

[22] A. V. Le, R. Parween, P. T. Kyaw, R. E. Mohan, T. H. Q. Minh, and C. S. C. S. Borusu, "Reinforcement learning-based energy-aware area coverage for reconfigurable hRombo tiling robot," *IEEE Access*, vol. 8, pp. 209750–209761, 2020.

[23] P. T. Kyaw et al., "Energy-efficient path planning of reconfigurable robots in complex environments," *IEEE Trans. Robot.*, vol. 38, no. 4, pp. 2481–2494, Aug. 2022.

[24] S. Roy and S. Baldi, "On reduced-complexity robust adaptive control of switched Euler–Lagrange systems," *Nonlinear Anal., Hybrid Syst.*, vol. 34, pp. 226–237, Nov. 2019.

[25] D. Liberzon, *Switching in Systems and Control*, vol. 190. Boston, MA, USA: Birkhauser, 2003.

[26] T. C. Lee and Z. P. Jiang, "Uniform asymptotic stability of nonlinear switched systems with an application to mobile robots," *IEEE Trans. Autom. Control*, vol. 53, no. 5, pp. 1235–1252, Jun. 2008.

[27] Z. Li, J. Deng, R. Lu, Y. Xu, J. Bai, and C.-Y. Su, "Trajectory-tracking control of mobile robot systems incorporating neural-dynamic optimized model predictive approach," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 6, pp. 740–749, Jun. 2016.

[28] R. D. Yadav, V. N. Sankaranarayanan, and S. Roy, "Adaptive sliding mode control for autonomous vehicle platoon under unknown friction forces," in *Proc. 20th Int. Conf. Adv. Robot. (ICAR)*, Dec. 2021, pp. 879–884.

[29] M. M. Rayguru, S. Roy, and I. N. Kar, "Time-scale redesign-based saturated controller synthesis for a class of MIMO nonlinear systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 8, pp. 4681–4692, Aug. 2021.

[30] M. M. Rayguru, R. E. Mohan, R. Parween, L. Yi, A. V. Le, and S. Roy, "An output feedback based robust saturated controller design for pavement sweeping self-reconfigurable robot," *IEEE/ASME Trans. Mechatronics*, vol. 26, no. 3, pp. 1236–1247, Jun. 2021.

[31] P. Cheng et al., "Asynchronous fault detection observer for 2-D Markov jump systems," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 13623–13634, Dec. 2022.

[32] N. Nedić, D. Pršić, C. Fragassa, V. Stojanović, and A. Pavlovic, "Simulation of hydraulic check valve for forestry equipment," *Int. J. Heavy Vehicle Syst.*, vol. 24, no. 3, pp. 260–276, 2017.

[33] V. Djordjevic, V. Stojanovic, H. Tao, X. Song, S. He, and W. Gao, "Data-driven control of hydraulic servo actuator based on adaptive dynamic programming," *Discrete Continuous Dyn. Syst. S*, vol. 15, no. 7, p. 1633, 2022.

[34] S. Tong, Y. Li, and S. Sui, "Adaptive fuzzy output feedback control for switched nonstrict-feedback nonlinear systems with input nonlinearities," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 6, pp. 1426–1440, Dec. 2016.

[35] Y. M. S. S. Li and S. C. Tong, "Adaptive fuzzy control design for stochastic nonlinear switched systems with arbitrary switchings and unmodeled dynamics," *IEEE Trans. Cybern.*, vol. 47, no. 2, pp. 403–414, Jun. 2017.

[36] K. P. Cheng, R. E. Mohan, N. H. K. Nhan, and A. V. Le, "Graph theory-based approach to accomplish complete coverage path planning tasks for reconfigurable robots," *IEEE Access*, vol. 7, pp. 94642–94657, 2019.

[37] A. A. Hayat, R. Parween, M. R. Elara, K. Parsuraman, and P. S. Kandasamy, "PANTHERA: Design of a reconfigurable pavement sweeping robot," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 7346–7352.

[38] P. Coelho and U. Nunes, "Path-following control of mobile robots in presence of uncertainties," *IEEE Trans. Robot.*, vol. 21, no. 2, pp. 252–261, Apr. 2005.

[39] S. Roy, S. Nandy, R. Ray, and S. N. Shome, "Robust path tracking control of nonholonomic wheeled mobile robot: Experimental validation," *Int. J. Control Autom. Syst.*, vol. 13, no. 4, pp. 897–905, 2015.

[40] J. P. Hespanha and A. S. Morse, "Stability of switched systems with average dwell-time," in *Proc. 38th IEEE Conf. Decis. Control*, vol. 3, Dec. 1999, pp. 2655–2660.

[41] J.-J. E. Slotine et al., *Applied Nonlinear Control*, vol. 199, no. 1. Englewood Cliffs, NJ, USA: Prentice-Hall, 1991.

[42] C. Dai, T. Guo, J. Yang, and S. Li, "A disturbance observer-based current-constrained controller for speed regulation of PMSM systems subject to unmatched disturbances," *IEEE Trans. Ind. Electron.*, vol. 68, no. 1, pp. 767–775, Jan. 2021.

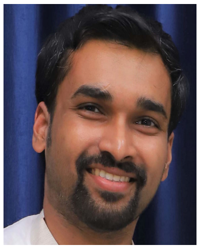
[43] F. Plestan, Y. Shtessel, V. Brégeault, and A. Poznyak, "New methodologies for adaptive sliding mode control," *Int. J. Control*, vol. 83, no. 9, pp. 1907–1919, 2010.

[44] S. Roy, S. Baldi, and L. M. Fridman, "On adaptive sliding mode control without *a priori* bounded uncertainty," *Automatica*, vol. 111, Jan. 2020, Art. no. 108650.

[45] T. Tao, S. Roy, and S. Baldi, "The issue of transients in leakage-based model reference adaptive control of switched linear systems," *Nonlinear Anal., Hybrid Syst.*, vol. 36, May 2020, Art. no. 100885.

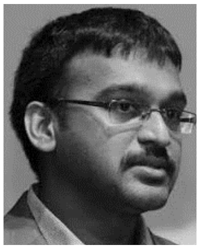
[46] S. Roy, E. B. Kosmatopoulos, and S. Baldi, "On vanishing gains in robust adaptation of switched systems: A new leakage-based result for a class of Euler–Lagrange dynamics," *Syst. Control Lett.*, vol. 144, Oct. 2020, Art. no. 104773.

Madan Mohan Rayguru received the B.Tech. degree from BPUT Odisha, India, the master's degree in control and automation from NIT Rourkela, and the Ph.D. degree in control systems from the Indian Institute of Technology Delhi, Delhi, India. He was working as a Research Fellow with the Engineering Product Development Pillar, Singapore University of Technology and Design (SUTD). He is currently an Assistant Professor with the Department of Electrical Engineering, Delhi Technological University, India. His research interests include robotics, convergent systems, and saturated controller design.



Spandan Roy received the B.Tech. degree in electronics and communication engineering from Techno India, West Bengal University of Technology, India, in 2011, the M.Tech. degree in mechatronics from the Academy of Scientific and Innovative Research, India, in 2013, and the Ph.D. degree in control and automation from the Indian Institute of Technology Delhi, India, in 2018. He is currently an Assistant Professor with the Robotics Research Center, International Institute of Information Technology Hyderabad, Hyderabad, India. Previously, he was a Post-Doctoral Researcher with the Delft Center for System and Control, Delft University of Technology, The Netherlands. His research interests include adaptive-robust control and switched systems and its applications in Euler–Lagrange systems. He is also a Subject Editor of *International Journal of Adaptive Control and Signal Processing*.

Lim Yi received the B.Eng. and M.Eng. degrees from the Engineering Product Development (EPD) Pillar, Singapore University of Technology and Design (SUTD), in 2017 and 2020, respectively. He is currently pursuing the Ph.D. degree in robotics path planning. He is also working with the Robotics and Automation Research (ROAR) Laboratory, SUTD. His current research interests include robotics, robotics control, robotics perception, robotics vision, and robotics navigation. He was awarded the Singapore University of Technology and Design President's Graduate Fellowship (Computing and Information Science Disciplines).



Mohan Rajesh Elara received the M.Sc. degree in consumer electronics from Nanyang Technological University, Singapore, and the Ph.D. degree in electrical and electronics engineering. He is currently an Associate Professor with the Engineering Product Development Pillar, Singapore University of Technology and Design (SUTD). Before joining SUTD, he was a Lecturer with the School of Electrical and Electronics Engineering, Singapore Polytechnic. His research interests are in robotics with an emphasis on self-reconfigurable platforms and research problems related to robot ergonomics and autonomous systems. He was a recipient of SG Mark Design Award in 2016, 2017, and 2018, respectively, Design Award in 2018, ASEE Best of Design in Engineering Award in 2012, and Tan Kah Kee Young Inventors Award in 2010. He is also a Visiting Faculty Member of the International Design Institute, Zhejiang University, China.

Simone Baldi (Senior Member, IEEE) received the B.Sc. degree in electrical engineering and the M.Sc. and Ph.D. degrees in automatic control engineering from the University of Florence, Italy, in 2005, 2007, and 2011, respectively. He is currently a Professor with the School of Mathematics, Southeast University, with guest position with the Delft Center for Systems and Control, TU Delft, where he was an Assistant Professor. His research interests are adaptive and learning systems with applications in unmanned vehicles and smart energy systems. He was awarded an Outstanding Reviewer of *Applied Energy* (2016) and *Automatica* (2017). He is also a Subject Editor of *International Journal of Adaptive Control and Signal Processing*, an Associate Editor of IEEE CONTROL SYSTEMS LETTERS, and a Technical Editor of IEEE/ASME TRANSACTIONS ON MECHATRONICS.

2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT) | 978-1-6654-9414-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICITIIT57246.2023.10068658

Machine Learning Techniques for Autism Spectrum Disorder: current trends and future directions

Kainat Khan
Big Data Analytics and Web Intelligence Laboratory, Department of Computer Science & Engineering
Delhi Technological University
New Delhi, India
khankainat388@gmail.com

Rahul Katarya
Big Data Analytics and Web Intelligence Laboratory, Department of Computer Science & Engineering
Delhi Technological University
New Delhi, India
rahuldtu@gmail.com

Abstract— **ASD or autism spectrum disorder is a critical neuro-developmental disorder that hinders an individual's capability of social communication and interaction. This disorder has acquired considerable attention and importance due to its ubiquity among individuals covering all the countries worldwide. Individuals with ASD struggles in daily life activities. Detection of autism with the help of medical tests is a tedious and very costly task. However, detection and care of ASD still remains unfamiliar due to inadequate awareness, knowledge among the society, limited number of diagnostic devices and limited verbal therapy services for ASD patients. This paper investigates and displays reviews of various machine learning approaches on extracting useful data associated with distinctive characteristics of ASD such as brain functioning, hyperactivitperactivity, language disability, etc. Current researches reveal that analysis of biological traits by employing machine learning techniques have helped in the progress of early detection of ASD. ABIDE dataset is very much explored for the research in ASD. Additionally, numerous studies for the advancement of tools are still in progression. The presented research work can remarkably aid future studies on machine learning for ASD.**

Keywords— ASD, Artificial intelligence, Machine learning, Supervised learning

I. Introduction

ASD indicates a variety of characteristics that includes a particular degree of behavioral, communicational and linguistic disabilities. ASD is a diverse disorder in terms of severity, level of risk and response towards a treatment. ASD initiates in the early years of childhood and it stretches into adulthood. Studies prove that detection and treatment of ASD in early years is beneficial as it reduces the expense of treatment and time[1]. Another interesting detail that urges the research is the co-occurring issues faced by the people suffering from ASD. Anxiety, sensory issues, and depression are a few of the major lifetime problems that an ASD handles[2]. Brain functioning, hyperactivity, and language disability are some of the symptoms/characteristics that differentiate a person with ASD and without ASD.

At present, pace of autism all over the world is expanding rapidly. Reports evaluates that 1 in every 59 individuals is detected with ASD in USA. Every year millions are spent for the cure of ASD. According to the data, the rate of autism among boys is higher than that of girls, exposing the actuality that boys were four-times more likely to be detected with ASD than the girls[3]. Advanced technologies and smart thinking are the pillars of every methodology for the pattern discovery in the data. This includes a procedural to approach breakdown the issue, discover the pattern, eradicate the irrelevant data and formulate the solution. This shifts our concern towards the benefits of deep learning, machine learning, Artificial intelligence that consists of methods and algorithms to make a model learn automatically from the existing data[4]. Machine learning is classified as unsupervised, supervised, reinforcement learning and semi-supervised learning. The motive of supervised method is to make sure that the model can predict on unseen data based on the labeled data used for training. In unsupervised learning model was exposed to unlabeled data with least supervision[5]. Association and clustering are two main kinds of unsupervised learning.

- The objective of this research survey is to shed light on machine learning state-of-the art techniques for autism spectrum disorder.
- Discussed ASD-ML based works done between 2017 to 2022 to encourage further works in the mentioned domain.

The organization of this work is as follows. Section 2 reveals the general criteria for conducting the research. Section 3 discusses the exployed machine learning approaches for ASD. At last, section 4 concludes the overall paper.

II. State-of-the-art

IEEExplore, Google scholar and Endnote were explored to fetch articles related to the following search terms: Autism spectrum disorder, machine learning, computational intelligence, AI in ASD, deep learning in ASD. The abstracts, methodology, and results portion of the studies were reviewed thoroughly, and by using a systematized rubric, a study was incorporated if it followed the criteria mentioned below.

1. Comprised of group/people with ASD.
2. Machine learning algorithms were employed as the fundamental approach of analysis.
3. Research studies must be in English.

We hope that conclusions drawn from this study imply the current state of machine learning techniques in ASD.

The literature survey generated 40 research publication from (1) Ieeexplore, (2) Google Scholar, (3) Endnote. After the abstracts and methodologies were reviewed, 30 research papers were incorporated. Of the incorporated studies, 18

978-1-6654-9414-4/23/$31.00 ©2023 IEEE

included the utilization of ML to detect, predict and classify ASD sufferers. Figure 1 represents the overall research flow chart to analyse eligible research studies for the work presented.

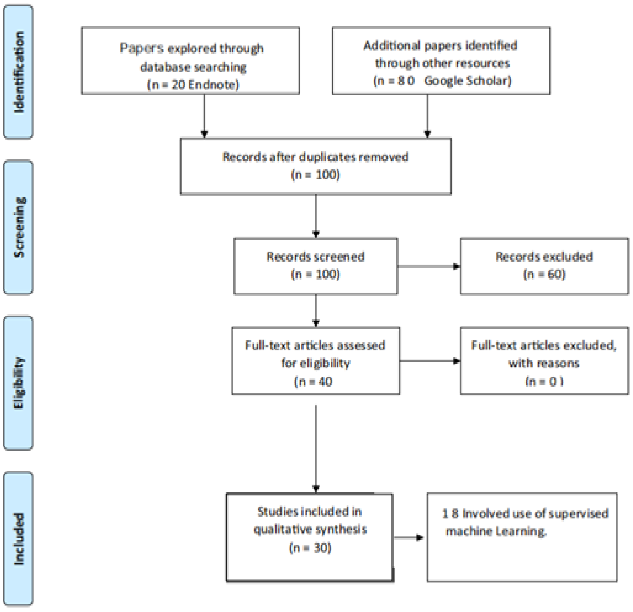

Identification
Screening
Eligibility
Included
Papers explored through database searching (n = 20 Endnote)
Additional papers identified through other resources (n = 8 0 Google Scholar)
Records after duplicates removed (n = 100)
Records screened (n = 100)
Records excluded (n = 60)
Full-text articles assessed for eligibility (n = 40
Full-text articles excluded, with reasons (n = 0)
Studies included in qualitative synthesis (n = 30)
1 8 Involved use of supervised machine Learning.

Fig.1. PRISMA Flowchart for selection of papers on machine learning for ASD

III. Machine Learning Approaches in ASD Research

In this work, supervised learning approaches were reviewed to analyze patterns for a data-set which further leads to improved and accurate prediction for a classification problem[6]. Studies inspected in this paper employed various techniques of machine learning, with outperforming results among other approaches[7].

Of the 30 papers incorporated in this work, with many papers reporting higher accuracy with Logistic Regression, SVM and Random Forest. The brief of the supervised approaches comprised in this work is mentioned in Table 2.

Literature survey of the ASD studies recommends categorizing the state-of-art into three sections:

- Data centered approach
- Algorithmic centered approach
- Conventional ML framework adopted in existing studies

A. Data Centered Approach

The ASD state-of-the-art is categorized into textual data and image data. The most conventional dataset used in the ASD research is the ABIDE dataset consisting of images of both ASD and Typical Control (TC) individuals. Out of 30 paper, 8 have implemented dataset from ABIDE. This dataset includes phenotypic data and RS-fMRI from various international websites, generating diversified sample data of large size[8]. This shows potentiality for analyzing patterns in ASD classification. Due to experimental differences, diversity in data and statistical noise classification becomes complex[9]. This recommends the necessity of a reliable technique that can efficiently deal with huge dataset and treat the other risk factors as well.

The second most acknowledged dataset is of UC Irvine machine learning repository having datasets related to children, adults and adolescents. The mentioned dataset was produced by utilizing the results of AQ-10 test and proved to be beneficial in examining performances of classifiers.

The third kind of dataset extracted through questionnaires and medical centers. This dataset is mostly used to analyse classifiers performance in order to distinguish ASD and non-ASD individuals. Kaggle is also one of the platforms for obtaining ASD dataset.

B. Algorithmic Centered Approach

This approach has engrossed on two objectives, i.e., to find biomarkers and to discover efficient classifier to differentiate between TC patients and ASD patients[10]. In order to find biomarkers, ABIDE dataset seems to be the best source for information extraction and uncover markers. Conventional approaches are not capable enough to mine and uncover data related to biomarkers. The common structural framework employed in this part is depicted in figure 2. The survey described in this study shows the extent for in-depth research with advanced computational approaches to detect the biomarkers

C. Conventional Framework Adopted in Existing ASD Studies

The diversified framework/architecture adopted for autism detection via machine learning is presented in figure 2. The experimental works and review studies considered for literature survey identifies various steps and machine learning startegies for ASD. The basic flow for the detection of ASD via ML includes collection of data, pre-processing of data (handling imbalanced data), feature selection process, model training & classification, and model testing & evaluation. The following sub-sections describes different strategies used at each step while classifiying a person as healthy or autistic.

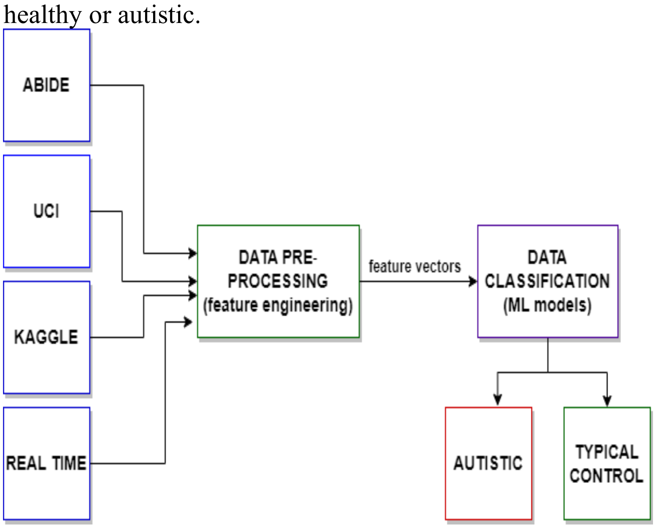

ABIDE
UCI
KAGGLE
REAL TIME
DATA PRE-PROCESSING (feature engineering)
feature vectors
DATA CLASSIFICATION (ML models)
AUTISTIC
TYPICAL CONTROL

Fig.2. Conventional framework for existing approaches in ASD

- ***Dataset collection and utilization for ASD***

Over the years, research in the domain of ASD has increased significantly. Brain magnetic imaging data (brain MRI), electroencephalography (EEG signals), questionnaire are the types of dataset that has been mostly utilized in autism studies. ABIDE dataset is one such dataset that is widely used for autism detection and prediction due to the functional, structural phenotypic and MRI data [11]

TABLE I. DETAILED DESCRIPTION OF ASD DATASETS

Dataset Name	Instances	Attributes	ASD : Normal	Dataset Type	Male : Female
Autism toddler dataset	1054	18	735 : 319	Questionnaire	735 : 319
Autism adult dataset	1118	23	358 : 760	Questionnaire	596 : 522
ABIDE I	-	-	539 : 573	R-fMRI images	-
ABIDE 11	-	-	521 : 593	R-fMRI images	-

Table 1 gives the detailed description of ASD datasets available on kaggle, uci repository, etc. Figure 3 portrays the utilization of datasets for conducting research on machine learning techniques for ASD.

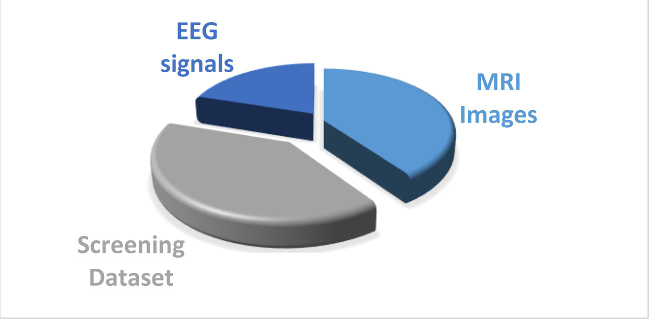


Fig.3. Utilization rate of existing datasets for ASD-ML based works

- **Data pre-processing**

Pre-processing of the dataset is an integral section which is used to enhance/ improve the quality of the data. To make the model learn better, crucial features are fed. In ASD by with the help of feature engineering we select, alter & convert information into useful featues. İn ASD sufferes, brain biomarkers can be elucidated by adopting convolutional neural network model fetching strategy.

- **Classification**

For ASD classification, various automatic classification & detection models has been adopted till now to accurately classify brain MRI images. As per the literature survey, machine learning classifiers like, support vector machine, k-nearest neighbor, convolutional neural network, random forest, logistic regression, etc are mostly utilized.

- **Evaluation**

The effectiveness of various classification models & feature extraction techniques was discovered through the examination of numerous works on the detection of ASD. Considering the purpose & the techniques that wereemployed, there appears to be a sigificant disparity in the outcomes. Table 2 summarizes the studies published on ASD detection and classification

TABLE II. OVERVIEW OF RUBRIC PREPARED ON MACHINE LEARNING TECHNIQUES FOR AUTISM SPECTRUM DISORDER

Authors	Year	Aim	Dataset	Methodology	Results	Limitations
Malathi et. al [12]	2022	To implement Adaptive whale optimization-based SVM for ASD prediction	ASD screening dataset	AWO-SVM was utilized for ASD prediction	Achieved an accuracy of 90.162%	Accuracy can be improved
Bala et al. [13]	2022	Development of ML model to identify ASD at various age levels in order to detect ASD accurately	ASD dataset of children, toddlers, adults and adolescents	Feature subset was generated via feature selection performed on ASD data and various ML classifiers were applied to develop the model	SVM achieved best results with feature selection strategies	Model was not properly trained for handling high dimensional datasets
Karim et al. [14]	2022	Proposed a novel fuzzy-semi-supervised learning technique for ASD meltdown prediction	ASD dataset of children	Worked by implementing divide & conquer method on 10 different classifiers	Fuzzy-semi-supervised classifier performed the best among other classifiers	Smaller dataset, could not diagnose ASD in early stage
Ajmi et al. [15]	2022	Exploring existing trends in ASD detection strategies	Reviewed 18 studies on ASD	Analysing ASD detection techniques on the basis of classification strategy, performance metrics	Advised to explore graph neural network method for ASD	Limited studies were explored
Pang et al. [16]	2022	Proposed an optimal cascaded classifier for diagnosing ASD	fMRI images	Extracted brain features from fMRI images which were enhanced via empirical kernel	Proposed method achieved	Accuracy can be improved, deep learning techniques like

		and parkinson disease		mapping. Enhanced features were fed to the model to make an optimized framework	decent accuracy	graph convolutional network should be adopted
Hosseini et. al [17]	2022	To deploy a cost-effective and efficient model for analyzing ASD	Image dataset from Kaggle	Deep learning methods were employed for facial analysis of ASD images	Achieved a test accuracy of 94.64%	Data pre-processing was not performed
Sofia et. al [18]	2022	To examine early detection of ASD using eye-tracking and DL	10 research works were explored	Eye-tracking technique using deep learning were explored	Recommends that ASD can be diagnosed using images	Few studies were examined
Zhang et al. [19]	2022	Developed a framework for ASD classification	fMRI images	Worked on a novel feature selection technique i.e., DSDC & built deep learning modal via two strategies of pre-training & MLP fine tuning	Achieved accuracy of 78.1%	Explainability of the model were not considered
Azbari et al. [20]	2022	Experimental study on facial emotion expression on children having autistic traits	Worked on a primary dataset	Performed multi model experiments on facial emotion classification and processing eye gaze	Results were true to the hypothesis made	Data collected in their work included very few samples
Rahaman et al. [21]	2022	To enhance classification of the disorder and identification of underlying mechanism behind mental disorders	fMRI, sMRI and genome data	Developed neural network modal for learning features & deploying adaptive control unit which included autoencoders, multilayer network & LSTM unit	Model gained accuracy of 92%	Accuracy can be improved
Sriram et. al [5]	2021	To predict ASD using ML	Hybrid autism screening dataset	Decision tree, Naïve Bayes, RF, KNN, LR were used to build a model	Accuracy was improved in comparison to other studies.	Small sample size
Hassan et.al [22]	2021	Comparative study between three metaheuristic approaches	High-dimensional microarray data	Used three-phase hybrid method using PPMCC and meta-heuristic techniques	BPSO outperformed genetic algorithm	Worked on a smaller dataset
Mehdi hosseinzadeh et. al [23]	2021	To present a systematic review of IoT and ML approaches for diagnosing ASD	28 Research studies were examined	IoT and ML-based techniques were explored	Advanced devices can be developed using IoT and ML	Less number of research studies were explored
Akter et. al [24]	2021	To present efficient ML-based model to diagnose ASD more accurately	AQ-10 dataset was used	ANN, RN, DT, Gradient boost, KNN, LR, NB, SVM, MLP, Xgboost were employed	Logistic regression outperforms among all the other classifiers	No screening tools diagnosed ASD correctly
Wu et. al [25]	2020	To detect ASD from videos using ML	2000 videos dataset	Followed two-stage process having image model and facial feature model	Achieved an accuracy of 82% in diagnosing ASD	Accuracy can be improved
Lee et. al [26]	2020	To develop pre-trained auto-encoder model for	Audio data from SNUBH	Performed an experiment to find ways	Deep learning-based feature extraction can	Research was focused only on infants

		feature extraction & optimization		to do feature extraction using auto-encoder	be useful in improving the accuracy	
Elbattah et. al [27]	2020	To learn about sequential patterns in saccadic eye movement to detect ASD	Data of children in the age group 3-12 years	Used NLP based transformation for processing raw data to fetch essential information	Accuracy of classification was 0.84 (ROC-AUC)	Small data was small and caters people from a certain age group only
Shahid, Singh [28]	2019	To address issues and current trend for medical diagnosis and prognosis	75 Research works were explored	Analysis of computational intelligence approaches	Hybrid approaches performs better than the CI techniques	Studies only based on CI & hybrid techniques were explored
Abitha, Vennila [29]	2019	To propose a swarm method based on symmetrical uncertainty and PSO	ASD children dataset	Feature selection and optimized techniques are used to propose an approach	SSU-FS outperforms than the existing algorithms	The study focuses on children dataset only
Hyde et.al [6]	2019	To explain the supervised ML trends in ASD	45 studies were analysed	Analysis of supervised ML methods	Supervised ML methods possesses great value in ASD	Limited research studies were explored
Omar et. al [7]	2019	To deploy efficient model for ASD prediction using ML techniques	AQ-10 dataset was used	Implemented Random Forest-CART and Random Forest-ID3	Developed model shows improved results for all accuracy measures	Limited to the findings of related work
Puli, Kushki [2]	2019	Automatically detect anxiety in ASD patients	ASD dataset of children and youth	Multiple model Kalman filter was proposed	Achieved an arousal detection accuracy of 93%	Sample dataset was small
Eman et.al [30]	2019	To review ML classifiers for ASD	16 studies were examined	Survey of ML classifiers	ML methods outperforms in most cases	Few research works were surveyed
Yang et.al [1]	2018	Multi-modal picture book recommendation for children with ASD	Picture book image dataset	CA and MCA algorithms were used	A novel recommender system was proposed	Only focuses on technical aspects
Sadock et.al [4]	2018	To present deep learning method for stereotypical motor movements (SMM) recognition in ASD	SMM time-series data	Implemented SMM detection framework using transfer learning, SVM	Time series and frequency domain techniques can be useful in many cases	SMM data is rare to find & doesn't have labelled instances in it

Lessons Learnt from the literature survey performed:

A. Recent researches proves that machine learning techniques like SVM, RF and logistic regression outperforms in ASD research.

B. Need of ample amount of medical data in the domain of autism spectrum disorder in order to generalize results as data available on ASD is limited.

C. Need of techniques to identify autism spectrum disorder in quick manner as early detection of ASD helps the individual to cure and limited research is done to detect ASD at early stages.

Ç. Need of researches that caters all age group people with a generalized technique. Most of the existing works are based on models made for a certain age group (either on toddlers or child or adults). Model should be able to capture features from the data of every age group.

D. Due to data complexity, the computation power and time complexity of existing techniques seems to be very high which needs to be considered.

E. Interpretability and identification of ASD biomarkers is need to addressed.

F. Precision, Recall and Accuracy are most used performance metrics on ASD works.

IV. Conclusion and Future Work

The motive of this literature survey was to shed light on the current trend to detect ASD using ML. A total of 30 articles that employed ML as a base were incorporated. The findings portray that the fundamental approaches and the advanced approaches need to work in such a manner that deploys a novel approach to handle the current limitations mentioned in the literature survey table. The need of reliable technique to collect data is still a challenge. This demands efficient algorithms for the detection of biomarkers. It is extremely important to build a prediction model that will be generalize to every age group patient suffering from ASD.

References

[1] Z. Zheng *et al.*, “Design of an Autonomous Social Orienting Training System (ASOTS) for Young Children with Autism,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 668–678, 2017, doi: 10.1109/TNSRE.2016.2598727.

[2] A. Puli and A. Kushki, “Toward Automatic Anxiety Detection in Autism: A Real-Time Algorithm for Detecting Physiological Arousal in the Presence of Motion,” *IEEE Trans. Biomed. Eng.*, vol. 67, no. 3, pp. 646–657, 2020, doi: 10.1109/TBME.2019.2919273.

[3] M. D. Samad, N. DIawara, J. L. Bobzien, J. W. Harrington, M. A. Witherow, and K. M. Iftekharuddin, “A Feasibility Study of Autism Behavioral Markers in Spontaneous Facial, Visual, and Hand Movement Response Data,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 353–361, 2018, doi: 10.1109/TNSRE.2017.2768482.

[4] T. Gadi and E. H. Essoufi, “A Novel Deep Learning Approach for Recognizing Stereotypical Motor Movements within and,” *Comput. Intell. Neurosci.*, vol. 2018, 2018.

[5] G. A. G. Y. M. Y. S. Dhanyatha Sriram, “IRJET- Prediction of Autism Spectrum Disorder based on Machine Learning Approach,” *Irjet*, vol. 8, no. 7, 2021.

[6] C. M. Parlett-Pelleriti, E. Stevens, D. Dixon, and E. J. Linstead, “Applications of Unsupervised Machine Learning in Autism Spectrum Disorder Research: a Review,” *Rev. J. Autism Dev. Disord.*, 2022, doi: 10.1007/s40489-021-00299-y.

[7] K. S. Oma, P. Mondal, N. S. Khan, M. R. K. Rizvi, and M. N. Islam, “A Machine Learning Approach to Predict Autism Spectrum Disorder,” *2nd Int. Conf. Electr. Comput. Commun. Eng. ECCE 2019*, no. February, 2019, doi: 10.1109/ECACE.2019.8679454.

[8] A. Z. Guo, “Automated Autism Detection based on Characterizing Observable Patterns from Photos,” *IEEE Trans. Affect. Comput.*, vol. X, no. X, pp. 1–6, 2020, doi: 10.1109/TAFFC.2020.3035088.

[9] S. Raj and S. Masood, “Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques,” *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 994–1004, 2020, doi: 10.1016/j.procs.2020.03.399.

[10] S. Gracia, “ScienceDirect ScienceDirect Algorithmic Algorithmic Approaches Approaches to to Classify Classify Autism Autism Spectrum Spectrum Disorders : Disorders : Perspective A Research Perspective,” *Procedia Comput. Sci.*, vol. 201, pp. 470–477, 2022, doi: 10.1016/j.procs.2022.03.061.

[11] N. Wang, D. Yao, L. Ma, and M. Liu, “Multi-site clustering and nested feature extraction for identifying autism spectrum disorder with resting-state fMRI,” *Med. Image Anal.*, vol. 75, p. 102279, 2022, doi: 10.1016/j.media.2021.102279.

[12] A. I. Technology and S. Malathi, “ADAPTIVE WHALE OPTIMIZATION BASED SUPPORT VECTOR MACHINE FOR PREDICTION OF AUTISM,” vol. 100, no. 9, pp. 2862–2870, 2022.

[13] M. Bala, M. H. Ali, M. S. Satu, K. F. Hasan, and M. A. Moni, “Efficient Machine Learning Models for Early Stage Detection of Autism Spectrum Disorder,” *Algorithms*, vol. 15, no. 5, pp. 1–22, 2022, doi: 10.3390/a15050166.

[14] S. Karim, N. Akter, and M. J. A. Patwary, “Predicting Autism Spectrum Disorder (ASD) meltdown using Fuzzy Semi-Supervised Learning with NNRW,” *2022 Int. Conf. Innov. Sci. Eng. Technol. ICISET 2022*, no. February, pp. 367–372, 2022, doi: 10.1109/ICISET54810.2022.9775860.

[15] N. S. Ajmi, D. A. George, M. B. Megha, and J. Mohan, “A Review of Machine Learning Techniques for Detecting Autism Spectrum Disorders,” *Int. Conf. Sustain. Comput. Data Commun. Syst. ICSCDS 2022 - Proc.*, pp. 148–155, 2022, doi: 10.1109/ICSCDS53736.2022.9760909.

[16] C. Pang *et al.*, “Improving model robustness via enhanced feature representation and sample distribution based on cascaded classifiers for computer-aided diagnosis of brain disease,” *Biomed. Signal Process. Control*, vol. 79, no. P1, p. 104047, 2023, doi: 10.1016/j.bspc.2022.104047.

[17] M. P. Hosseini, M. Beary, A. Hadsell, R. Messersmith, and H. Soltanian-Zadeh, “Deep Learning for Autism Diagnosis and Facial Analysis in Children,” *Front. Comput. Neurosci.*, vol. 15, no. January, pp. 1–7, 2022, doi: 10.3389/fncom.2021.789998.

[18] M. S. S and N. Mohanan, “A Review on Early Detection of Autism Spectrum Disorder using Eye Tracking and Deep Learning,” vol. 11, no. 04, pp. 302–304, 2022.

[19] F. Zhang, Y. Wei, J. Liu, Y. Wang, W. Xi, and Y. Pan, “Identification of Autism spectrum disorder based on a novel feature selection method and Variational Autoencoder,” *Comput. Biol. Med.*, vol. 148, 2022, doi: 10.1016/j.compbiomed.2022.105854.

[20] S. Bagherzadeh-Azbari *et al.*, “Multimodal Evidence of Atypical Processing of Eye Gaze and Facial Emotion in Children With Autistic Traits,” *Front. Hum. Neurosci.*, vol. 16, no. February, 2022, doi: 10.3389/fnhum.2022.733852.

[21] M. A. Rahaman *et al.*, “Deep multimodal predictome for studying mental disorders,” *Hum. Brain Mapp.*, no. July, pp. 1–14, 2022, doi: 10.1002/hbm.26077.

[22] S. S. Hameed, W. H. Hassan, L. A. Latiff, and F. F. Muhammadsharif, “A comparative study of nature-inspired metaheuristic algorithms using a three-phase hybrid approach for gene selection and classification in high-dimensional cancer datasets,” *Soft Comput.*, vol. 25, no. 13, pp. 8683–8701, 2021, doi: 10.1007/s00500-021-05726-0.

[23] M. Hosseinzadeh *et al.*, “A review on diagnostic autism spectrum disorder approaches based on the Internet of Things and Machine Learning,” *J. Supercomput.*, vol. 77, no. 3, pp. 2590–2608, 2021, doi: 10.1007/s11227-020-03357-0.

[24] A. N. Optimal *et al.*, “Article an Optimal Metaheuristic Based Feature Selection With Deep Learning Model for Autism Spectrum,” vol. 12, pp. 1–7, 2021.

[25] C. Wu *et al.*, “Machine learning based autism spectrum disorder detection from videos,” *2020 IEEE Int. Conf. E-Health Networking, Appl. Serv. Heal. 2020*, 2021, doi: 10.1109/HEALTHCOM49281.2021.9398924.

[26] H. Sewani and R. Kashef, “An autoencoder-based deep learning classifier for efficient diagnosis of autism,” *Children*, vol. 7, no. 10, 2020, doi: 10.3390/children7100182.

[27] M. Elbattah, J. L. Guerin, R. Carette, F. Cilia, and G. Dequen, “NLP-Based Approach to Detect Autism Spectrum Disorder in Saccadic Eye Movement,” *2020 IEEE Symp. Ser. Comput. Intell. SSCI 2020*, pp. 1581–1587, 2020, doi: 10.1109/SSCI47803.2020.9308238.

[28] A. H. Shahid and M. P. Singh, “Computational intelligence techniques for medical diagnosis and prognosis: Problems and current developments,” *Biocybern. Biomed. Eng.*, vol. 39, no. 3, pp. 638–672, 2019, doi: 10.1016/j.bbe.2019.05.010.

[29] R. Abitha and S. Mary Vennila, “A Swarm Based Symmetrical Uncertainty Feature Selection Method for Autism Spectrum Disorders,” *Proc. 3rd Int. Conf. Inven. Syst. Control. ICISC 2019*, no. Icisc, pp. 665–669, 2019, doi: 10.1109/ICISC44355.2019.9036454.

[30] D. Eman and A. W. R. Emanuel, “Machine learning classifiers for autism spectrum disorder: A review,” *2019 4th Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2019*, vol. 6, pp. 255–260, 2019, doi: 10.1109/ICITISEE48480.2019.9003807.

doi: 10.1002/hbm.26077.

Short Papers

Measuring Influence of Indices in DN Planning

Shubham Gupta, Vinod Kumar Yadav, *Senior Member, IEEE*, and Madhusudan Singh, *Senior Member, IEEE*

***Abstract*—The optimal distribution network (DN) setup in the modern framework baffles the DN planner, as various technoeconomic impacts need to be analyzed. The prevailing studies urge multiobjective problem formulation to quantify the wide range of technoeconomic performance indices; however, no adequate method is posited to measure the relative influence coefficient of these indices in DN planning. This article proposes a novel method based on the Shannon entropy formula to determine the relative influence level of indices in multiobjective optimally distributed generation unit(s) placement problems. Numerical results have been presented to validate the efficacy of the proposed approach over the previously published strategies.**

***Index Terms*—Distribution network (DN), DN planning, Shannon entropy (SE), weighted-sum multiobjective analysis.**

I. Introduction

The shifting from a vertically integrated structure to unbundling of power system brought fringe benefits but increased complexity in the planning, maintenance, and operational activities. In the new liberalized paradigm, the consumers have become prosumers as the traditional distribution networks (DNs) have now transcended to active DNs [1]. Renewable energy technologies (RETs), such as solar photovoltaic, wind, hydro, and others, are continuously improving and becoming cheaper and more efficient. Therefore, the integration of RETs in power DNs is escalating massively. However, due to the increased distributed generation (DG) penetration, the liabilities of the distribution network planner (DNP) are increasing as well [2]. The impact analysis of DGs in the power network has become essential so that the reliability and power availability do not get degraded. Unplanned and arbitrary DG placement may result in negative improvements in DNs' quantities, viz., a surge in power losses, system voltage drops, and diminished loading capability [3].

Conventionally, power loss reduction and voltage maximization are two common objectives for the DG placement problem. Finally, the extensive use of power electronics' devices in the DN has caused power quality and voltage fluctuation issues, raising concerns about finding new DG placement solutions for reliable and economical operations in a power DN [4]. The DG placement has been formulated as a multiobjective problem (MOP) to take different technical and economic viewpoints simultaneously. The stud krill herd optimization technique is used in [5] to solve a weighted-sum MOP for the ideal location and sizing of DGs, where the appropriate weight preferences for the problem objectives have been taken subjectively. An analytical hierarchical process (AHP) technique is widely utilized for obtaining the weight coefficients in the weighted-sum method of MOP [6], [7], [8], [9]. However, the pairwise comparison matrix in the AHP technique is formed based on the decision maker's choice. The weighted normalized decision matrix is constructed based on predetermined weight coefficients in the technique for order preference by similarity to ideal solution to find the preferred alternative most near to the positive ideal solution while solving MOP for optimal siting and sizing of DGs in [9] and [10]. The selection of appropriate weight factors in weighted-sum MOP is still challenging and poses ambiguity [11].

Various methods have recently been developed based on the Pareto dominance concept to obtain tradeoff solutions among the different objectives in a MOP, such as Harris Hawks optimizer [11], nondominated sorting genetic algorithm [12], and multiobjective differential evolution [13]. However, relatively fewer efforts have been laid into presenting strategies for selecting the best tradeoff from a set of nondominated solutions, which still requires further research. Often, distance-based methods, such as gray relational analysis (GRA) [11] and fuzzy decision [12], [13], [14], are utilized instead of investigating the inherent relationship (i.e., relative influence) among objectives to find the best alternative.

Finally, it can be concluded that the relative influence coefficients (RICs) of the investigated indices in MOP formulations for optimal placement of DG are usually assessed empirically in previous studies, although it is worth noting that the impact of an objective (or indices) varies with the change in the network's physical properties. Therefore, achieving pragmatic solutions seems deceptive without impartially evaluating the RICs of indices in a MOP. This article presents a novel method based on the Shannon entropy (SE) formula to evaluate the relative influence of each considered index objectively and effectively in DN planning.

II. Problem Formulation

In generic, a weighted-sum multiobjective optimization problem for DG placement in DN can be written as follows:

$$\text{minimize} \quad \sum_{i=1}^{N} \xi_i * f_i(x) \tag{1}$$

$$\text{s.t.} \quad h(x) = 0 \tag{2}$$

$$g(x) \geq 0 \tag{3}$$

where N in (1) is the total number of objectives considered. In this article, ξ_i is referred to as the RIC representing the degree of importance of the ith objective function relatively in MOP formulation. In (2) and (3), the equality and inequality constraints represent the power flow balance, active and reactive power limits, voltage limits, and other operational and security bounds, as described in (4)–(8). In (1), ξ_i of each considered objective is usually based on the decision maker's choice, resulting in bias. This article proposed a new strategy based

Manuscript received 21 June 2022; revised 24 October 2022 and 7 January 2023; accepted 24 February 2023. *(Corresponding author: Shubham Gupta.)*

The authors are with the Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India (e-mail: shubham.gupta6691@gmail.com; vinodkumar@dtu.ac.in; madhusudan@dce.ac.in).

Digital Object Identifier 10.1109/JSYST.2023.3250226

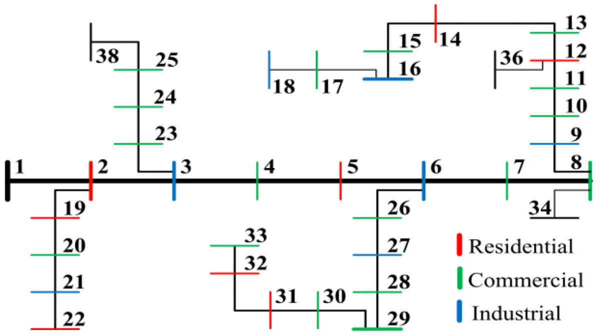


Fig. 1. Single line diagram of 38-node test system depicting residential, commercial, and industrial loads.

on the SE formula for assessing the RIC of each objective function equitably in a multiobjective optimally DG placement problem.

The DN power flows' equations [15] and DG operating constraints to be satisfied $\forall f \in \Omega_n \ \forall \ell = ft \in \Omega_L$, and $\forall i \in \Omega_{\text{DG}}$ are

$$-P_{\text{inj},t} = -\left(P_{g,t} - P_{d,t}\right)$$
$$= P_{ft} - \frac{r_{ft}}{V_f^2}\left\{P_{ft}^2 + Q_{ft}^2\right\} - \sum_{c:t\to c} P_{tc} \tag{4}$$

$$-Q_{\text{inj},t} = Q_{d,t} = Q_{ft} - \frac{x_{ft}}{V_f^2}\left\{P_{ft}^2 + Q_{ft}^2\right\} - \sum_{c:t\to c} Q_{tc} \tag{5}$$

$$V_t^2 = V_f^2 + \frac{r_{ft}^2 + x_{ft}^2}{V_f^2}\left\{P_{ft}^2 + Q_{ft}^2\right\}$$
$$-2\left\{r_{ft}P_{ft} + x_{ft}Q_{ft}\right\} \tag{6}$$

$$V_f^{\min} \le V_f \le V_f^{\max} \tag{7}$$

$$P_{g,i}^{\min} \le P_{g,i} \le P_{g,i}^{\max} \tag{8}$$

where $P_{\text{inj},t}$ and $Q_{\text{inj},t}$ in (4) and (5) are the net injected real and reactive power to the tth bus, respectively. $P_{d,t}$ and $Q_{d,t}$ are the active and reactive power demands at the tth bus, respectively. P_{ft} and Q_{ft} are the active and reactive power flow between buses f and t, respectively. V_f, $V_f^{\min}$, and $V_f^{\max}$ in (7) are the voltage magnitude and lower/upper limits of V of the fth bus, respectively. r_{ft} and x_{ft} are the resistance and reactance of the line connecting buses f and t, respectively. $P_{g,i}$ in (8) is the active power generated by the connected DG at the ith bus.

III. SE Incorporating DNP Priority Information (PI)

The SE is the most practiced method in information theory to measure the value dispersion in multicriteria decision making [16]. The fundamental principle of the SE is that the greater the dispersal in the measured value, the greater the differentiation of the criteria; as a result, more information can be obtained, moreover, the higher the importance of that criteria, and vice-versa. The SE formula is expressed as follows:

$$e_i = -e_0 \sum_{k=1}^{s} \bar{\text{I}}_{ki}.\ln \bar{\text{I}}_{ki} \tag{9}$$

here e_o is termed as the entropy constant and is equal to $(\ln s)^{-1}$, $\bar{\text{I}}_{kl}$ is the normalized kth value of the ith index. From (9), the RI coefficient can be measured as follows:

$$\xi_i = \frac{(1-e_i)}{\sum_{i=1}^{N}(1-e_i)}. \tag{10}$$

TABLE I
DN Indices Consider for DG Placement Problem

	DN Indices	Mathematical Notation
Technical Indices [17]	APLI	$f_1 = 1 - \left(\frac{\sum_{t\in U_t} P_{Loss}^{k,t}}{\sum_{t\in U_t} P_{Loss}^{0,t}}\right)$
	Reactive Power Loss Index	$f_2 = 1 - \left(\frac{\sum_{t\in U_t} Q_{Loss}^{k,t}}{\sum_{t\in U_t} Q_{Loss}^{0,t}}\right)$
	MVDI	$f_3 = 1 - \frac{\sum_{b\in U_b}\left(\left\lvert\bar{V}_o\right\rvert - \left\lvert\bar{V}_b^k\right\rvert / \left\lvert\bar{V}_o\right\rvert\right)^4}{\sum_{b\in U_b}\left(\left\lvert\bar{V}_o\right\rvert - \left\lvert\bar{V}_b^0\right\rvert / \left\lvert\bar{V}_o\right\rvert\right)^2}$
	LCLI	$f_4 = 1 - \frac{1}{T}\sum_{t\in U_t} \max\left(\left\lvert\bar{S}_l^{k,t}\right\rvert / S_l^{\max}\right)^{l\in U_l}$
Economic Index	Annual Equivalent Cost Index	$f_5 = 1 - \left(\sum_{t\in U_t} C_{cost}^{k,t} / \sum_{t\in U_t} C_{cost}^{0,t}\right)$ where $C_{cost}^{k} = \left(P_{DG}^{k} \times C_I^{DG} \times CRF\right)$ $+\left(P_{DG}^{k} \times C_{OM}^{DG}\right) + \left(C_E\left(P_{Loss}^{k} + P_L\right)\right)$
Reliability Index [19]	Average Energy Not Supplied Index	$f_6 = 1 - \left(\sum_{q=1}^{L} P_q^k \times U_q / \sum_{q=1}^{L} P_q^0 \times U_q\right)$

In some given conditions, the DNP strategically sets a preference for an index; thus, RI coefficients must be changed accordingly. This PI can be incorporated by modifying the ξ_i in (10) for "n" indices with PI as in (11) and "m" indices without PI as in (12).

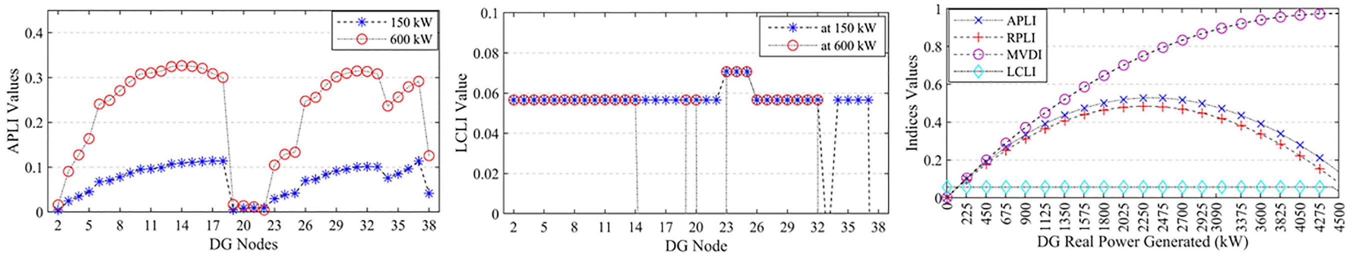
$$\xi_{n\in N} = \frac{(1-e_n) + \left(\zeta_n * \sum_{i=1}^{N}(1-e_i)\right)}{\sum_{i=1}^{N}(1-e_i)} \tag{11}$$

$$\xi_{m\in\{N\backslash n\}} = \frac{(1-e_m)}{\sum_{i=1}^{N}(1-e_i)} - \frac{\sum_{n\in N}\zeta_n * (1-e_m)}{\sum_{i=1}^{N}(1-e_i) - \sum_{n\in N}(1-e_n)} \tag{12}$$

The AHP technique provides flexibility to customize the weights associated with indices in a MOP, which is mainly useful for subjective analysis [17]. On the contrary, DN planning is a type of complex problem where the RICs of indices need to be assessed objectively based on their physical significance, which can be handled very efficiently by the proposed method. Furthermore, it allows the incorporation of PI given by the DNP in certain situations.

IV. Solution Methodology

Finally, numerous methodologies have been suggested for the optimal placement of DG in power DNs, and in them, many methods focus on multiobjective optimization. However, no adequate strategy is directed to quantity the RIC of each objective resulting in a chance of getting a pseudo-optimal solution. Therefore, this article presents a novel approach based on the SE formula to solve the multiobjective optimal DG placement problem. The SE formula relatively quantifies the degree of importance of indices in formulated MOP. The proposed approach is solved in the following steps.



(a) (b) (c)

Fig. 2. Technical indices scores for a 38-node test system in different DG placement problems. (a) APLI scores. (b) LCLI scores. (c) Indices scores at sixth bus.

TABLE II
SIMULATION RESULTS FOR ALL FOUR CONSIDERED CASES

Cases	RI Coefficients						PI	(ζ_i) Optimal Solution	DN Quantities				
	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6			APL (kW)	RPL (kVAr)	$V_{\min}$ (pu.)	AEC (M$)	AENS
Base	-	-	-	-	-	-	-	-	152.80	101.50	18/0.9252	-	1.6732
Case-1	0.24	0.27	0.10	0.39	-	-	-	Sixth bus/2523 kW	72.23	52.74	18/0.9616	1.718	1.6236
Case-2	0.34	0.23	0.09	0.34	-	-	ζ_1 =10%	Sixth bus/2480 kW	72.06	52.58	18/0.9611	1.724	1.6227
Case-3	0.29	-	-	-	0.34	0.37	ζ_1 =10%	Sixth bus/2834 kW	75.04	54.92	18/0.9660	1.678	1.6165

TABLE III
COMPARISON AMONG DIFFERENT EXISTING METHODS AND PROPOSED APPROACH

Methods		AHP	GRA*	FDM*	Proposed Approach
RIC/ Preference Weights	ξ_1	0.54	-	-	0.19
	ξ_5	0.16	-	-	0.39
	ξ_6	0.30	-	-	0.42
Optimal Solution		Sixth bus/ 2492	Sixth bus/ 1995	Sixth bus/ 1995	Sixth bus/ 3180
APL (kW)		72.11	73.52	73.52	81.09
RPL (kW)		52.63	52.95	52.95	59.22
$V_{\min}$ (p. u.)		18/0.9613	18/0.9543	18/0.9543	18/0.9707
AEC (M$)		1.722	1.788	1.788	1.636
AENS		1.6234	1.6332	1.6332	1.6096

* GRA and FDM strategies determine the best-compromised solution from a set of nondominated solutions of Pareto-front.

1) Gather the input data and solve the load flow problem for the base case without the placement of DG for comparison and analysis of results with the proposed approach.
2) Define the objectives of MOP and evaluate their RIC using the SE-based strategy, as described in Section III, satisfying the constraint as given in the following:

$$\sum_{i=1}^{N} \xi_i = 1. \tag{13}$$

3) Formulate the weighted-sum multiobjective optimal DG placement problem and obtain the optimal solution

$$\text{maximize} \quad \sum_{i=1}^{N} \xi_i * f_i'(x) \tag{14}$$

$$\text{s.t.} \quad (4) - (8). \tag{15}$$

In (14), the problem objectives $f_i(x)$ are reconstructed to model them unitless for the relative measurement of RIC, where the $f_i'(x)$ is elucidated as the index in this article

$$f_i' = 1 - \left(\frac{f_i^k(x)}{f_i^{\text{base}}(x)}\right). \tag{16}$$

The formulated problem is a mixed-integer programming problem that simultaneously finds the optimal location and size of the DG in DN planning and could be solved through any classical or nature-inspired multiobjective optimization technique. This work primarily focuses on objectively determining the relative influence of indices in DN planning.

V. CASE STUDY

The proposed method is examined on a 38-node test system [18], consisting of the total active and reactive power demand of 3501.6 kW and 1870.2 kVAr, respectively. The system configuration has different types of loads, viz., industrial, commercial, and residential, as shown in Fig. 1. For the simulation, the base quantities S_{base} and kV_{base} are taken for analysis as 100 MVA and 12.66 kV, respectively. Considering many indices for measuring the alteration in the network quantities during the DG placement problem analysis is a colossal need to tend toward the practical application. This work considers four technical indices, one economic and one reliability index, to measure their influence on DN planning while formulating the DG placement problem. Mathematically, these indices are expressed in Table I.

A. *Influence of DN Indices on DG Placement Problem*

It is a well-known fact that the DG placement alters the DN quantities, viz., change in real and reactive power losses, reshaping the voltage profile, and regulating the line capacity, which can be elucidated through Fig. 2(a)–(c). The greater the degree of dispersion in DN quantity, the higher the impact of the designated index in the MOP problem. From Fig. 2(a) and (b), it can be observed that the active power loss index (APLI) scores are more dispersed compared with line capacity limit index (LCLI) scores when one DG of 150 kW is placed in DN. Hence, APLI relatively influences more the MOP problem than LCLI; therefore, RIC should be more for APLI. On the contrary, when a DG of size 600 kW is placed, the line capacity limit is violated for many configurations; therefore, priority should be given to LCLI, and the RIC of LCLI should be put higher than to APLI. Fig. 2(c) represents the various indices value when DG size is varied for a node location (i.e., sixth bus). In Fig. 2(c), the LCLI scores barely changed; in contrast, the maximum voltage deviation index (MVDI) shows the highest dispersion among all indices.

B. *Simulation Results of the Proposed Approach*

The following four cases are simulated and analyzed for the DG placement problem to illustrate the effectiveness of the proposed approach.

- Case-1: Optimal placement and sizing of DG, considering technical indices (f_1, f_2, f_3, and f_4).
- Case-2: Optimal placement and sizing of DG, considering technical indices (f_1, f_2, f_3, and f_4) and index PI.
- Case-3: Optimal placement and sizing of DG, considering technical, economic, and reliability indices (f_1, f_5, and f_6) and index PI.

The problem formulation is simulated for the placement of a single DG in DN; however, the proposed approach can obtain the solution for multiple DGs in both simultaneous or sequential placement manner.

C. *Comparison Study*

This article compares the proposed methodology with the previously published methods, i.e., AHP, GRA, and fuzzy decision method (FDM), to validate the efficacy. The comparison results are reported in Table III. The optimal DG size and location are assessed considering the active power loss (f_1), cost factor (f_5), and reliability factor (f_6) indices.

The prevailing studies did not articulate the fact that the RICs differ as the problem formulation changes for the DG placement problem, which can be deduced from the simulated results in Tables II and III. On this basis, it could be said that the proposed approach can produce an objective and comprehensive solution with physical significance.

VI. Conclusion

A novel framework to measure the influence of indices on the DG placement problem during DN planning is proposed, which utilizes the SE formula to quantify the relative degree of importance objectively and effectively. The main advantages of the proposed method are that incorporates the index PI given by DNP in inevitable situations and is easily executable, and fit in with other similar multiobjective optimization problem.

References

[1] L. Tziovani, L. Hadjidemetriou, P. Kolios, A. Astolfi, E. Kyriakides, and S. Timotheou, "Energy management and control of photovoltaic and storage systems in active distribution grids," *IEEE Trans. Power Syst.*, vol. 37, no. 3, pp. 1956–1968, May 2022.

[2] S. Riaz and P. Mancarella, "Modelling and characterisation of flexibility from distributed energy resources," *IEEE Trans. Power Syst.*, vol. 37, no. 1, pp. 38–50, Jan. 2022.

[3] S. Maity, S. Paul, H. Karbouj, and Z. H. Rather, "Optimal sizing and placement of wind farm in a radial distribution network considering reliability, operational, economic and environmental factors," *IEEE Trans. Power Del.*, vol. 36, no. 5, pp. 3043–3054, Oct. 2021.

[4] P. Prakash and D. K. Khatod, "Optimal sizing and siting techniques for distributed generation in distribution systems: A review," *Renewable Sustain. Energy Rev.*, vol. 57, pp. 111–130, May 2016.

[5] S. A. Chithra Devi, K. Yamuna, and M. Sornalatha, "Multi-objective optimization of optimal placement and sizing of multiple DG placements in radial distribution system using stud krill herd algorithm," *Neural Comput. Appl.*, vol. 33, pp. 13619–13634, 2021.

[6] K. Babu B. and S. Maheswarapu, "A solution to multi-objective optimal accommodation of distributed generation problem of power distribution networks: An analytical approach," *Int. Trans. Elect. Energy Syst.*, vol. 29, no. 4, 2019, Art. no. e12093, doi: 10.1002/2050-7038.12093.

[7] P. Gangwar, S. N. Singh, and S. Chakrabarti, "Multi-objective planning model for multi-phase distribution system under uncertainty considering reconfiguration," *IET Renewable Power Gener.*, vol. 13, no. 12, pp. 2070–2083, Sep. 2019.

[8] S. S. Tanwar and D. K. Khatod, "Techno-economic and environmental approach for optimal placement and sizing of renewable DGs in distribution system," *Energy*, vol. 127, pp. 52–67, May 2017.

[9] C. Venkaiah and R. V. Jain, "Multi-objective JAYA algorithm based optimal location and sizing of distributed generation in a radial distribution system," in *Proc. IEEE PES Asia-Pacific Power Energy Eng. Conf.*, 2017, pp. 1–6, doi: 10.1109/APPEEC.2017.8308965.

[10] V. S. Galgali, M. Ramachandran, and G. A. Vaidya, "Multi-objective optimal placement and sizing of DGs by hybrid fuzzy TOPSIS and Taguchi desirability function analysis approach," *Electr. Power Compon. Syst.*, vol. 48, no. 19/20, pp. 2144–2155, 2020.

[11] A. Selim, S. Kamel, A. S. Alghamdi, and F. Jurado, "Optimal placement of DGs in distribution system using an improved Harris Hawks optimizer based on single- and multi-objective approaches," *IEEE Access*, vol. 8, pp. 52815–52829, 2020.

[12] W. Sheng, K.-Y. Liu, Y. Liu, X. Meng, and Y. Li, "Optimal placement and sizing of distributed generation via an improved nondominated sorting genetic algorithm II," *IEEE Trans. Power Del.*, vol. 30, no. 2, pp. 569–578, Apr. 2015.

[13] S. R. Behera and B. K. Panigrahi, "A multi objective approach for placement of multiple DGs in the radial distribution system," *Int. J. Mach. Learn. Cybern.*, vol. 10, pp. 2027–2041, 2019.

[14] M. Mosbah, S. Arif, and R. D. Mohammedi, "Multi-objective optimization for optimal multi DG placement and sizes in distribution network based on NSGA-II and fuzzy logic combination," in *Proc. IEEE 5th Int. Conf. Elect. Eng. - Boumerdes*, 2017, pp. 1–6, doi: 10.1109/ICEE-B.2017.8192171.

[15] H. Bagheri Tolabi, M. H. Ali, and M. Rizwan, "Simultaneous reconfiguration, optimal placement of DSTATCOM, and photovoltaic array in a distribution system based on fuzzy-ACO approach," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 210–218, Jan. 2015.

[16] Z. Liu, N. Wu, Y. Qiao, and Z. Li, "Performance evaluation of public bus transportation by using DEA models and Shannon's entropy: An example from a company in a large city of China," *IEEE/CAA J. Automatica Sinica*, vol. 8, no. 4, pp. 779–795, Apr. 2021.

[17] N. Munier and E. Hontoria, "Shortcomings of the AHP methods," in *Uses and Limitations of the AHP Method: A Non-Mathematical and Rational Analysis*. New York, NY, USA: Springer, pp. 41–90, 2021, doi: 10.1007/978-3-030-60392-2_5.

[18] D. K. Patel, D. Singh, and B. Singh, "A comparative analysis for impact of distributed generations with electric vehicles planning," *Sustain. Energy Technol. Assessments*, vol. 52, no. Part A, Aug. 2022, Art. no. 101840, doi: 10.1016/j.seta.2021.101840.

[19] M. Bilal, M. Rizwan, I. Alsaidan, and F. M. Almasoudi, "AI-based approach for optimal placement of EVCS and DG with reliability analysis," *IEEE Access*, vol. 9, pp. 154204–154224, 2021.

International Journal of Mechanical and Production
Engineering Research and Development (IJMPERD)
ISSN(P): 2249–6890; ISSN(E): 2249–8001
Vol. 10, Issue 3, Jun 2020, 7625–7638
© TJPRC Pvt. Ltd.

MECHANICAL PROPERTIES OF COLD METAL TRANSFER (CMT) WELDED 202 AND 304 STAINLESS STEELS

PRATEEK RAJ & N. YUVARAJ

Department of Mechanical Engineering, Delhi Technological University

ABSTRACT

Welded joints of dissimilar stainless steel sheets have many industrial applications due to their cost effectiveness, light-weight, and high efficiency. Thin austenitic stainless steel sheets of different thicknesses are extensively used in the automation industry. Conventional welding techniques due to their high heat input and high spatters have always posed problems such as burn-through and distortion for welding these joints of thin austenitic steels. The CMT welding technique is effectively used for the joining of thin sheets due to its characteristics of lower distortion rate and low heat input. In this research work, austenitic stainless steels of grade SS202 and SS304 of thickness 1.2 mm and 2 mm respectively were welded by CMT welding technique and studied the mechanical characteristics of dissimilar stainless steel joints. Taguchi L9 optimization technique was used to find the optimized process variables for obtaining the maximum tensile strength. The maximum tensile strength of dissimilar joint welded at115 A current, 4 mm/s welding speed, and 10% arc length correction factor was found to be 291 MPa. The maximum micro-hardness value of the weld zone was achieved and the lower value was observed in the heat-affected zone (HAZ). The X-ray diffraction (XRD) technique was utilized to detect the residual stresses of the welded joint. The residual stress was observed in compressive nature in the weld zone and base plate and of tensile nature in the heat-affected zone (HAZ). CMT welding process can produce high strength dissimilar austenitic steel joints of different thicknesses.

KEYWORDS: CMT; SS 202/304; Tensile Strength; Hardness; Residual Stresses

Received: Jun 08, 2020; **Accepted:** Jun 28, 2020; **Published:** Aug 11, 2020; **Paper Id.:** IJMPERDJUN2020725

INTRODUCTION

Fusion welding is considered as one of the most important techniques for the manufacturing of various components. Mostly a complete product may require varied properties at different positions like one area may need to be corrosion resistant while the other must be resistant to heat. One area may require high toughness while others may require high strength. It is very difficult to manufacture any product without using the joining process due to these technological limitations. Several parts of the products are typically assembled by the fusion practices and mostly help to fabricate economically [1]. Joining different metals is preferred to help in giving benefits of various materials and that offer distinctive solutions to different engineering requirements [2]. A reduction in weight and cost of the product without hindering the structural requirements and safety is one of the basic advantages of fusing different materials

There are many joining processes for dissimilar materials that have gained remarkable consideration in recent years. The dissimilar fusion weld must acquire satisfactory tensile and ductility test results so that the joint will be successful within the weld [3]. MIG/MAG welding technique is the most preferable process for the joining of dissimilar ferrous and non-ferrous metals due to its supreme weldment characteristics. This method is especially preferred in applications related to the automotive industry. However, with the recent shift of these industries

Original Article

towards environmental sustainability and safety of passengers, different grades of steel are now being used for fabrications. Thin sheet materials alloys have a high coefficient of thermal conductivity and thermal expansion and so they pose some problems like burn through and distortion during arc welding. Welding dissimilar materials of different thicknesses having limitations with the conventional welding process due to high heat input and high spatter. Controlled heat input is an essential parameter to avoid such difficulties [4,5]. Thus a need for a welding technique arises which can be used to join thin sheets and eradicate these problems. CMT techniques lessen these difficulties to a great extent. CMT welded joint has narrow HAZ with lesser distortion makes this technique preferable for joining thin plates with improved productivity [6].

Cold Metal Transfer (CMT) welding is a newly introduced process of joining thin sheets based upon the conventional short-circuiting (CSC) transfer technique established via "Fronius of Austria". It is a technological enhancement to Gas Metal Arc Welding (GMAW) process and is highly superior to GMAW in terms of lesser spatter, distortion, burns-through, and welding cost due to its unique feature known as low heat input. In this welding process, as the arcing starts the electrode moves toward the weld puddle. As the tip of the electrode interacts with the molten metal within the weld pool, an arc is extinguished. The value of current reduces to a non-zero value and this in a way circumvents the chances of spatter. The dropped welding current value results in a significant reduction in the thermal heat input. This becomes the most favorable conditions and a feasible technique to weld thin sheets with no or very less distortion, a lower rate of dilution and lower stresses in the weld zone. It also provides high gap bridge-ability which is highly appropriate for automation. CMT welding is an automatic welding technique having a controlled deposition of material during the short-circuiting of the work-piece to an electrode and is well described for its working with a low heat input [7]. This results in less damage to the base metal, lower deformation, and residual stress. In a conventional arc welding process, the filler wires move continuously in the outward direction till a short-circuit takes place. Whereas in a CMT process, the filler wire is both pushed as well as retracted during welding, and thus it is called an intelligent system. In this process, the movement of the feeding wire with an oscillating frequency up to 70 Hz is mostly used [8].

Stainless steels have gained popularity as one of the most useful materials in industrial applications due to high resistance to corrosion. Their exceptional physical properties and design codes have made them functional in engineering applications such as structural applications, heat exchangers, and thermal power plants [9-13].Yan et. al. (2010) [14] reported that the physical and metallurgical properties of the stainless steel TIG welded joint improved due to the presence of delta ferritic and gamma ferritic phases. During diffusion, the transformation of the austenitic phase to martensite occurs and the martensite phase improves high strength [15]. Amongst the various grades of austenitic steels available, SS304 a prominent member of 300 series is an important and most used grade due to its good corrosion resistance, higher strength, and ductility. SS202 is one of the most preferred stainless steel of the 200 grade series. They are similar to 300 series sheets of steel except for the low nickel content in them. SS 202 steel grade is economical material with the use of extensive applications due to the property of excellent toughness at a lower temperature. The welding of dissimilar grades of SS304 and SS202 has an excellent future scope and industrial applications. Both SS304 and SS202 are austenitic grade steels containing gamma iron under equilibrium cooling conditions. But during rapid cooling, an incomplete transformation occurs which leads to the formation of some meta-stable comprising delta iron [16-17]. The physical properties of inoconel 718 welded with SS316 have concluded that dissimilar weld gave a higher tensile strength than the parent SS316 metal [18]. Investigations and analysis have shown that most of the failures occurred in the heat-affected zone (HAZ) [19]. Sathiyaet. al. (2005) [20] have studied the friction welding of SS 304 and concluded based on fractography that fissure occurred most of the time at the joint zone rather than the base metal. Tensile reports show an inverse relation between friction time and joints strength. Kumaret. al. (2016) [21] in his experimentation on the effects of CMT welding process on aluminium found that an increased heat input and better fluidity can be achieved at a lower welding speed. Varghese et. al. (2019) [22] in his experimentation of coating in conel 617 M on austenitic stainless steel by CMT welding process found a direct relationship between the heat input per unit length and the welding current. Sammaiahet. al. (2010) [23] in his study on metallurgical properties of friction welded aluminium and austenitic steels showed increased tensile strength and reduced toughness with an increase in the pressure due to friction. Mishra et. al. (2014) [24] in his investigations about the strength of mild steel welded with different grades of steel found that the welded joints of SS 202 and mild steel gave the best tensile strength value with TIG and MIG welding. Proper selection of filler wire is an important criterion in fusion welding. The strength and hardness of the welded zone depend upon the selection of the filler wire. SS308 filler wire is the most suitable and recommended filler wire material for joining SS304, SS32, and SS347 [25]. Fusion welding depends on various input variables such as welding current, voltage, wire feed rate, welding speed, contact tip to workpiece distance (CTWD), etc. Arc length also controls the microstructure and the strength of the weld joints. The arc length correction while joining must be non-linear and results in a controlled dilution in the weld zone which has an advantage during the joining of sheets [26]. Most of the investigations on the welding are carried on the sheets or plates of similar equal thickness. Though, most of the welding required between the automotive parts is between materials of dissimilar thickness [27-28].

Joining dissimilar grades of steel is difficult due to the difference in melting temperatures. The different thermal conductivities and uneven temperature distribution on the weld surface result in the generation of residual stresses. The residual stresses are the internal stress that remains in the bodies when subjected to uneven or non-uniform temperature conditions when no external load is applied [29-32]. These are macroscopic stresses and static quantities, the value of which varies from zero to the material yield point.

The measurement of residual stress in the welded joint is an important factor for evaluating the mechanical characteristics of materials. The distribution and location of the residual stresses, types of stress in the material and, fatigue properties in specimens can be predicted easily [33-34].The residual stresses are also likely to change the vulnerability for various modes of failures such as corrosion fatigue stress, corrosion fracture, and cracking. X-ray diffraction method is used for measuring the residual stresses and for analyzing the mechanical structures. The static behavior can be studied from a microscopic and macroscopic level change [35].

From the literature welding of dissimilar austenitic steel has become challenging and more investigations require for maintaining perfect arc length and higher edge tolerances. For welding of thin dissimilar materials from conventional MIG welding is a difficult task. CMT welding has become more popular for welding of thin sheets which is required for most industrial applications. In the present work, SS 304/202 and thickness of 2 mm and 1.2 mm respectively were welded. The Taguchi optimization technique is an effective method for finding the optimized parameter along with their relationship [36]. In this study, Taguchi L9 method is adopted for finding the optimized welding parameters in order to achieve the maximum tensile strength of the welded joint. The mechanical characteristics such as tensile strength, hardness, and, residual stresses of the welded joint were investigated.

EXPERIMENTAL PROCEDURE

Material and Methods

In the present research, the austenitic stainless steel grades of SS 304 and SS 202 and thickness of 2 mm and 1.2 mm respectively were CMT welded with SS 308 of 1.2 mm diameter filler wire. The CMT welding experimental setup is shown in fig. 1. Recent advancements in the modern industry have found many applications of tailor welded blanks (TWB) which are made from single sheets of steel of dissimilar thickness, coating, and, strength which are welded together[37].Flexible part designs are allowed in this manufacturing procedure and it is ensured that the right amount of material is used in the right place. The shielding gas of 98% argon and 2% carbon mixture has been used for welding. During the process of welding, the shielding gas usually interacts with the filler metal which results in enhancement of mechanical as well as corrosion resistance properties of the weld deposits. An increase in CO_2 content % in the Argon + CO_2 mixture improves the wet ability of molten filler wire and fusion volume. It also leads to increased spatter rates and a decrease in the ferrite numbers [38]. The chemical composition of SS 202 sheet and SS 304 sheet is given in table 1 and table 2 respectively.

Table 1: Chemical Composition of SS 202

Fe	C	Si	Mn	P	S	Cr	Mo	Ni	Al
73.9	0.103	0.490	10.5	0.0730	0.0179	12.8	0.303	0.205	<0.002

Table 2: Chemical Composition of SS 304

Fe	C	Si	Mn	P	S	Cr	Mo	Ni	N
71.9	0.0585	0.219	0.837	0.0426	0.0166	18.3	0.157	8.28	0.1

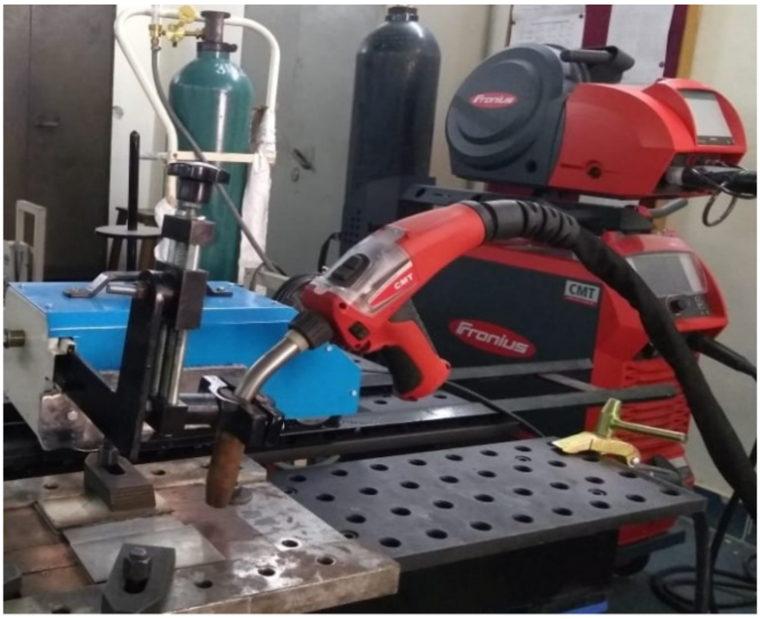


Figure 1: CMT Welding Experimental Set Up.

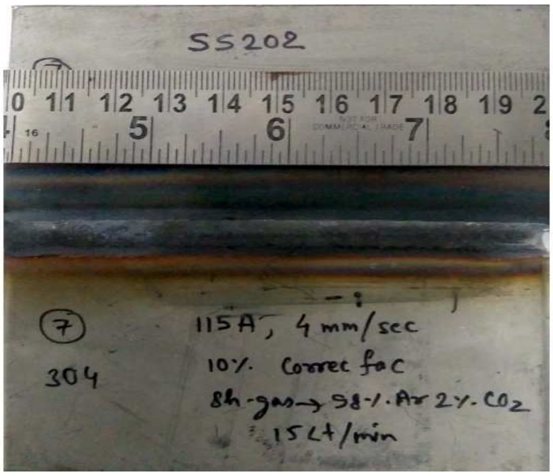


Figure 2: Welded Sample C7 at 115 A Current, 4 mm/s Welding Speed, 10% Arc Length Correction Factor.

Welding Parameters

A number of preliminary trial runs were conducted to set the welding parameters. The parameters are chosen in a way that the plates are welded correctly without any damage or burn through. The thickness ratio of the two plates is 1.67, which may result in local stress concentration and a shift in the neutral axis. The sheets are perfectly clamped in welding fixtures and necessary care has been taken to avoid distortion. Taguchi L9 orthogonal array has been applied here to optimize the welding parameters. Welding parameters such as welding current, welding speed and, arc length correction factor have been taken for welding of test specimens. Table 3 shows the process parameter values have taken for 3 different levels as per the L9 Taguchi technique. The shielding gas flow rate 15 L/min and contact tip to work-piece distance (CTWD) 10 mm is kept constant for all the samples. Fig.2 shows a sample welded at 115 A current, 4 mm/s welding speed and, a 10% arc length correction factor.

Table 3: Welding Process Parameters

Sample No.	I (A)	W.S (mm/s)	A.C.F (%)	CTWD (mm)
C1	75	4	-10	10
C2	75	5	0	10
C3	75	6	10	10
C4	95	4	0	10
C5	95	5	10	10
C6	95	6	-10	10
C7	115	4	10	10
C8	115	5	-10	10
C9	115	6	0	10
I = Current, W.S = Welding Speed, A.C.F = Arc Correction Factor, Flow Rate Of Shielding Gas = 15 Ltr/Min				

RESULTS AND DISCUSSIONS

Tensile Test (UTM)

The welded specimen for the tensile strength testing was cut as per ASTME8 standard using wire EDM process. The size of the tensile test specimen is shown in fig. 3. The tensile specimens before and post testing are shown in fig. 4 and fig. 5 respectively. TINIUS OLSEN H50KS tensile testing machine capacity of 50 KN was used to investigate the tensile properties of the weld specimens. The strain rate of 1 mm per min at room temperature is fixed for tensile testing.

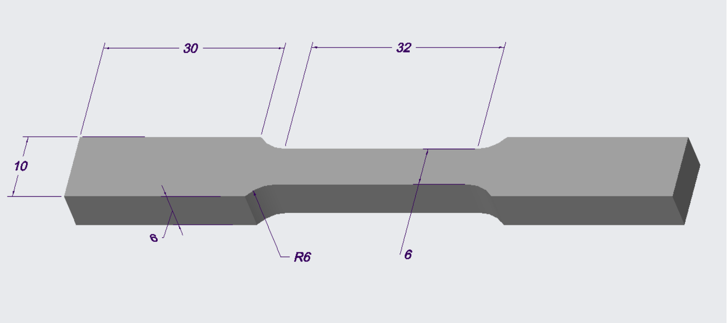


Figure 3: Tensile Testing Specimen as per ASTM E8 std.

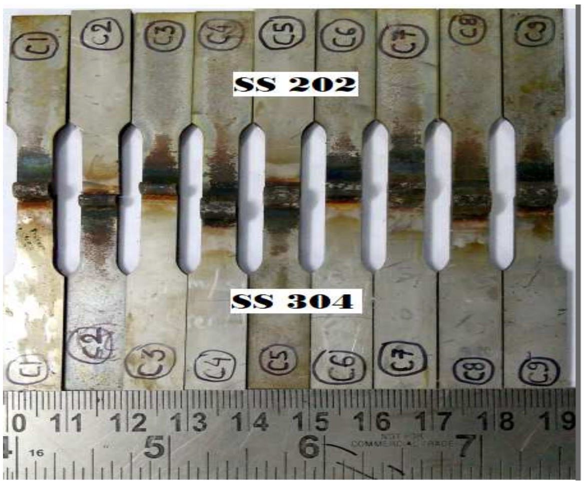


Figure 4: Tensile Specimens Before Testing.

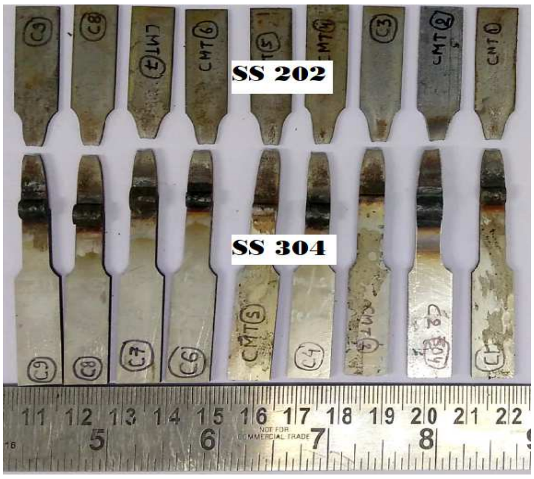


Figure 5: Specimens Post Tensile Test.

Welding entails the melting and solidification of the base metal. Welding of dissimilar materials involves fusion of two different materials having different solidification rate. This ultimately changes the microstructure and the grain size. The weld zone bead has high strength due to the alloy genesis. The filler wire along with the melted base metal forms this alloy thus making it the strongest portion. The typical tensile stress-strain graph of welded samples (C1, C4, C5, C7, C8) is shown in fig. 6. The tensile strength and elongation results obtained for the welded sample are shown in table 4. The maximum tensile strength achieved is 291 MPa at 115 A current, 4 mm/s welding speed, and 10 % arc length correction factor. The tensile results showan increase in weld strength as the welding current is increased and no significant change in the elongation was observed. Prakash et al. (2017) [39], reported that sample SS202-SS316 of 1.5 mm sheets spot welded specimen tensile strength was 268 MPa. In this study, the maximum tensile strength value of 291MPa was achieved. This strength was achieved at 4 mm/s welding speed and 10 %arc length correction factor. Arc length is the distance between the end of the filler wire and workpiece material. An arc length correction factor is also an important process parameter to determine and maximize the tensile strength of welded material. A positive arc length provides better strength and perfect penetration. Kannan et.al. (2019) [40] results support for a significant increase in strength for samples C1, C5 and, C7 with positive arc length correction factor. It was observed that the maximum elongation observed with a higher positive arc length correction factor welded sample.

Table 4: Tensile Properties

Sample No.	C1	C2	C3	C4	C5	C6	C7	C8	C9
Tensile Strength(MPa)	274	270	272	275	284	276	291	289	278
Elongation (%)	18.3	19.2	20.4	22.4	23.4	21.3	23.3	22.8	21.7

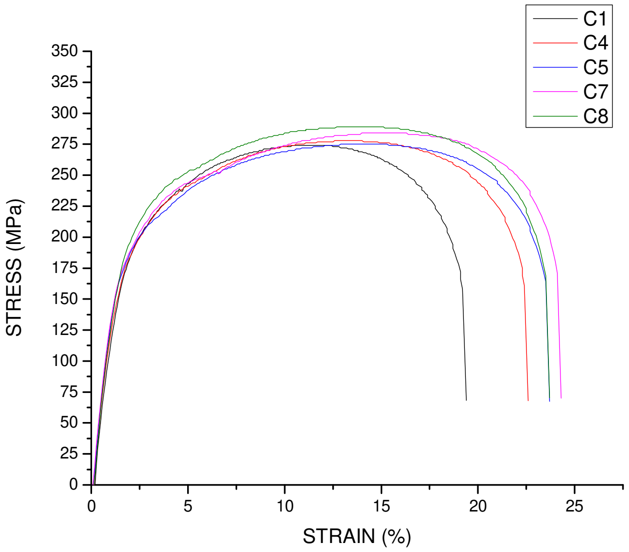


Figure 6: Typical Stress vs Strain Graph of Welded Specimen.

The tensile fracture of the welded specimen clearly shows the formation of a cup-cone shape when necking initiates or due to surface slip occurs. Joining dissimilar material thickness leads to uneven heat distribution and a higher amount of heating takes place in thicker sheets in comparison with smaller thickness sheets. These results lead to maximum penetration in the thicker plate and less take place in thin sheets [41].Marashiet. al.(2008) [42] observed that in spot welding of dissimilar thickness material's necking initiated in lower thickness sheets. The thinner sheets occur severe necking due to lesser force. The finite element results have shown the concentration of stress in the thinner section due to plastic distortion in the thinner part [43]. The efficiency of the welded joint was 100 % as all the failures occurred in the heat-affected zones (HAZ) of the thinner section of the SS 202 material. These results revealed that evidence of low heat input characteristics of CMT welding provides higher strength joints for dissimilar thickness.

Micro-Hardness (HV)

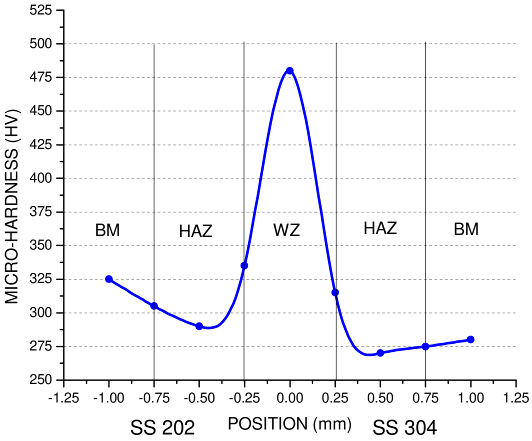


Figure 7: Micro-Hardness Variation for Sample C7 Welded at 115 A Current, 4 mm/s Welding Speed, 10% Arc Length Correction Factor.

Fig.7 shows the micro-hardness variations for the welded sample C7 welded at 115 A current, 4 mm/s welding speed, and 10 % arc length correction factor with respect to the positions. The higher hardness value was observed on the welded region and decreasing trend in the base metal (BM) and heat-affected zone (HAZ) for both the steel sheets. The micro-hardness of base material SS 202 and SS 304 are 325 HV and 280 HV respectively. The hardness in the material depends upon the carbon content and it controls the presence of cementite. The carbon content in SS 202 is slightly higher than SS 304 thus a difference in the hardness value of the base metal is observed. The micro- hardness of both SS 202 and SS 304 in the heat-affected zone (HAZ) reduces to 290 HV and 270 HV respectively. Sabooniet. al (2015) [44] reported that Friction stir welding of SS304 softening of HAZ takes place due to the recovery of coarser grains. According to the hall-petch relationship, the hardness and strength of a material are correlated to grain size. The reduction of micro-hardness in HAZ due to coarser grains. The hardness of the material relates to tensile properties in a material. The low hardness in HAZ is an important factor for the initiation of crack and fracture took place in this region during the tensile testing. Hardness is increasing towards the weld region. The increasing trend between HAZ and WZ is due to enhanced refinement grains of the austenitic steels. Similar type of pattern reported in the welding of SS 202 and SS 304 by TIG welding [45]. The hardness value achieved by CMT welding is higher than TIG welding. The hardness value of material depends upon the heat input that occurs during welding. Lesser heat input results in harder increases in weld region. The diffusion of chromium element between the two stainless steel sheets increases the hardness of weld zone. The highest micro-hardness value at the weld zone was 480 HV achieved for the C7 sample which is welded at the highest current value (115 A), slow welding speed (4 mm/s) and a positive arc length correction factor (10 %). The micro-hardness value in the welded region in the range of 460-480 HV due to its finer grain structure and excellent fusion of filler wire in the weld zone. Besides that the high rate of cooling in CMT process and an incomplete austenitic transformation in weld zone forms delta-ferrite phase. These are reasons for increasing the hardness in the welded region. Fig. 8 shows similar types of results for samples C1, C4, C5, and C8.

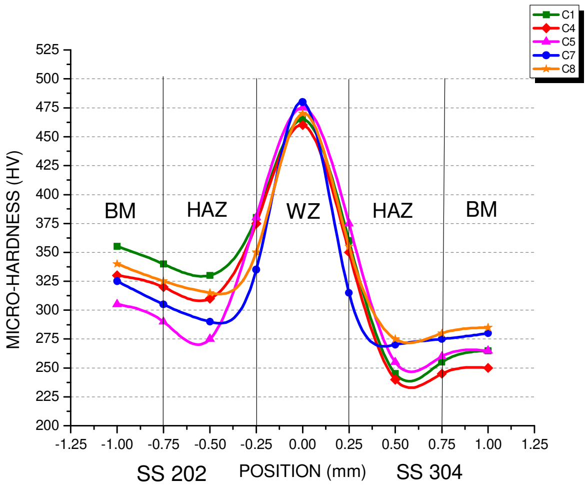


Figure 8: Micro-Hardness Variation with the Distance from the Weld.

Residual Stress (MPa)

Residual stresses are the stresses that are present within a body or a material after the process of manufacturing and material processing in the absence of temperature gradients or external loads [46].These stresses which are produced due to non-uniform distribution of temperature are measured by PULSTEC micro-X360n Full 2D X-ray residual stress analyzer, which operates on X-ray diffraction technique. It is associated with cos alpha method that uses a single exposure to collect the entire diffraction cone via a 2D detector. This system consisted of a sensor unit attached to a computer for output result and a power system. The sensor unit uses cos alpha method to calculate the residual stress. Amongst the number of available non-destructive techniques for measuring the residual stresses, X-ray diffraction is suitable for thin plates as its penetration is about 10µm with spatial resolution in the range of 10µm to 1mm, thus suitable for thin stainless steel plates. Some positions were marked comprising of the base metal plate (BM), heat affected zone (HAZ) and the weld zone (WZ) on both plates. Full debye-scherrer ring at each position was acquired through X-ray exposure. These rings determined the strain and finally the residual stress value at each position was produced [47].

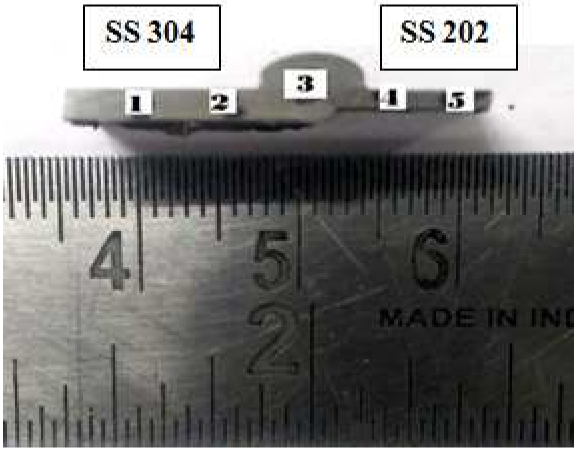


Figure 9: Residual Stress Points at Sample C7.

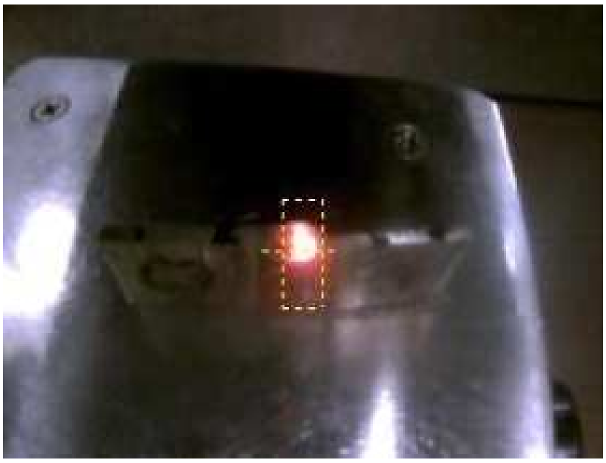


Figure 10: Position 3 (Sample C7) under X-Ray Diffractor.

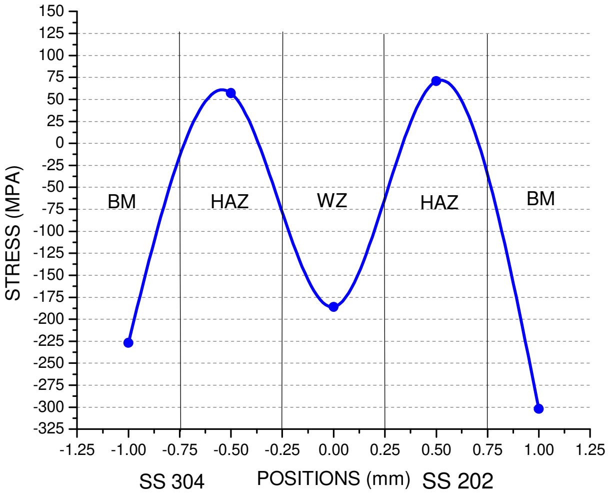


Figure 11: Residual Stress Variation at Different Zones of Sample C7 Welded at 115 A, 4 mm/s Welding Speed and 10 % Arc Length Correction Factor.

Residual stresses were taken at different points which are shown in Fig 9.X-ray diffractor position 3 is shown in Fig. 10. Standard chromium (Cr) material X-ray tube is used having collimator size of 1mm diameter with 30kV and 1mA specification for determining these stresses. The achieved residual stress values for the C7 welded sample for different regions are as shown in the graph in fig.11. Compressive residual stress was produced in the base plate and weld zone. Tensile residual stress was produced in the heat affected zone (HAZ). During the process of joining by CMT process, due to low heat input, upper and the lower weld zone surfaces experiences a higher rate of solidification due to rapid cooling than the material within the weld pool and the heat-affected zone (HAZ). This uneven rate of cooling leads to differential thermal distribution thus causing expansion in the heat affected zones and contraction in the weld zone. Due to this, the residual stress in the weld zone tends to become negative (compressive) due to shrinkage of the grain size (fine grains). The compressive stress in a way is desirable as it helps in avoiding the formation of cracks and notches. Relief from stress corrosion cracking also observed in the weld region. The slow cooling and coarser grains in the heat-affected zone (HAZ) results in a positive or tensile residual stress. This tensile residual stress present in the heat-affected zone is detrimental and results in fatigue failure. There may be chances of crack initiation and thus material fails most of the time in this region which leads to degradation of mechanical properties. The plots obtained in fig. 12for samples C1, C4, C5, C7, C8 between SS 304 and SS 202 was nearly the mirror images of each other. A similar type of results was observed in all the samples.

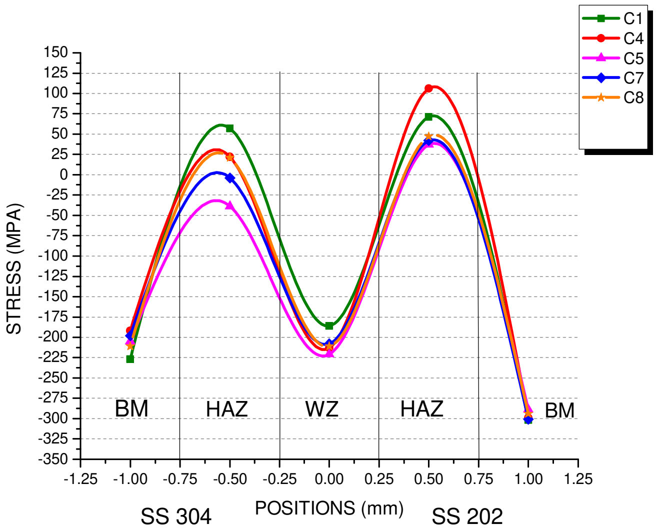


Figure 12: Residual Stress Variation at Different Zones of Welded Samples.

CONCLUSIONS

The paper investigates the mechanical characteristics of welded joints of dissimilar austenitic steels SS202 and SS304 thin sheets by cold metal transfer welding technique. The below mentioned are the conclusions drawn from this experimental study:-

- CMT welding is a suitable technique for welding thin stainless steel plates of dissimilar grades and different thicknesses.
- A tailor welded blank of blank ratio 1.67 can be made of stainless steel sheets of grades SS 304 and SS 202 with optimum strength and hardness.
- The highest strength equal to 291 MPa and highest hardness equal to 480 HV is achieved at 115 A current, 4 mm/s welding speed, and 10 % arc length correction factor.
- Low heat input and rapid cooling in CMT welding results in compressive (negative) residual stress in the weld zone making it the strongest and hardest zone.
- The HAZ is the most affected portion of the joint due to coarser grains and tensile residual stress because of the slow cooling here thus making it a hotspot for failure. The necking starts in the thinner section and results in cup-cone shaped fracture.

REFERENCES

1. *Tseng K, Lin Po-Yu. (2014), UNS S31603 Stainless Steel Tungsten Inert Gas Welds Made with Microparticle and Nanoparticle Oxides,7, 4755-4772*
2. *Gungor B., Kaluc E., Taban E ,and Aydin S. I. K. (2014), Mechanical and micro structural properties of robotic Cold Metal Transfer (CMT) welded 5083-H111 and 6082-T651 aluminum alloys. Materials & Design (1980-2015), 54, 207-211.*

3. *Ghosh N., Pal P. K., and Nandi G. (2017), GMAW dissimilar welding of AISI 409 ferritic stainless steel to AISI 316L austenitic stainless steel by using AISI 308 filler wire. Engineering Science and Technology, an International Journal, 20(4), 1334-1341.*

4. *Feng J., Zhang H., He P. (2009), The CMT short-circuiting metal transfer process and its use in thin aluminium sheets welding, Materials & Design, v. 30, pp. 1850–1852.*

5. *PickinC. G., Williams S. W., Lunt M. (2011), Characterization of the cold metal transfer (CMT) process and its application for low dilution cladding, Journal of Materials Processing Technology, v. 211, n. 3, pp. 496-502.*

6. *Rambabu, V., J. Ramarao, and S. Ravibabu. "Enhancement of Heat transfer in Shell and Tube heat exchanger by using nano fluid." International Journal of Mechanical and Production Engineering Research and Development, 7 (5) (2017): 191-198.*

7. *Koli Y., Yuvaraj N., Aravindan S., Vipin (2020), Multi-response mathematical modelling for prediction of weld bead geometry of AA6061-T6 using response surface mrthodology, 73(3):645-666*

8. *Pickin C. G., Young K. (2006), Evaluation of cold metal transfer (CMT) process for welding aluminium alloy, SciTechnol Weld Joi, v. 11, n. 5, pp. 583-585.*

9. *Thakur, Gaurav, and Gurpreet Singh. "Experimental invistigation of heat transfer characteristics in Al2O3-water based nanofluids operated shell and tube heat exchanger with air bubble injection." International Journal of Mechanical and Production, 7, 263 273 (2017).*

10. *Gungor B., Kaluc E., Taban E. (2014) Mechanical and microstructural properties of robotic Cold Metal Transfer (CMT) welded 5083-H111 and 6082-T651 aluminum alloys, Materials & Design, v. 54, pp. 207–211.*

11. *Gardner L.(2005) The use of stainless steel in structures. Progress in Structural Engineering and Materials, 7(2), 45-55.*

12. *Sun Z, Karppi R. (1996) The application of electron beam welding for the joining of dissimilar metals: an overview. Journal of Materials Processing Technology, 59(3), 257-67.*

13. *Rahman, M. T., et al. "Heavy metal contaminations in vegetables, soils and river water: A comprehensive study of Chilmari, Kurigram, Bangladesh." Int. J. Environ. Ecol. Fam. Urban Stud 5 (2015): 29-42.*

14. *Joseph A, Rai SK, Jayakumar T, Murugan N. (2005) Evaluation of residual stresses in dissimilar weld joints. International Journal of Pressure Vessels and Piping, 82(9), 700-5.*

15. *Jang C, Lee J, Kim JS, Jin TE. (2008), Mechanical property variation within Inconel 82/182 dissimilar metal weld between low alloy steel and 316 stainless steel. International Journal of Pressure Vessels and Piping, 85(9), 635-46.*

16. *Muransky O, Smith MC, Bendeich PJ, Edwards L. (2011), Validated numerical analysis of residual stresses in Safety Relief Valve (SRV) nozzle mock-ups. Computational Materials Science, 50(7), 2203-15.*

17. *Yan J.,Ming G., Xiaoyan Z. (2010), Study microstructure and mechanical properties of 304 stainless steel joints by TIG, laser, and laser-TIG hybrid welding, J Optics and Lasers in Eng 48, 512-517.*

18. *Adhoni, Shakeel Ahmed, Shanthanu M. Raikar, and C. T. Shivasharana. "Bioremediation of Industrial Effluents with Heavy Metals using Immobilised Microalgae." International Journal of Applied and Natural Sciences (IJANS) 7.5 (2018): 67-84.*

19. *Martin D.S., Castillo P.E.J, Peekstok E., Zwaag S. (2007), A new etching route for revealing the austenite grain boundaries in an 11.4% Cr precipitation hardening semi-austenitic stainless steel, Material Characteristics 58, 455–460.*

20. *Decroix JH. (1968), Deformation Under Hot Working Conditions. The Iron and Steel Institute, London.*

21. *Hanninen H, Romu J, IIola R, Tervo J, Laitinen A. (2001), Effects of processing and manufacturing of high nitrogen-containing stainless steels on their mechanical, corrosion and wear properties. J Mater Process Technol;117:424–30.*

22. *Ramkumar T, Selvakumar M, Narayanasamy P, Begam AA, Mathavan P, Raj AA.(2017), Studies on the structural property mechanical relationships and corrosion behavior of Inconel 718 and SS 316L dissimilar joints by TIG welding without using activated flux. J Manuf Process, 30:290–8.*

23. *A. Joseph, Sanjai K. Rai, T. Jayakumar, N. Murugan (2005), Evaluation of residual stresses in dissimilar weld joints. International Journal of Pressure Vessels and Piping, 82, 700–705.*

24. *Sathiya P, Aravindan S, Haq AN. (2005), Mechanical and metallurgical properties of friction welded AISI 304 austenitic stainless steel. The International Journal of Advanced Manufacturing Technology, 26(5-6), 505-11.*

25. *Kumar NP, Vendan SA, and Shanmugam NS. (2016), Investigations on the parametric effects of cold metal transfer process on the microstructural aspects in AA6061. J. Alloy Compd.,658, 255-264.*

26. *Varghese P, Vetrivendan E, Kamaraj M, Ningshen S and Mudali U K (2019), Weld overlay coating of Inconel 617M on type 316L stainless steel by cold metal transfer process Surf. Coat. Technol., 357-1004–13.*

27. *Sammaiah P, Suresh A, Tagore GR (2010), Mechanical properties of friction welded 6063 aluminum alloy and austenitic stainless steel. Journal of materials science 45(20), 5512-21.*

28. *Mishra RR, Tiwari VK, Rajesha S. (2014), A study of tensile strength of MIG and TIG welded dissimilar joints of mild steel and stainless steel. International Journal of Advances in Materials Science and Engineering, 3(2), 23-32.*

29. *Dupont J.N,Kusko C.S (2007), Technical Note: Martensite Formation in Austenitic/Ferritic Dissimilar Alloy Welds, Welding Journal, 51s-54s.*

30. *Cai M., Wu C., and Gao X (2018), IOP Conf. Ser.: Earth Environ. Sci. 170, 042106.*

31. *William N.T., Parker J.D. (2005), Int. Mater. Rev., 49, 45– 75.*

32. *Zhang H., Senkara J. (2005), Resistance welding: fundamentals and applications, Boca Raton FL, CRC Press, 196–201.*

33. *Okumura T, Taniguchi C (1982), Engenharia de Soldagem e Aplicacoes. Rio de Janeiro: LTC*

34. *Withers PJ, Bhadeshia HKDH. (2001), Residual stress. Part 1 - Measurement techniques. Materials Science and Technology, 17(4):355-365.*

35. *Wan Y, Jiang W, Li J, Sun G, Kim DK, Woo W (2017) Weld residual stresses in a thick plate considering back chipping: Neutron diffraction, contour method and finite element simulation study. Materials Science and Engineering, 699:62-70.*

36. *Magalhaes RR, Vieira AB Jr., Barra SR. (2013), The use of conventional strain gauges evaluation for measurements of residual stresses in welded joints. Journal of the Brazilian Society of MechanicalSciences and Engineering,36(1):173-180.*

37. *Lee C.H., Chang K.H. (2012), Temperature fields and residual stress distributions in dissimilar steel butt welds between carbon and stainless steels, Applied Thermal Engineering, S- 45–46:33–41*

38. *Withers P.J., Turski M., Edwards L. (2008),Recent advances I residual stress measurement Int. J. Press. Vessel. Pip., 85 (3), pp. 118-127*

39. *Koli Y., Yuvaraj N., Vipin, Aravindan S. (2019), Investigation on weld wead geometry and microstructure in CMT, MIG pulse synergic and Mig welding of AA6061-T6, MRX2-103311.R1*

40. *Shanmugasundar G., Karthikeyan B., Ponvell P.S., Vignesh V. (2019), Optimization of Process Parameters in TIG Welded Joints of AISI 304L -Austenitic Stainless Steel using Taguchi's Experimental Design Method, 16 1188–1195*

41. Kinsey B., Liu Z., Cao J. (2000), A Novel forming technology for tailor-welded blanks Journal of Materials Processing Technology, Vol 99, Issues 1–3, 1, Pages 145-153

42. Purnama D., Oktaditana H. (2019),Effect of Shielding Gas and Filler Metal to Microstructure of Dissimilar Welded Joint Between Austenitic Stainless Steel and Low Carbon Steel, Material science and engineering, 547-012003

43. Prakash A., Mangal D (2017), Study and Assessment Of Mechanical Properties Of Resistance Spot Weld Of Two Dissimilar Metals, 6(10) 2277-9655.

44. Kannan R., Shanmugam N.S., Naveen S (2019),Effect of Arc Length Correction on Weld Bead Geometry and Mechanical Properties of AISI 316L Weld ments by Cold Metal Transfer (CMT) Process ,18-3916–3921

45. Hasanbasoglu, A., &Kacar, R. (2007). Resistance Spot Weldability of Dissimilar Materials (AISI 316L-DIN EN 10130-99 Steels). Material and Design, 28, 1794-1800.

46. Marashi S.P.H., Pouranvari M., Salehi1 M., Abedi A., Kaviani S.(2008), Overload failure behaviour of dissimilar thickness resistance spot welds during tensile shear test, Mater. Sci. Eng. A480, 175–180.

47. Darwish S.M., Al-Samhan A.M.(2004), J. Mater. Process. Technol., 147, 51–59.

48. Sabooni S., Karimzadeh F., Enayati M.H., Ngan A.H.W (2015), Friction-stir welding of ultrafine grained austenitic 304L stainless steel produced by martensitic thermomechanical processing, Mater. Des. 76, 130–140

49. Keshari R.K., Sahu P.L.(2019) Mechanical characterization of dissimilar welded joint of SS202 and SS304 by tungsten inert gas welding, Vol-3, Issue-4, 245-256

50. Wan Y, Jiang W, Li J, Sun G, Kim DK, Woo W. (2017), Weld residual stresses in a thick plate considering back chipping: Neutron diffraction, contour method and finite element simulation study. Materials Science and Engineering, A 699, 62-70

51. Gupta A. (2019),Determination of residual stresses for helical compression spring through Debye-Scherrer ring method, Materials Today: Proceedings

Medical Image Denoising using Convolutional Autoencoder with Shortcut Connections

Manas Gupta
Dept of Computer Science and Engineering
Delhi Technological University
New Delhi, India
gmanas813@gmail.com

Kushal Goel
Dept of Computer Science and Engineering
Delhi Technological University
New Delhi, India
kushal8601@gmail.com

Jatin Kansal
Dept of Computer Science and Engineering
Delhi Technological University
New Delhi, India
jatinkansal81@gmail.com

Anurag Goel
Dept of Computer Science and Engineering
Delhi Technological University
New Delhi, India
anurag@dtu.ac.in

2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT) | 978-1-6654-7467-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSSIT55814.2023.10061131

***Abstract*—In recent times, medical image analysis has gained immense attention among researchers for diagnosing and treating deadly diseases, e.g., Cancer. Denoising of Images is considered as a crucial process in medical image analysis. Several Deep Learning techniques are effectively used for image denoising including Autoencoders. In this paper, a convolutional autoencoders based approach with shortcut connections is proposed for medical image denoising. The proposed approach is evaluated by using three medical images datasets. The results demonstrated that the proposed approach outperforms the current cutting-edge methods of medical image denoising on all the three datasets.**

Keywords*—*Medical Image Denoising, Convolutional Neural Network, Convolutional Autoencoders, Shortcut Connections.

I. INTRODUCTION

Noise refers to the distortions in an image. It can be in the form of brightness variations or color information in an image. Denoising is the technique used to eradicate these distortions.

Various techniques, including Ultrasound Imaging, Radiography, etc., are used to view the internal structure of the human body to diagnose various medical conditions. Such technologies are categorized as Medical Imaging and are vulnerable to noise. The vulnerability could arise from using distinct image acquisition methods to reduce patient exposure to radiation. The more the radiation exposure is reduced, the more the noise is increased. Other reasons could also result in a noisy image, like bit errors in transmission, statistical quantum fluctuation, etc. For medical analysis, medical image denoising is usually required to diagnose and treat diseases properly.

Being a classical problem, image denoising remains a popular topic of research. There are different image denoising approaches, such as Domain Transformation based models, Partial Differential Equation based models. The equation for the problem would be:

$$z = x + y \tag{1}$$

Here, x refers to the original image, y is the noise, and z is the image generated after the addition of noise in the original image. It is assumed that noise is generated from a distinct process, and using this assumption, y is calculated. Most of the models try to approximate x using z.

With the evolution of deep learning, the deep learning models perform better than the conventional denoising approaches. In deep learning models, Convolutional autoencoders are majorly implemented for image denoising task. Convolutional autoencoders consist of encoder-decoder network, which is built using Convolutional Neural Networks. An image denoising technique is used for medical dataset based on Convolutional autoencoders with shortcut connections.

The paper is divided into a number of sections. Related Work study is described in Section II. The planned approach is illustrated in Section III. The results of the work is shown in Section IV while in Section V conclusion is stated.

II. RELATED WORK

In [1], a review of various Convolutional Neural Network (CNN) based approaches for image denoising have been presented. In [1], it was shown that the CNN-based approaches gave a good performance. CNN models are designed with the information of noise in the dataset. In [2], the correlation between medical image denoising and medical image classification is explored using DenseNet-121 and CNN models.

In [2], it was concluded that image denoising significantly affects the performance of image classification. The approach proposed in [3] uses CNN for image denoising. It was found that performance can be equivalent to or higher than the methods established upon wavelets.

Autoencoder encodes an input into a latent space representation using the encoder network, and then the latent space reprsentation is converted back to the original input utilizing the decoder network. Through training, the model develops the ability to map various input images to precise locations in the latent space [4].

Vincent et al. [5] proposed Denoising Autoencoders, which are an extension to the traditional autoencoders, helping reconstruction of the input by introducing a noisy version. Later, the same authors proposed a stacked version of denoising autoencoders by stacking denoising autoencoders one over the other [6]. A Denoising Autoencoder based architecture is also proposed by Robinet et al. [7] for image denoising.

Convolutional encoding and decoding layers are added to a typical autoencoder architecture to create Convolutional autoencoders, proposed by Jonathan et al. [8]. Convolutional autoencoders are considered to be better for image processing as compared to classic autoencoders because they utilize the power of CNNs to exploit the contextual information in the image.

Lovedeep et al. [9] proposed Convolutional Neural Network based Denoising AutoEncoders (CNN DAE) and have shown that CNN DAE outperformed the median filter in terms of denoising performance on small datasets. It was suggested that these techniques could recover signals even in situations with high noise levels when the majority of denoising techniques would fall short. However, this straightforward network has problems recreating the original signal when the noise level is relatively high. However, this network is successful at partially generating actual images even when they are invisible to the human eye.

Prashanth et al. [10] have shown the effectiveness of Convolutional Autoencoder for image denoising. In Convolutional Autoencoders, convolutional layers are used instead of dense layers in standard autoencoders. In [10], it was demonstrated that the completely connected Autoencoder, that just uses dense layers, is not efficient to denoise the input; while Convolutional autoencoders output a nearly noise-free image. This is because, when mapping an image to a latent space, the convolutional layers extract and preserve the essential input features and eliminate noise.

From the current state-of-the-art technologies, it is observed that there is a need to reduce overfitting of the model. Deeper architectures sometimes prove to be less efficient because of memorizing the training data.

III. PROPOSED APPROACH

The approach suggested is based on Convolutional AutoEncoder with shortcut connections. Shortcut connections boost the performance of the model by reducing the depth of the architecture of the model. This helps in eliminating the problem of over training the model as well. The representation of a shortcut connection is shown in Fig. 1.

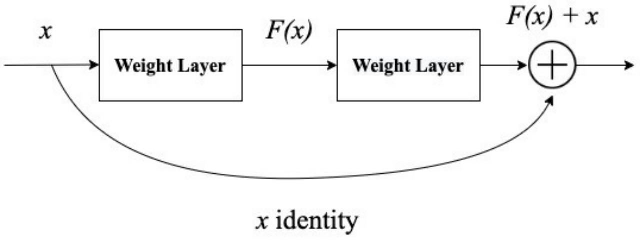


Fig. 1. Representation of Shortcut Connections

Fig. 2 illustrates the architecture of the proposed approach. The architecture includes an encoder network which is built using Convolutional layers and a decoder network consisting of Deconvolutional layers. Shortcut connections are introduced in the architecture. The goal of introducing the shortcut connections is to pass the features of an image from one level to another without interfering with in-between layers.

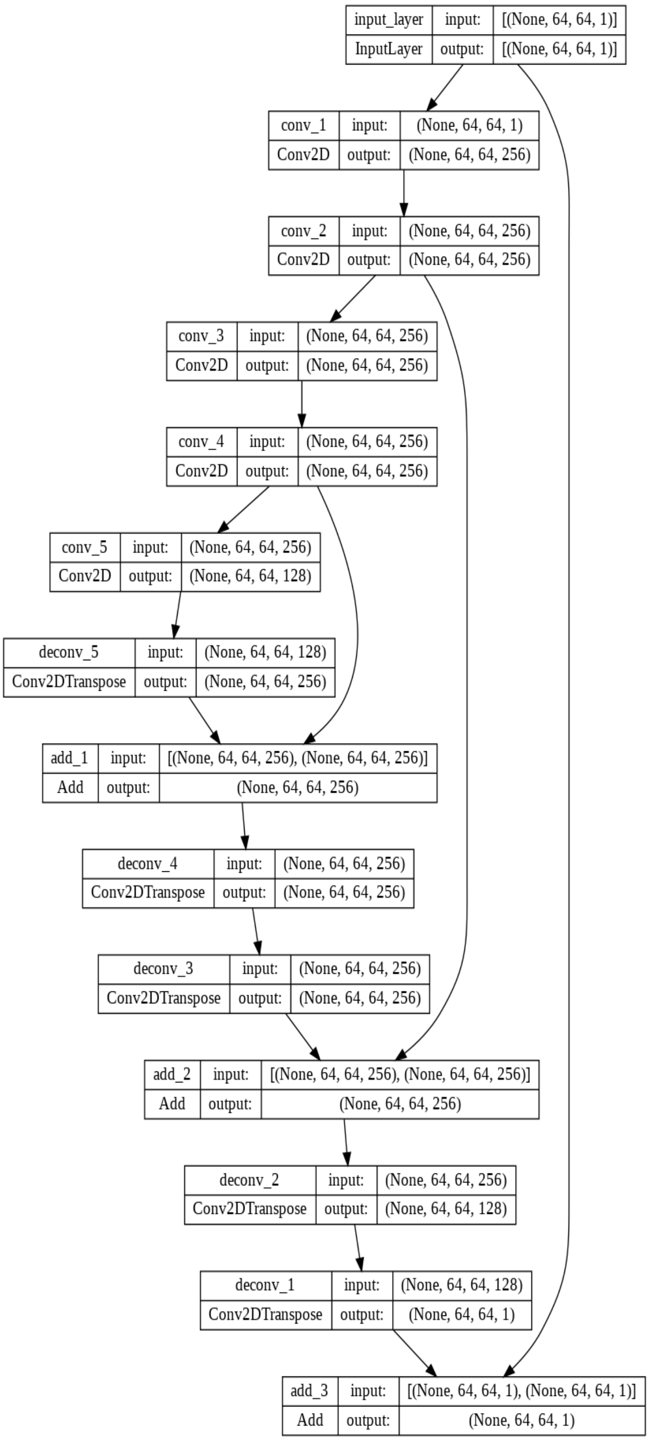

input_layer | input: | [(None, 64, 64, 1)]
InputLayer | output: | [(None, 64, 64, 1)]
conv_1 | input: | (None, 64, 64, 1)
Conv2D | output: | (None, 64, 64, 256)
conv_2 | input: | (None, 64, 64, 256)
Conv2D | output: | (None, 64, 64, 256)
conv_3 | input: | (None, 64, 64, 256)
Conv2D | output: | (None, 64, 64, 256)
conv_4 | input: | (None, 64, 64, 256)
Conv2D | output: | (None, 64, 64, 256)
conv_5 | input: | (None, 64, 64, 256)
Conv2D | output: | (None, 64, 64, 128)
deconv_5 | input: | (None, 64, 64, 128)
Conv2DTranspose | output: | (None, 64, 64, 256)
add_1 | input: | [(None, 64, 64, 256), (None, 64, 64, 256)]
Add | output: | (None, 64, 64, 256)
deconv_4 | input: | (None, 64, 64, 256)
Conv2DTranspose | output: | (None, 64, 64, 256)
deconv_3 | input: | (None, 64, 64, 256)
Conv2DTranspose | output: | (None, 64, 64, 256)
add_2 | input: | [(None, 64, 64, 256), (None, 64, 64, 256)]
Add | output: | (None, 64, 64, 256)
deconv_2 | input: | (None, 64, 64, 256)
Conv2DTranspose | output: | (None, 64, 64, 128)
deconv_1 | input: | (None, 64, 64, 128)
Conv2DTranspose | output: | (None, 64, 64, 1)
add_3 | input: | [(None, 64, 64, 1), (None, 64, 64, 1)]
Add | output: | (None, 64, 64, 1)

Fig. 2. Block Diagram Architecture of Proposed Approach

The proposed model takes an input image and fed it to a number of convolutional layers and transposed convolutional layers. The convolutional layers bring out the essential characteristics of the image and also help in removing noise from images. The kernel filters in the convolution layers processed over the input image to extract features and create a feature map. Deeper layer extracts more complex features from images.These feature maps store essential characteristics of the image needed to reconstruct it to the original form. In the proposed model, five convolutional layers have been implemented in which the first four layers use 256 convolutional kernel filters while the last layer uses 128 convolutional kernel filters. The proposed model also consists of five transposed convolutional layers (deconvolutional layers) used for the purpose of rebuilding the original image. The final output produced is a clean and clear image without noisebecause the shortcut connections eliminates the problem of vanishing gradient and deeper architecture.Also early stopping is used to avoid overfitting

The shortcut connections are implemented in between the deconvolutional layers. The shortcut connections are implemented with the help of add layer which adds the output from one layer to another layer directly skipping the layers in between.

In the proposed model architecture, pooling layers are not used because pooling may remove the essential information which becomes critical in the study of medical dataset. Using the shortcut connections, the feature maps extracted by the convolutional layers are passed directly to the deconvolutional layers by skipping some layers in between. This solves the problem of overfitting caused by the deep layer architecture of the model.

IV. EXPERIMENTS AND RESULTS

A. Datasets Used

As suggested in [14], large scale versatile images prove better than small scale similar images. The following datasets are used for performing experiments:

1) Chest X-Ray Dataset: In this dataset, there are total 247 Chest X-ray images [11]. The training set and the testing set conisists of 197 images and 50 images respectively. A sample image is shown in Fig. 3.

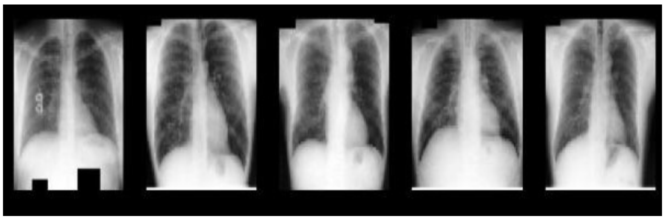


Fig. 3. Sample Image of Chest X-Ray Dataset

2) Dental X-Ray Dataset: In this dataset, there are total 120 images [12]. The training set and the testing set consists of 96 images and 24 images respectively. A sample image is shown in Fig. 4.

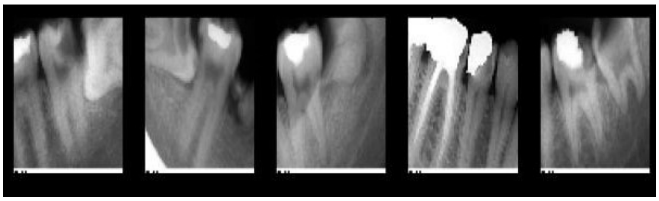


Fig. 4. Sample Image of Dental X-Ray Dataset

3) Covid CT Dataset: In this dataset, there are total 349 images of Computed Tomography (CT) scans of Covid patients as well as non-Covid patients [13]. For the training set and the testing set, 279 images and 70 images are used respectively. A sample image of this dataset is shown in Fig. 5.

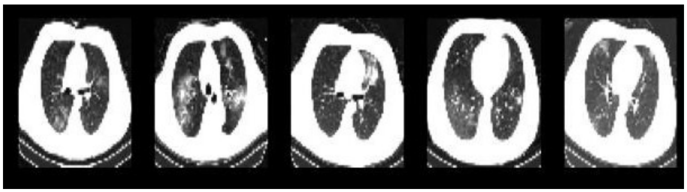


Fig. 3. Sample Image of Covid CT Dataset

B. Experimental Setup

The experiments have been performed on an Intel i5 processor running Ubuntu. The code has been written in Python language. For pre-processing, the images are loaded in grayscale mode with a target size of (64,64) and Gaussian noise (7%) is added to images due to the fact as stated in [15], Gaussian noise is additive in which shows contribution of Gaussian noise in each pixel. The images are used in the form of array with normalized values.

Number of epochs for which model is trained is 20 with batch size of 10.Early stopping is also used in the model for better performance.

C. Metrics Used

The results of the approach planned is matched with the current methods of denoising of medical dataset using the following metrics:

1) Peak Signal To Noise Ratio (PSNR): This evaluation metric is the ratio between the maximum power of the signal and the power of noise signal that affects reconstruced image quality. If the PSNR is high, reconstructed image would be better. The formula is given as follows:-

$$PSNR = 20\log_{10}\left(\frac{MAX_f}{\sqrt{MSE}}\right) \tag{2}$$

where MSE(Mean Squared Error) is calculated as follows:-

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}||f(i,j) - g(i,j)||^2 \tag{3}$$

where,

- f represents the data matrix of noiseless image;
- g represents the data matrix of noisy image;
- m represents the total number of rows of pixels in the image;
- n represents the total number of columns of pixels in the image;
- MAX_f is the maximum signal value exists in the noiseless image.

2) Structural Similarity Index Measure (SSIM): It is one of the evaluation metrics which measures the quality of the reconstructed image with that of the reference image. Hence, higher would be the SSIM score, better would be the quality of the reconstructed image relating to structural similarity with reference image. The formula is given as follows:-

$$SSIM = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \tag{4}$$

where:-

- μ_x is the mean of x;
- μ_y is the mean of y;
- σ_x^2 is the variance of x;

- σ_y^2 is the variance of y;
- σ_{xy} is the covariance of x and y;
- c$_1$=$(k_1L)^2$, c$_2$=$(k_2L)^2$ are to stabilize the division with weak denominator;
- L is the dynamic range of the pixel values;
- k_1 =0.01 and k_2 =0.03 (default values) are constants.

D. Results Analysis

Table I shows the SSIM scores and Table II depicts the PSNR scores. From the results, it can be interpreted that the proposed model outperforms the benchmarks on all the datasets by a significant margin.

TABLE I. SSIM SCORES

	CNN DAE [9]	CAE [10]	Proposed
Chest X-Ray Images Dataset	0.92	0.75	**0.93**
Dental X-Ray Images Dataset	0.86	0.66	**0.89**
Covid CT Dataset	0.74	0.70	**0.87**

TABLE II. PSNR SCORES

	CNN DAE [9]	CAE [10]	Proposed
Chest X-Ray Images Dataset	51.17	46.48	**52.51**
Dental X-Ray Images Dataset	49.35	45.61	**51.72**
Covid CT Dataset	43.44	42.55	**48.27**

For the Chest X-Ray dataset [11], Fig. 6, Fig. 7 and Fig. 8 show the results of CNN DAE model [9], CAE model [10] and the proposed model respectively. In each figure, the first row shows the noiseless image, the second row shows the noisy version of the corresponding first row image, and third row shows the output of the respective model on the corresponding first row image.

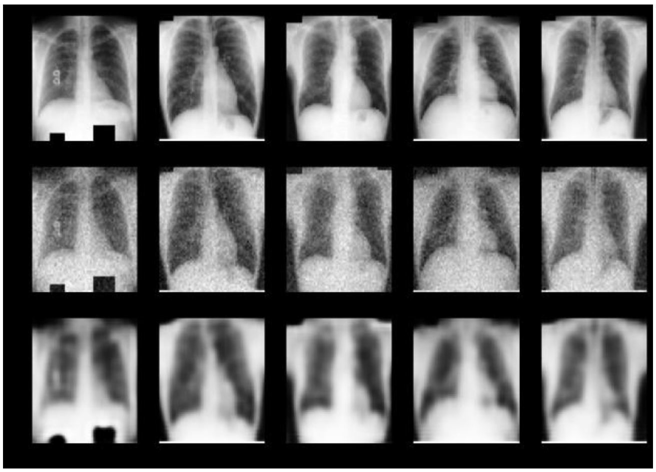


Fig. 6. Results of CNN DAE (First Row: Noiseless Image, Second Row: Noisy Image and Third Row: Output of CNN DAE)

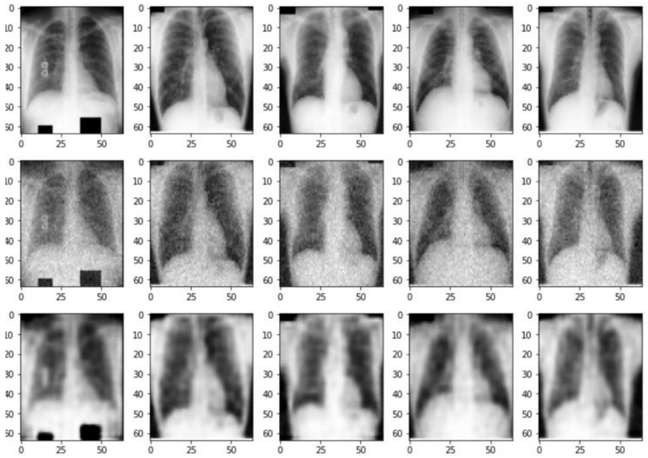


Fig. 7. Results of CAE (First Row: Noiseless Image, Second Row: Noisy Image and Third Row: Output of CAE)

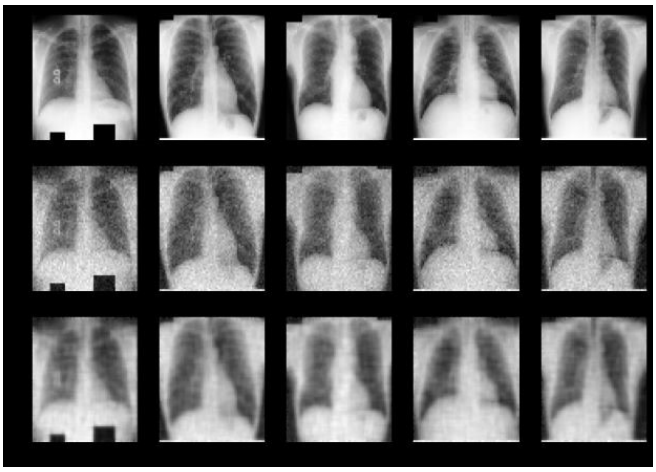


Fig. 8. Results of Proposed Model (First Row: Noiseless Image, Second Row: Noisy Image and Third Row: Output of Proposed model)

V. CONCLUSION AND FUTURE WORK

Medical image denoising is considered extremely crucial in medical image analysis. In this work, on a Convolutional Autoencoder based approach with Shortcut connections has been proposed for medical image denoising. The shortcut connections pass the features from one level to another by skipping the in-between layers. For the experiments, three medical images datasets have been used. The proposed approach is compared with several benchmark models of medical image denoising. The results demonstrated that the proposed approach outperforms the benchmarks by a significant margin on all the datasets.

Autoencoders have limitations of overfitting in case of large dataset due to their dual-network architecture of encoder and decoder. Since, the proposed model is based on Convolutional Autoencoder, it also faces the limitations of overfitting due to the high number of learning parameters of encoder as well as decoder. This limitation of dual-network architecture can be explored in future.

REFERENCES

[1] Ilesanmi, Ademola E., and Taiwo O. Ilesanmi, "Methods for image denoising using convolutional neural network: a review." Complex & Intelligent Systems 7, no. 5 (2021): 2179-2198.

[2] Michael, Peter F., and Hong-Jun Yoon, "Survey of image denoising methods for medical image classification.", In Medical Imaging 2020: Computer-Aided Diagnosis, vol. 11314, pp. 892-899. SPIE, 2020.

[3] Jain, Viren, and Sebastian Seung. "Natural image denoising with convolutional networks." Advances in neural information processing systems 21 (2008).

[4] Baldi, Pierre. "Autoencoders, unsupervised learning, and deep architectures." In Proceedings of ICML workshop on unsupervised and transfer learning, pp. 37-49. JMLR Workshop and Conference Proceedings, 2012.

[5] Vincent, Pascal, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. "Extracting and composing robust features with denoising autoencoders." In Proceedings of the 25th international conference on Machine learning, pp. 1096-1103. 2008.

[6] Vincent, Pascal, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, Pierre-Antoine Manzagol, and Léon Bottou. "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion." Journal of machine learning research 11, no. 12 (2010).

[7] L. Robinet. (2020), “Autoencoders and the denoising feature: From theory to practice,” [Online]. Available: https://towardsdatascience.com/autoencoders-and-the-denoising-feature-from-theory-to-practice-db7f7ad8fc78

[8] Masci, Jonathan, Ueli Meier, Dan Cireşan, and Jürgen Schmidhuber. "Stacked convolutional auto-encoders for hierarchical feature extraction." In International conference on artificial neural networks, pp. 52-59. Springer, Berlin, Heidelberg, 2011.

[9] Gondara, Lovedeep. "Medical image denoising using convolutional denoising autoencoders." In 2016 IEEE 16th international conference on data mining workshops (ICDMW), pp. 241-246. IEEE, 2016.

[10] Venkataraman, Prashanth. "Image Denoising Using Convolutional Autoencoder." arXiv preprint arXiv:2207.11771 (2022).

[11] Shiraishi, Junji, Shigehiko Katsuragawa, Junpei Ikezoe, Tsuneo Matsumoto, Takeshi Kobayashi, Ken-ichi Komatsu, Mitate Matsui, Hiroshi Fujita, Yoshie Kodera, and Kunio Doi. "Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules." American Journal of Roentgenology 174, no. 1 (2000): 71-74.

[12] Dental X-Ray Images Dataset: https://www.kaggle.com/datasets/parthplc/medical-image-dataset

[13] Yang, Xingyi, Xuehai He, Jinyu Zhao, Yichen Zhang, Shanghang Zhang, and Pengtao Xie. "COVID-CT-dataset: a CT scan dataset about COVID-19." arXiv preprint arXiv:2003.13865 (2020).

[14] Sharif, S. M. A., Rizwan Ali Naqvi, and Mithun Biswas. "Learning medical image denoising with deep dynamic residual attention network." Mathematics 8, no. 12 (2020): 2192.

[15] Verma, Rohit, and Jahid Ali. "A comparative study of various types of image noise and efficient noise removal techniques." International Journal of advanced research in computer science and software engineering 3, no. 10 (2013).

Circuits, Systems, and Signal Processing
https://doi.org/10.1007/s00034-023-02346-x

Memristor-Based Architectures for PFSCL Circuit Realizations

Shikha[1,2] · Neeta Pandey[1] · Kirti Gupta[2]

Received: 30 April 2022 / Revised: 4 March 2023 / Accepted: 4 March 2023

Abstract

Positive feedback source-coupled logic (PFSCL) circuits are imperative in high-resolution mixed-signal applications. In this paper, memristors are introduced in PFSCL style to amalgamate their advantages in circuit design. Two new memristor-based PFSCL architectures are presented. The first proposed architecture (Mem-PA1) employs memristors with a PFSCL inverter, while the other uses them with PFSCL NOR/OR gate (Mem-PA2). The proposed circuits eliminate the restriction of implementing the logic function in NOR/OR form or sum of two minterms in existing architectures. The reduced gate count gives significant advantage in power and propagation delay values. Different functions are implemented and simulated using CMOS PTM 180 nm technology parameters and a TiO_2 memristor. A maximum improvement of 33%, 80% and 86% in propagation delay, power and power delay product, respectively, is observed for Mem-PA1 exclusive-OR(XOR3) gate with respect to existing architectures-based gates. Further, it is observed that Mem-PA2 XOR3 gate shows 5.16% improvement in delay over the Mem-PA1 gate for same value of power consumption. Lastly, the performance is also examined with technology scaling and at various process corners.

Keywords Mixed-signal design · Digital circuits · PFSCL · Fundamental cell · Memristor

✉ Kirti Gupta
kirtigupta22@gmail.com

Shikha
ishushikha234@gmail.com

Neeta Pandey
n66pandey@rediffmail.com

[1] Department of Electronics and Communication Engineering, Delhi Technological University, Delhi 110042, India

[2] Department of Electronics and Communication Engineering, Bharati Vidyapeeth's College of Engineering, New Delhi 110063, India

Published online: 25 March 2023

1 Introduction

Nowadays, many systems demand unification of analog and digital operations in high-resolution mixed-signal applications. The integration of both kinds of circuitry on a single chip has led to numerable advantages, such as improvement in the system reliability, reduction in cost, power consumption as well as the size of the system. The design for such integrated circuits is challenging as the functionality of analog subsystem is severely limited due to switching noise generated in digital subsystem implemented using popular and acclaimed complementary MOS (CMOS) logic family [6, 15, 17, 22, 30]. Further, the power consumption increases proportionately with rising operating frequency. Therefore, researchers have explored alternate logic styles to address these issues at different levels of abstraction such as placement and routing, layout and logic. The positive feedback source-coupled logic (PFSCL) style is a variant among the alternative logic families that work on current steering principle. The PFSCL gates have constant current source such that the current drawn from the power supply remains nearly constant, thereby reducing switching noise significantly and making the power consumption independent of operating frequency [4, 5, 16, 19, 36].

The conventional PFSCL style is based on implementing NOR/OR realization of a logic function by employing PFSCL NOR/OR gates [4]. It, however, infers high gate and current source count for complex function realization which mitigates high-speed advantage of PFSCL style. In an effort to reduce the number of stages, a PFSCL fundamental cell as a circuit element has been introduced in [13, 14] which facilitates sum-of-minterms realization in a single gate. Despite being an attractive concept, it, however, requires larger footprint due to wider control transistors as well as restricted two minterms forms in a single fundamental cell. Recently, memristor, a new element, has emerged as an eye of designers with growing opportunities to explore and develop new avenues as research in the field is still in nascent stage [2, 3, 7, 8, 10, 11, 18, 20, 21, 23–29, 31, 33–35, 39, 41–44]. Due to its inherent features of smaller footprint and compatibility with CMOS technology, the memristor has been used in applications such as neural networks [10, 23, 24, 39], non-volatile static random-access memory [18, 20, 26, 29, 31, 33, 35], combinational circuits [7, 25, 28, 34, 42–44], analog signal processing and generation circuits [21], etc. The memristor-based logic circuits may be broadly viewed as state-based logic circuits and conventional voltage-based logic circuits. The former is based on storing the output as a resistance state in logic styles such as material implication logic (IMPLY) [26], memristor-aided logic (MAGIC) [25] and scouting logic [41]. It requires peripheral circuits to read memristor state and provide output, thus mitigating its benefits. In conventional voltage-based logic, the logic level is defined by voltage level. The memristor ratio logic (MRL) [27] and memristor CMOS (MCM) [8] integrate memristor with CMOS technology and are benefited from easier translation using design abstractions. An interesting logic style to provide OR and AND functions is suggested in [11]. Subsequently, it is extended by putting a CMOS inverter to provide universal NOR and NAND gates and is named as MRL [27]. Another variant of MRL is presented in [8] that realizes AND–OR–INVERT function by memristor-based AND gate which is followed by a resistive load-based NOR gate. Here nanowire resistor is used. In [34] the memristors are combined with only NMOS transistors for designing priority encoders, decoders and code converters. The MRL-based crossbar (X-MRL) design is yet another variant wherein conventional CMOS logic realized with MRL and mapped into crossbar structure is proposed [3]. The concept is illustrated with help of one-bit full adder. In an effort to reduce the number of memristors, a modified MRL is proposed in [2]. It uses a single memristor along with one/two transistors to design NOT, AND, and NOR gates. Compared with the CMOS logic, the memristor-based logic circuits not only reduce the number of transistors, but also greatly reduce the power consumption and improve the operating speed of the whole circuit due to the reduction of associated capacitances. Considering the benefits of PFSCL over CMOS logic in mixed-signal applications, an amalgamation of memristor and PFSCL is put forward in this paper. Two memristor-based PFSCL architectures with the objective to reduce gate and current source count for complex function realization and simultaneously overcoming the restrictions of limited (two) minterms in the existing PFSCL architectures are proposed in this work.

The paper is organized in five sections including the introduction section. An overview to existing PFSCL architectures along with a brief introduction to memristor is presented in Sect. 2. The memristor is then introduced in PFSCL style, and two architectures are suggested in Sect. 3—the former uses a cascade of memristor-based logic block followed by a PFSCL inverter, while the latter embeds memristor-based logic block in PFSCL NOR/OR gate. The mathematical formulation for computing output voltages of basic gates is put forward which can be extended to other gates as well. Section 4 details the functional verification using 180 nm CMOS technology parameters and the memristor models suggested in [7] and [38]. This section also includes performance comparison of proposed architectures with their existing PFSCL counterparts. The conclusion is drawn in 5 section.

2 Existing Work

In this section, the existing works related to PFSCL circuit design are presented which is followed by a brief description on memristor.

2.1 PFSCL Style

A PFSCL gate belongs to the single-ended source-coupled logic family wherein the input transistors are source-coupled to a transistor with gate output as feedback to it as shown in Fig. 1a [4]. The presence of positive feedback improves the switching speed of the gate in comparison to the conventional single-ended source-coupled logic [4]. A generic N-input PFSCL NOR gate shown in Fig. 1a comprises three basic components, namely pull-down network (PDN), current source and load transistors. The logic function is realized in PDN wherein depending on the inputs, the bias current (I_{SS}) is steered in one of the two source-coupled branches. The load transistors then perform the current-to-voltage conversion to produce the required output voltage level. It is pertinent to mention here that a generic N input PFSCL gate of Fig. 1a realizes NOR functionality. Further, OR functionality can be obtained from the same gate by

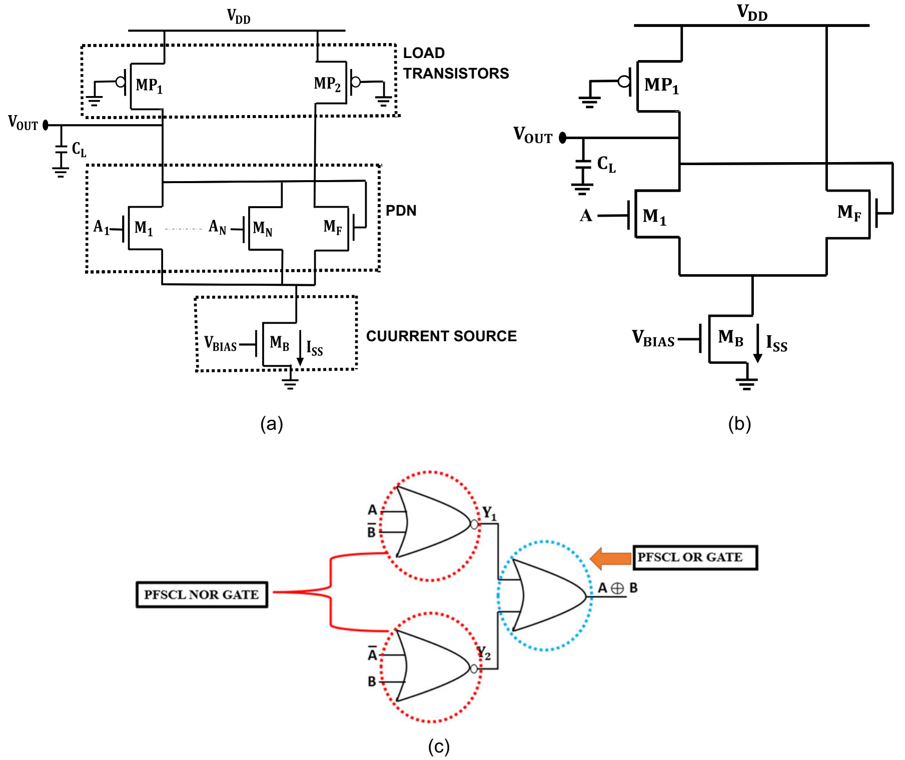


Fig. 1 **a** Generic N-input PFSCL NOR gate [4] **b** Conventional PFSCL inverter [4] **c** Gate level schematic for NOR/OR-based realization PFSCL of XOR2 gate

obtaining the output from the drain terminal of MP$_2$. A PFSCL inverter is drawn in Fig. 1b to illustrate the operation [4]. For a low value of input A, $\mathrm{I_{SS}}$ gets steered from $\mathrm{M_F}$ such that there exists no voltage drop across load transistor MP$_1$; therefore, a high value is obtained at the output ($V_{OH} = V_{DD}$). Alternatively, for high value of input A, I_{SS} is steered through M_1 and a low output voltage is obtained at the output ($V_{OL} = V_{DD} - V_{SWING}$), where V_{SWING} corresponds to the voltage drop across load MP$_1$ and is referred to as the voltage swing. Thus, PFSCL style is a reduced voltage swing logic and has static power consumption due to the flow of constant current from power supply throughout the operation.

The generic PFSCL gate realizes a function in a NOR/OR representation, so the functions requiring sum of minterms or AND–OR representations infer cascading of multiple generic PFSCL gates. This can be attributed to the basic nature of PFSCL style of involving a single level of source-coupled transistors in the PDN. A NOR/OR-based two-input exclusive-OR (XOR2) gate realization is shown in Fig. 1c for illustration. It may be noted that three PFSCL NOR/OR gates are required for implementation such that the total power consumption would be thrice of a single gate realization (= 3 times $V_{DD}I_{SS}$).

An alternate method for realizing sum of minterms is suggested in [13, 14] which employs a PFSCL fundamental cell (FC) comprising two triple-tail cells (TTCs). This method facilitates realization of minterms in a single gate. Figure 2a shows the circuit diagram of a generic PFSCL FC. It has two TTCs (TTC$_1$: M_1, M_2, M_{F1}) and (TTC$_2$: M_3, M_4, M_{F2}) which are biased by two separate current sources of $I_{\mathrm{SS}}/2$ in order to maintain the total current drawn by the single gate as I_{SS} [14]. At any given instance, one-third of the TTC is deactivated by turning its respective control transistor in the middle branch ON, while the output voltage level is decided by the activated TTC. The symbolic representation of a FC is shown in Fig. 2b for simplification. In general, the output of a FC is obtained by combining any one of the two output nodes of TTC cell. For instance, the output Q_1 and Q_4 of TTC$_1$ and TTC$_2$ can be combined to realize an XOR function by connecting A at both the inputs X$_1$ and X$_2$ and X$_3$ as B as shown in Fig. 2c [14].

$$V(\mathrm{FC}-\mathrm{XOR2}) = \overline{A}.B + A.\overline{B} \tag{1}$$

Thus, a single PFSCL gate realizes XOR2 gate functionality in contrast to three PFSCL NOR/OR gates of existing NOR/OR-based architecture (Fig. 1c), leading to significant savings in power consumption. The gates combining NOR/OR gates and FC are also used in system design [12, 37]. It is to be noted that a design having N number of PFSCL gates would employ N current sources leading to N times increase

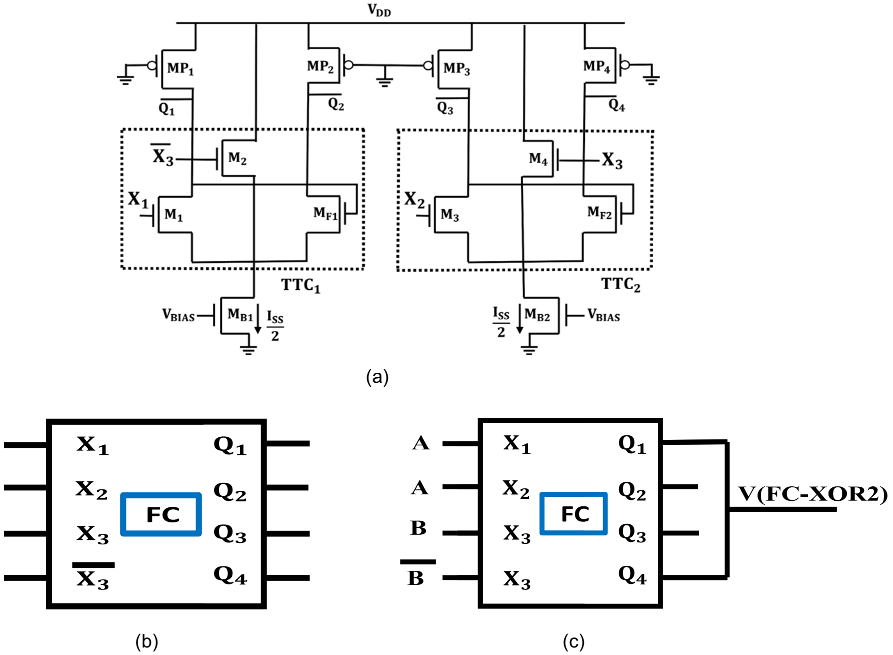


Fig. 2 **a** MOS schematic of generic PFSCL FC [13] **b** Symbolic representation **c** FC-based XOR2 gate implementation

in power consumption in comparison to a single gate. Additionally, TTC in FC has a wider control transistor in the middle branch for proper activation/deactivation and is restricted to two sum-of-minterms forms in a single fundamental cell. Thus, these issues must be dealt in and an efficient method that addresses the above issues may be useful.

2.2 Basic Concept of Memristor

Memristor is a two-terminal passive element which stores the data in terms of resistance. It was postulated by Leon Chua in 1971 as a fourth circuit element, besides the other three elements, namely, resistor, inductor and capacitor [9]. It remained a hypothetical component for a substantial amount of time till William and his co-worker from HP laboratories introduced a memristor physically in 2008 [1]. As shown in Fig. 3a, a memristor has a titanium dioxide (TiO_2) layer sandwiched between platinum (Pt) electrodes. The doped part of the TiO_2 layer is oxygen-deficient, i.e., TiO_{2-x}, and acts as a conductor, whereas the undoped region is solely an insulator (TiO_2). Figure 3b shows a symbolic representation of a memristor, with the shaded portion representing the undoped region and the other end representing the doped region.

When a positive voltage is applied across the doped side of the memristor, the oxygen deficiencies of the doped TiO_{2-x} layer will repel toward the undoped TiO_2 layer. It increases the width of TiO_{2-x} layer resulting in lowering of memristor resistance. In this condition, the memristor is defined to be in low resistance state (LRS) with resistance as R_{ON}. Conversely, if a negative voltage is applied across the memristor, it attracts oxygen deficiencies, resulting in an increase in the width of the undoped TiO_2 layer. The high resistance condition is the name given to this memristor state (HRS) with resistance referred to as R_{OFF}. The oxygen deficits will remain unchanged if no voltage is placed across the memristor; this is how the memristor retains its former condition. If voltage $v_m(t)$ is applied across the memristor, then $v_m(t)$ can be expressed as:

$$v_m(t) = M(t) i_m(t) \tag{2}$$

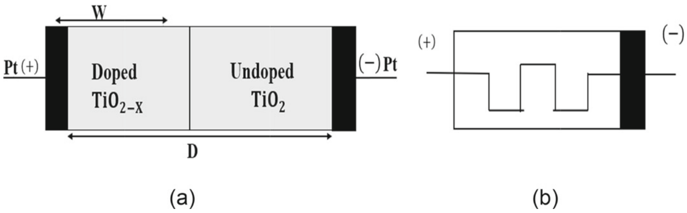


Fig. 3 Representation of memristor **a** Physical **b** Symbolic

where $i_m(t)$ is the current passing through the memristor and $M(t)$ stands for memristance, or memristor resistance, written as:

$$M(t) = R_{\mathrm{ON}}\frac{w(t)}{D} + R_{\mathrm{OFF}}\left(1 - \frac{w(t)}{D}\right) \tag{3}$$

where D is the thickness of the semiconductor film sandwiched between two Pt electrodes and μ_v is the oxygen ions mobility. The width of the doped region $\mathrm{TiO}_{2-\mathrm{x}}$ is represented by the state variable $w(t)$. The resistance of the memristor is denoted by R_{ON} when the doped region extends to the full length (i.e., $W/D = 1$). Similarly, if the undoped zone extends the entire length of the device (i.e., $W/D = 0$), the overall resistance is represented by R_{OFF}.

3 Proposed Memristor-Based PFSCL Architectures

In this section, an amalgamation of memristor in PFSCL style is presented, and two architectures are proposed. The first architecture embeds a memristor network with a PFSCL inverter, while the latter uses PFSCL NOR/OR gate.

3.1 Proposed Architecture 1: Memristor logic with PFSCL Inverter/Buffer (Mem-PA1)

The proposed architecture 1 (Mem-PA1) implements the logic function using memristor-based network whose output is fed as an input to the PFSCL inverter/buffer. The same is illustrated through a block diagram in Fig. 4. The Mem-PA1-based logic

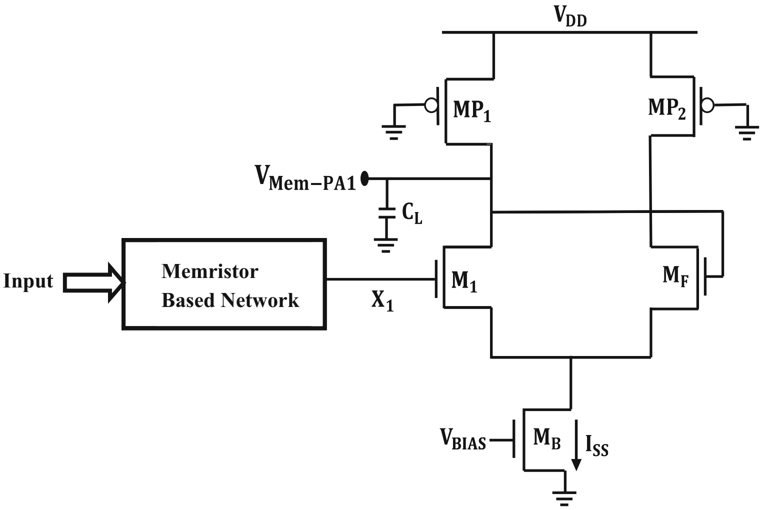


Fig. 4 Block diagram of proposed Mem-PA1 architecture

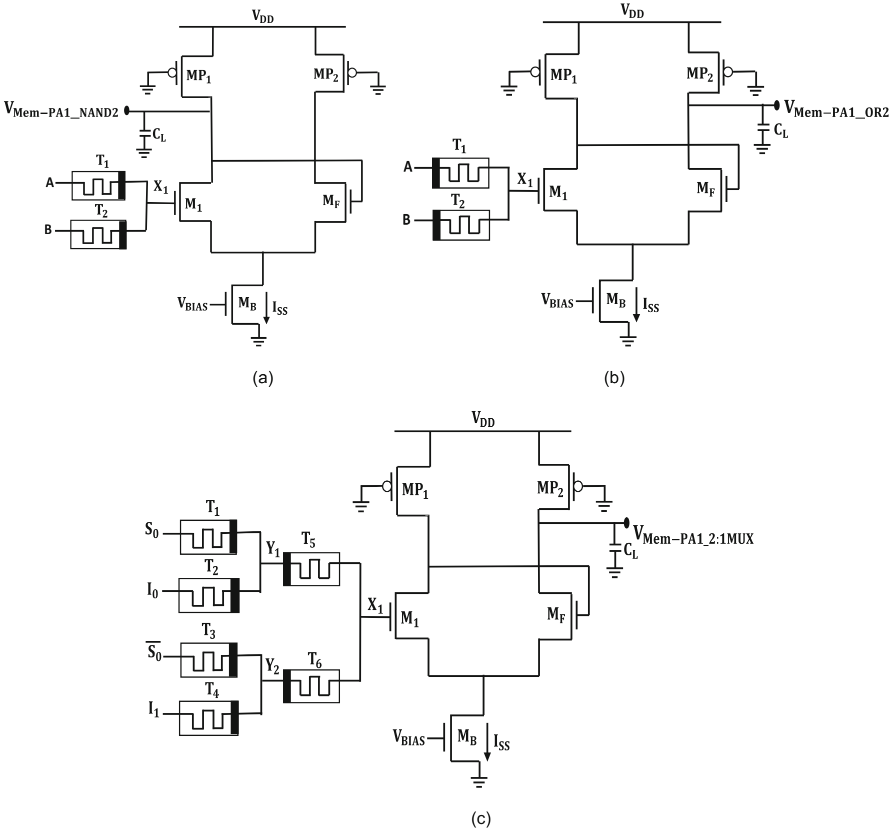


Fig. 5 Block diagram of **a** Mem-PA1 NAND2 **b** Mem-PA1 OR2 **c** Mem-PA1 2:1 MUX

function realization of basic gates, namely two-input NAND (Mem-PA1 NAND2) gate, two-input OR (Mem-PA1 OR2) gate and 2:1 multiplexer MUX (Mem-PA1 2:1 MUX), is shown in Fig. 5a-c. For a NAND2 gate (Fig. 5a), the inputs A and B are connected to the memristors T_1 and T_2, respectively. The intermediate output X_1 drives the transistor M_1 of PFSCL inverter/buffer. For the case when both the inputs A and B are at low voltage level, then no current flows through memristors T_1 and T_2 such that node X_1 attains low voltage level. The low potential of node X_1 makes the bias current I_{SS} to get steered in transistor M_{F}. Thus, there does not exist any potential drop across MP_1 and a high output voltage level is obtained at the output, $V_{\mathrm{Mem}\text{-}\mathrm{PA1\_NAND2}}$. Alternatively, when input A is at low and input B is at high voltage levels, there exists a current flow from T_2 to T_1, causing an increase in the potential of X_1. This high voltage level at node X_1 causes steering of the bias current I_{ss} through M_1 such that a low output level is obtained at the output. Analogously, the operation for the remaining two input cases of the proposed Mem-PA1 NAND2 can be worked upon.

Considering the voltage levels of PFSCL style, the voltage level of MBN at node X_1 for different input combinations is analyzed here. Further, the resistances of the memristor in high resistance state (HRS) and low resistance state (LRS) are assumed to be R_{OFF} and R_{ON}, respectively. The voltages at different nodes may now be found by simply applying the voltage divider rule. A NAND2 gate has two inputs; therefore, a total of 4 combinations are possible and are given attention below.

Case-1: Inputs A and B are both at low voltage level.

The memristors T_1 and T_2 are driven by low input voltage so there is no current flow and node X_1 is maintained at V_{OL}. Thus, for a low voltage level (V_{OL}) at the input of the PFSCL inverter, the bias current I_{SS} gets steered through M_F and a high voltage level ($V_{\mathrm{OH}} = V_{\mathrm{DD}}$) is obtained at the output node $V_{\text{MEM-PA1\_NAND2}}$.

Case-2: Inputs A and B are low and high voltage levels, respectively.

The memristors T_1 and T_2 driven by A and B are in HRS and LRS states, respectively. This condition enables a current flow from T_2 to T_1. Thus, the voltage at node X_1 can be expressed as:

$$V(X_1) = \frac{(V_{\mathrm{DD}} - V_{\mathrm{SWING}})R_{\mathrm{OFF}} + V_{\mathrm{DD}}R_{\mathrm{ON}}}{R_{\mathrm{ON}} + R_{\mathrm{OFF}}} \tag{4}$$

As $R_{\mathrm{OFF}} \gg R_{\mathrm{ON}}$, Eq. (4) can be reduced to

$$V(X_1) = V_{\mathrm{DD}} - V_{\mathrm{SWING}} = V_{\mathrm{OL}} \tag{5}$$

As low voltage level is obtained at node X_1, the bias current I_{SS} will steer through M_F. Thus, high a voltage level ($V_{\mathrm{OH}} = V_{\mathrm{DD}}$) is obtained at the output node $V_{\text{MEM-PA1\_NAND2}}$.

Case-3: Inputs A and B are high and low voltage levels, respectively.

The memristors T_1 and T_2 have LRS and HRS states, respectively. This condition enables a current flow from T_1 to T_2. Thus, the voltage at node X_1 can be expressed as:

$$V(X_1) = \frac{V_{\mathrm{DD}}R_{\mathrm{ON}} + (V_{\mathrm{DD}} - V_{\mathrm{SWING}})R_{\mathrm{OFF}}}{R_{\mathrm{ON}} + R_{\mathrm{OFF}}} \tag{6}$$

With $R_{\mathrm{OFF}} \gg R_{\mathrm{ON},}$ Eq. (6) can be reduced as:

$$V(X_1) = \frac{V_{\mathrm{DD}}(R_{\mathrm{ON}} + R_{\mathrm{OFF}})}{R_{\mathrm{ON}} + R_{\mathrm{OFF}}} - V_{\mathrm{SWING}}\frac{R_{\mathrm{OFF}}}{R_{\mathrm{OFF}} + R_{\mathrm{ON}}} \tag{7}$$

$$V(X_1) = V_{\mathrm{DD}} - V_{\mathrm{SWING}} = V_{\mathrm{OL}} \tag{8}$$

Since $V(X_1)$ is low, a high voltage level ($V_{\mathrm{OH}} = V_{\mathrm{DD}}$) is obtained at the output node $V_{\text{MEM-PA1\_NAND2.}}$

Case 4: Inputs A and B are both at high voltage levels.

The memristors T_1 and T_2 are driven by high inputs, so there is no current flow and node Y_1 will be at V_{OH}. Thus, the bias current I_{SS} will steer through M_1 and a low voltage level ($V_{\mathrm{OL}} = V_{\mathrm{DD}} - V_{\mathrm{SWING}}$) is obtained at the output node $V_{\text{Mem-PA1\_NAND2}}$. It can be observed that the required functionality is achieved.

Table 1 Summary of node voltages for Mem-PA1 OR2 PFSCL gate

Inputs		Intermediate output	$V_{\text{Mem-PA1\_OR2}}$
A	B	X_1	
V_{OL}	V_{OL}	V_{OL}	V_{OL}
V_{OL}	V_{OH}	V_{OH}	V_{OH}
V_{OH}	V_{OL}	V_{OH}	V_{OH}
V_{OH}	V_{OH}	V_{OH}	V_{OH}

where $V_{OH} = V_{DD}$ and $V_{OL} = V_{DD} - V_{SWING}$

Table 2 Summary of node voltages for Mem-PA1 2:1 PFSCL MUX

Inputs			Intermediate outputs			$V_{\text{Mem-PA1\_MUX2:1}}$
S_0	I_0	I_1	Y_1	Y_2	X_1	
V_{OL}	V_{OL}	V_{OL}	V_{OL}	V_{OL}	V_{OL}	V_{OL}
V_{OL}	V_{OL}	V_{OH}	V_{OL}	V_{OH}	V_{OH}	V_{OH}
V_{OL}	V_{OH}	V_{OL}	V_{OL}	V_{OL}	V_{OL}	V_{OL}
V_{OL}	V_{OH}	V_{OH}	V_{OL}	V_{OH}	V_{OH}	V_{OH}
V_{OH}	V_{OL}	V_{OL}	V_{OL}	V_{OL}	V_{OL}	V_{OL}
V_{OH}	V_{OL}	V_{OH}	V_{OL}	V_{OL}	V_{OL}	V_{OL}
V_{OH}	V_{OH}	V_{OL}	V_{OH}	V_{OL}	V_{OH}	V_{OH}
V_{OH}	V_{OH}	V_{OH}	V_{OH}	V_{OL}	V_{OH}	V_{OH}

where $V_{OH} = V_{DD}$ and $V_{OL} = V_{DD} - V_{SWING}$

Similarly, Mem-PA1 OR2 realizes the functionality of a OR gate (Fig. 5b). The concept is extended to the design of 2:1 MUX shown in Fig. 5c. The minterms S_0I_0 and $\overline{S_0}$ I_1 are realized by memristors (T_1, T_2) and (T_3, T_4), respectively. The corresponding outputs Y_1, Y_2 are further fed as inputs to the memristor (T_5, T_6) arranged for OR realization. The intermediate output X_1 drives the PFSCL buffer to obtain the multiplexer functionality. Tables 1 and 2 summarize the node voltages for different input combinations in Mem-PA1 and Mem-PA2 2:1 MUX.

3.2 Proposed Architecture 2: Memristor Logic with PFSCL NOR/OR Gate

The proposed architecture 2 (Mem-PA2) is based on incorporating the advantages of PFSCL NOR/OR gate with memristor network. In this architecture, the minterms defined in a logic function are realized using a memristor network, and the corresponding intermediate outputs are then fed to the input transistors of a PFSCL NOR/OR gate as depicted in Fig. 6a. An implementation of 2:1 MUX in Mem-PA2 is drawn in Fig. 6b for illustration. The two minterms S_0I_0 and $\overline{S_0}$ I_1 are realized by memristors (T_1, T_2) and (T_3, T_4), respectively. The corresponding outputs Y_1 and Y_2 are further

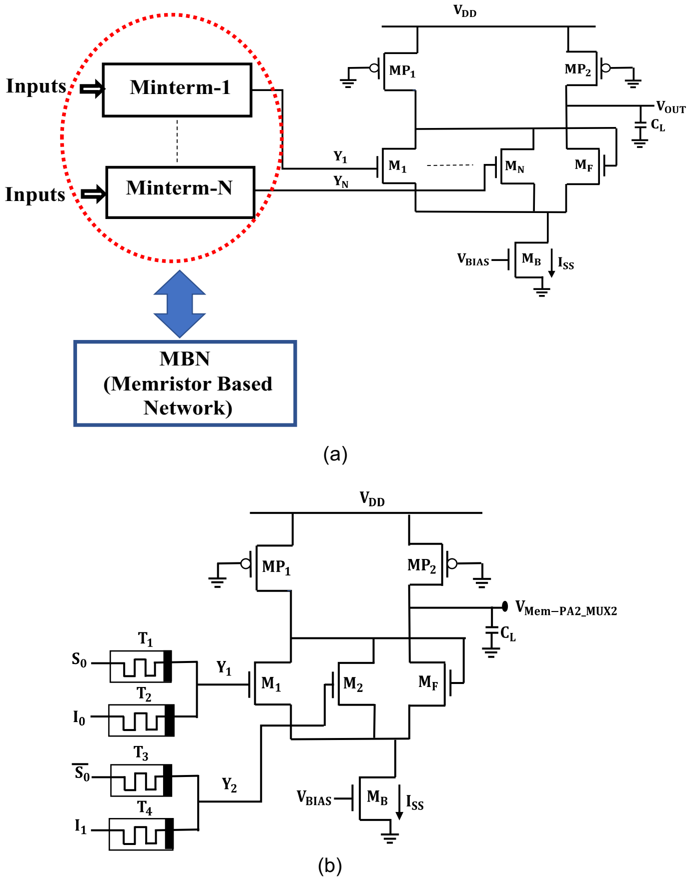


Fig. 6 **a** Block diagram of Mem-PA2 **b** Mem-PA2 2:1 MUX

fed to the input transistors (M_1, M_2) of the PFSCL OR2 gate. A summary of the same is presented in Table 3.

3.3 Design Examples

In this subsection, a three-input exclusive-OR (XOR3) gate and a 4-bit ripple carry adder (RCA) are chosen to illustrate the benefits of employing proposed architectures in circuit design.

3.3.1 Three-Input Exclusive-OR (XOR3) Gate

The NOR-based implementation of a XOR3 gate may be expressed as:

$$F = \overline{\overline{(A + B + C)} + \overline{(A + \overline{B} + \overline{C})} + \overline{\left(\overline{A} + B + \overline{C}\right)} + \overline{\left(\overline{A} + \overline{B} + C\right)}} \qquad (9)$$

Table 3 Summary of node voltages for Mem-PA2 2:1 PFSCL MUX

Inputs			Intermediate outputs		$V_{\text{Mem-PA2\_2:1MUX}}$
S_0	I_0	I_1	Y_1	Y_2	
V_{OL}	V_{OL}	V_{OL}	V_{OL}	V_{OL}	V_{OL}
V_{OL}	V_{OL}	V_{OH}	V_{OL}	V_{OH}	V_{OH}
V_{OL}	V_{OH}	V_{OL}	V_{OL}	V_{OL}	V_{OL}
V_{OL}	V_{OH}	V_{OH}	V_{OL}	V_{OH}	V_{OH}
V_{OH}	V_{OL}	V_{OL}	V_{OL}	V_{OL}	V_{OL}
V_{OH}	V_{OL}	V_{OH}	V_{OL}	V_{OL}	V_{OL}
V_{OH}	V_{OH}	V_{OL}	V_{OH}	V_{OL}	V_{OH}
V_{OH}	V_{OH}	V_{OH}	V_{OH}	V_{OL}	V_{OH}

where $V_{\text{OH}} = V_{\text{DD}}$ and $V_{\text{OL}} = V_{\text{DD}} - V_{\text{SWING}}$

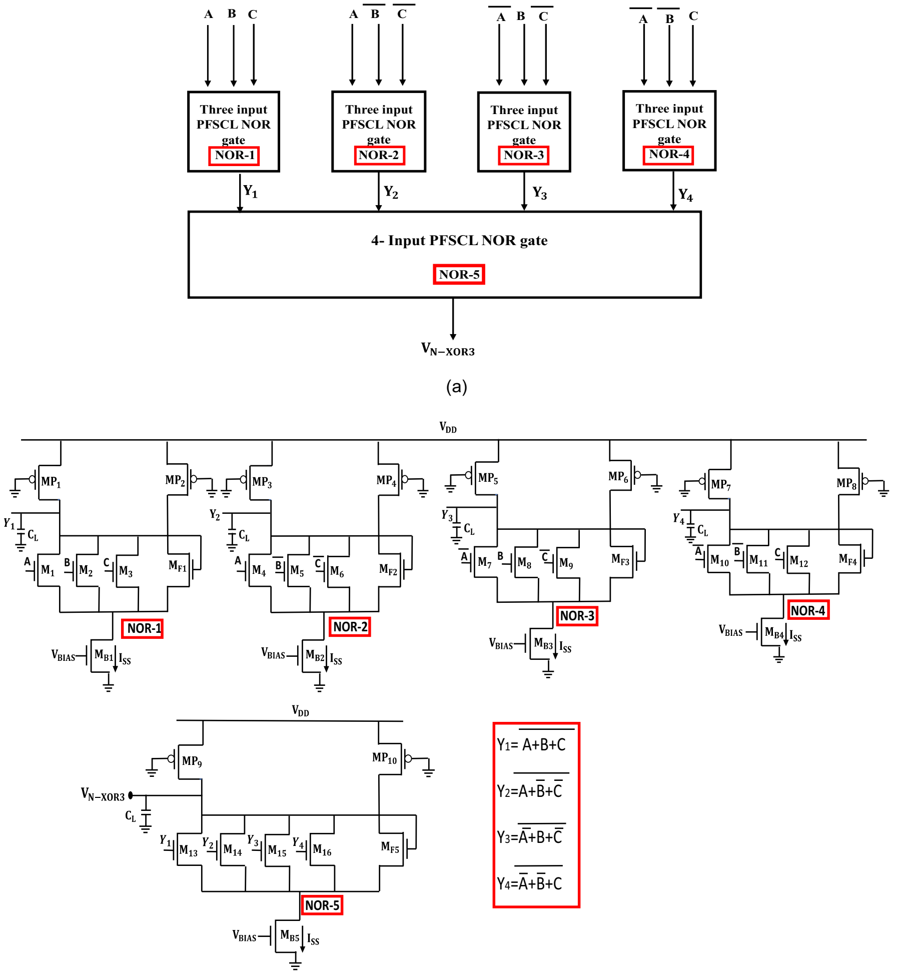
$$F = \overline{Y_1 + Y_2 + Y_3 + Y_4} \tag{10}$$

The block diagram of NOR/OR XOR3 gate-based design is shown in Fig. 7a. It requires a four input PFSCL NOR/OR gate driven by the outputs of the three-input PFSCL NOR/OR gates. The MOS-based realization for XOR3 gate is shown in Fig. 7b. The minterms are realized by NOR/OR PFSCL gates (NOR-1, NOR-2, NOR-3 and NOR-4), and the corresponding outputs (Y_1–Y_4) are further fed to the NOR/OR PFSCL gate (NOR-5) to realize XOR3 gate functionality. The MOS-based realization requires 31 transistors and 5 current sources. The involvement of multiple current sources leads to a design with high power dissipation which makes this approach unattractive.

The FC-based XOR3 gate may be realized by cascading two FC-based XOR2 gates. The block diagrammatic representation is depicted in Fig. 8a which uses two FCs of Fig. 2c. The complete schematic shown in Fig. 8b uses 20 transistors and 4 current sources. This number is significantly lesser than 31 transistors and 5 current sources that are used in NOR/OR-based counterpart. However, the requirement of wider control transistors in the middle branch (M_2, M_4, M_6 and M_8) in FC for proper activation/deactivation possesses a limitation in terms of larger footprints. It is relevant to mention here that FC-based XOR3 gate uses 4 current sources of $I_{\text{SS}}/2$ value each in comparison to 5 current sources (of I_{SS} value each) in its NOR/OR-based counterpart. Thus, there will be a significant saving on power consumption due to the reduced number of current sources which carry half bias current.

Alternatively, the minterms defined in a logic function are realized using a memristor network and the corresponding intermediate outputs are then fed to the input transistors in proposed Mem-PA1/ Mem-PA2 gate. In minterms, the XOR3 gate function may be expressed as:

$$V_{\text{N}-\text{XOR3}} = A.B.C + A.\overline{B}.\overline{C} + \overline{A}B\overline{C} + \overline{A}.\overline{B}.C \tag{11}$$

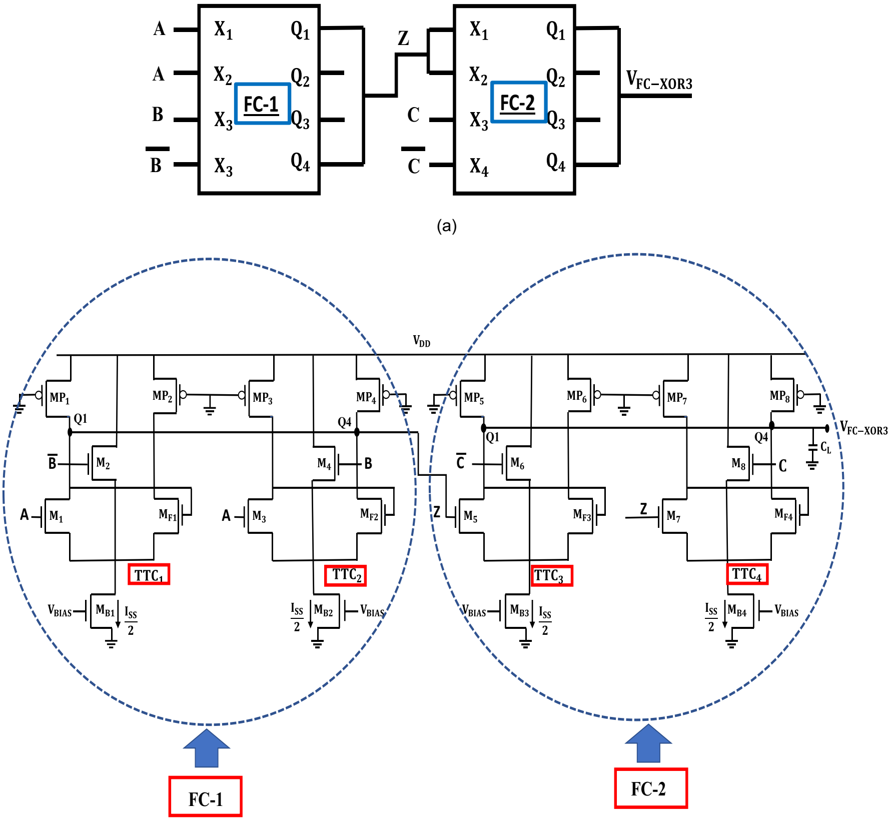


(a)

(b)

Fig. 7 Existing NOR/OR-based XOR3 gate implementation **a** Block diagram **b** Circuit diagram

The proposed Mem-PA1-based XOR3 gate implementation is drawn in Fig. 9. The minterms $A.B.C$, $A.\overline{B}.\overline{C}$, $\overline{A}.B.\overline{C}$ and $\overline{A}.\overline{B}.C$ are realized by the memristors (T_1–T_3), (T_4 –T_6), (T_7–T_9), (T_{10}–T_{12}), respectively, and the corresponding outputs (m_1–m_4) are further fed to the memristors (T_{13}–T_{16}) to realize OR functionality. The output of the memristor network drives the PFSCL inverter. The proposed Mem-PA2-based XOR3 gate implementation is depicted in Fig. 10. Similar to Mem-PA1-based XOR3 gate implementation, the minterms $A.B.C$, $A.\overline{B}.\overline{C}$, $\overline{A}.B.\overline{C}$ and $\overline{A}.\overline{B}.C$ are realized by the memristors (T_1–T_3), (T_4–T_6), (T_7–T_9) and (T_{10}–T_{12}), respectively. The corresponding intermediate outputs are then fed to the input transistors M_1–M_4 in a PFSCL NOR/OR gate. A closer look of Figs. 9 and 10 reveals that the proposed architecture-based realizations of XOR3 gate require a single PFSCL gate in contrast to multiple



(a)

(b)

Fig. 8 Existing FC-based XOR3 gate implementation **a** Block diagram **b** Circuit diagram

gates in the existing schemes. Therefore, gate count in the proposed scheme is minimum among PFSCL variants, leading to better performance of the circuits.

3.3.2 Bit RCA Design

A 4-bit RCA is chosen as a representation for larger functions with multiple levels, as shown in Fig. 11. It consists of four full adders with two inputs as ($A_0A_1A_2A_3$), ($B_0B_1B_2B_3$) and input carry (C_0). The final sum bits are taken from individual full adders ($S_0S_1S_2S_3$), and the carry output (C_4) from the last adder. The full adder is designed in existing NOR (Fig. 12a), FC (Fig. 12b)-, proposed Mem-PA1 (Fig. 12c)- and the proposed Mem-PA2 (Fig. 12d)-based architectures. It can be observed that NOR-based full adder design involves six levels with nine PFSCL NOR gates, while FC full adder design has three levels with seven PFSCL FCs. Alternatively, the full adder design with proposed Mem-PA1 has two levels with three PFSCL inverters, whereas the proposed Mem-PA2 uses four PFSCL NOR gates arranged in two levels.

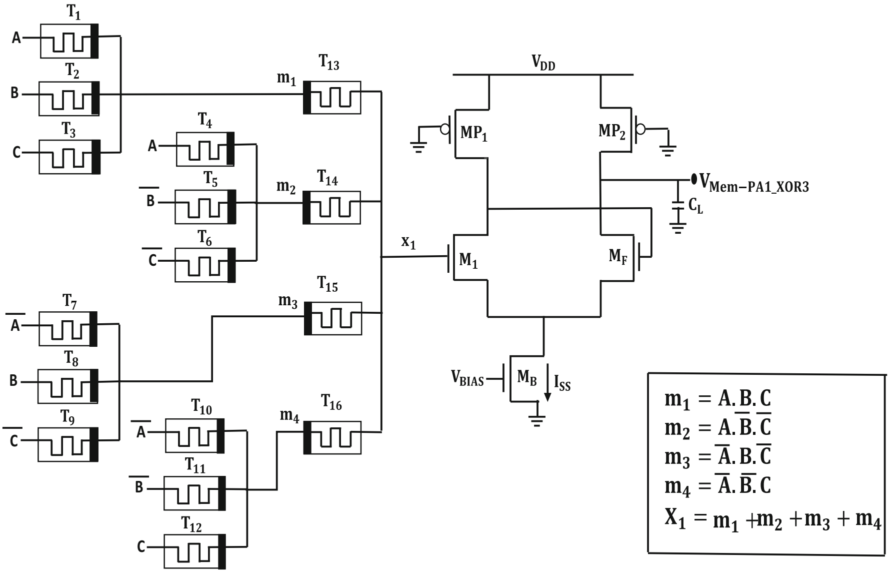


Fig. 9 Proposed Mem-PA1 XOR3 gate

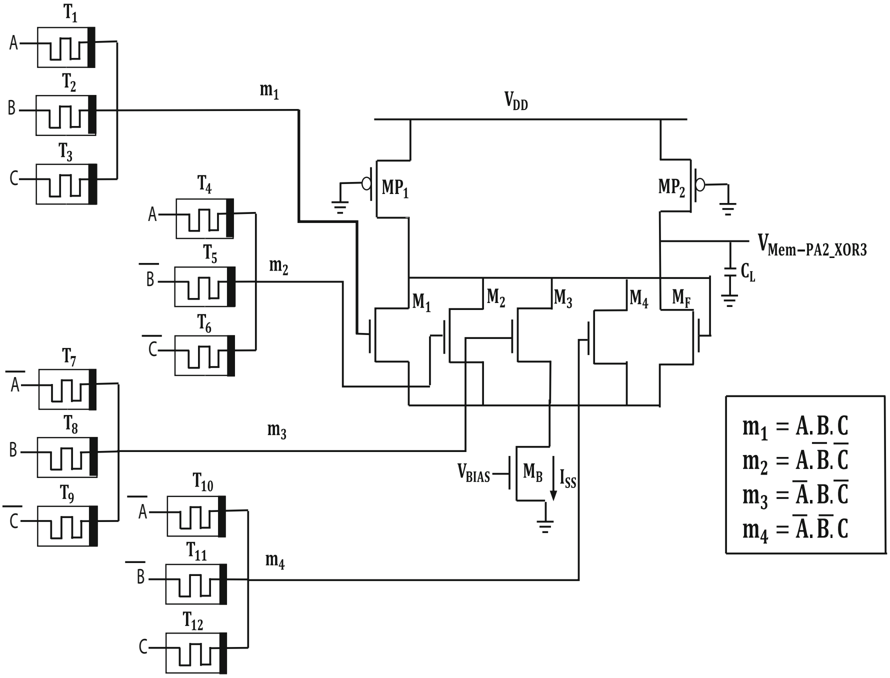


Fig. 10 Proposed Mem-PA2 XOR3 gate

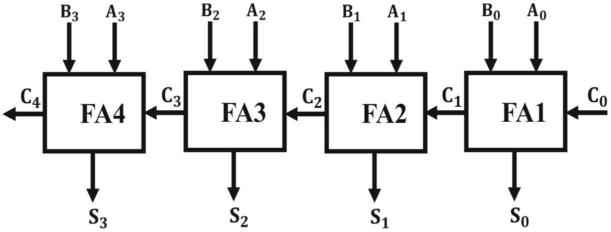


Fig. 11 Bit RCA design

4 Simulation Section

In this section, the functionality of the gates, namely two-input AND (AND2) gate, OR (OR2), two-input exclusive-OR gate (XOR2), 2:1 MUX and three-input exclusive-OR gate (XOR3), based on proposed Mem-PA1 and Mem-PA2 is verified, and their performance under process variations is compared with existing counterparts. The simulation results for 4-bit RCA are also presented and compared. All the simulations are carried out in LT spice by using PTM 180 nm CMOS technology parameters and with a power supply, bias current load capacitance, fan-out (FO) and voltage swing of 1.1 V, 100μA, 50fF, 4 and 400 mV, respectively. Two memristor models suggested in [7] and [38] are used, and the layouts are drawn in microwind to put forward the comparison in pre- and post-layout simulation results.

4.1 Functional Verification

All the gates based on existing and proposed architectures are simulated, and the timing waveforms are observed in Fig. 13. In the timing waveforms of a XOR2 gate, it can be observed that when both the inputs are at same logic level (either high or low), then all the gates produce a low ($V_{\mathrm{OL}} = V_{\mathrm{DD}} - V_{\mathrm{SWING}} = 0.7$ V) output voltage. However, when the inputs have different voltage levels as (high, low), then a high voltage ($V_{\mathrm{OH}} = V_{\mathrm{DD}} = 1.1$ V) is obtained at the output node. Similarly, the correct functionality is exhibited by each gate in all the architectures.

To highlight the impact of proposed architectures on delay, the time taken for the output to switch from high logic level to low logic level in a XOR3 (Fig. 13c) is marked. By observing time intervals (A, B, C, D), it can be inferred that time intervals (C-D) of Mem-PA1 and Mem-PA2 XOR3 gate are shorter in comparison to the gate based on existing architectures. Similar curves are obtained for the rest of the gates.

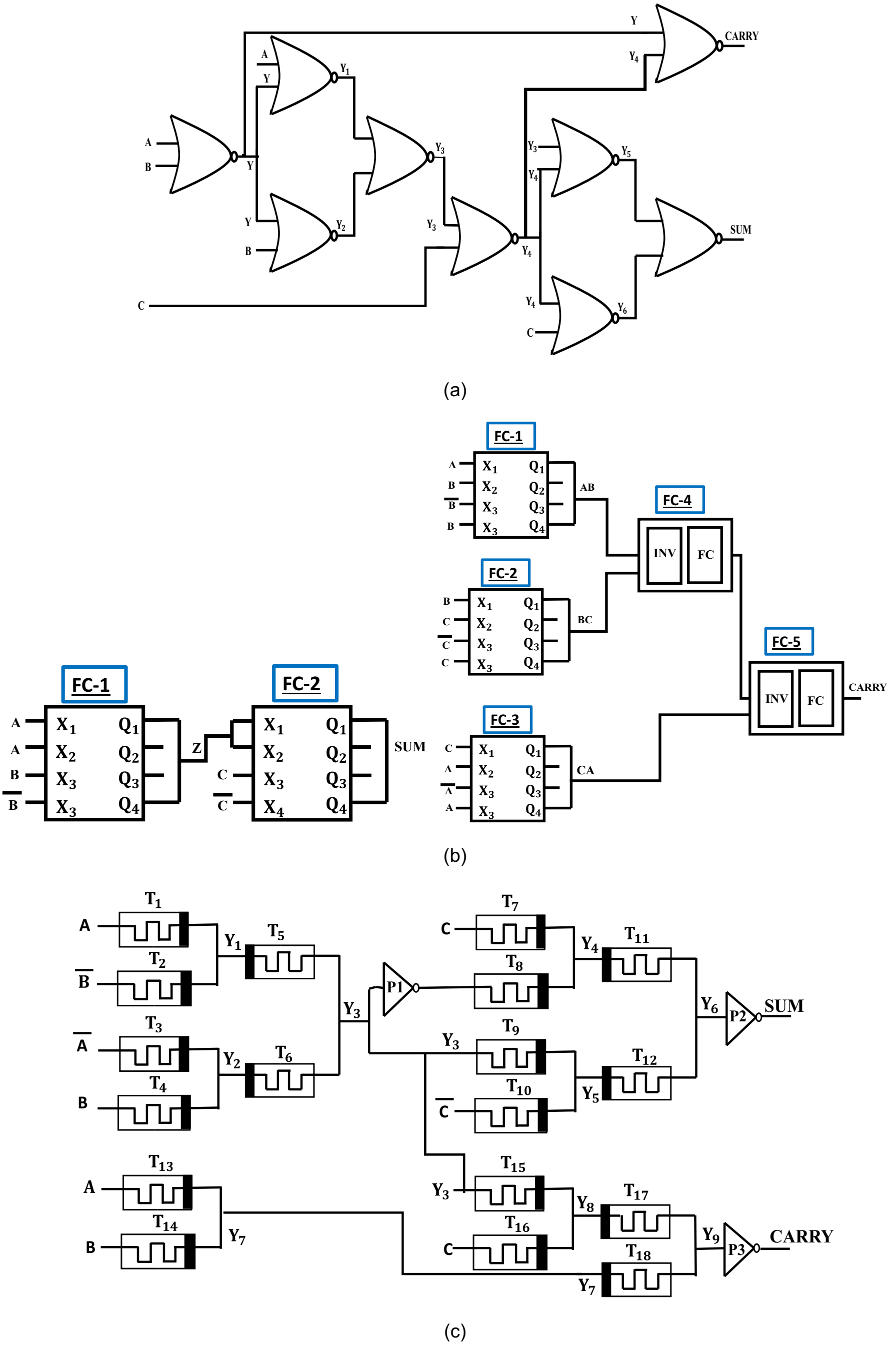
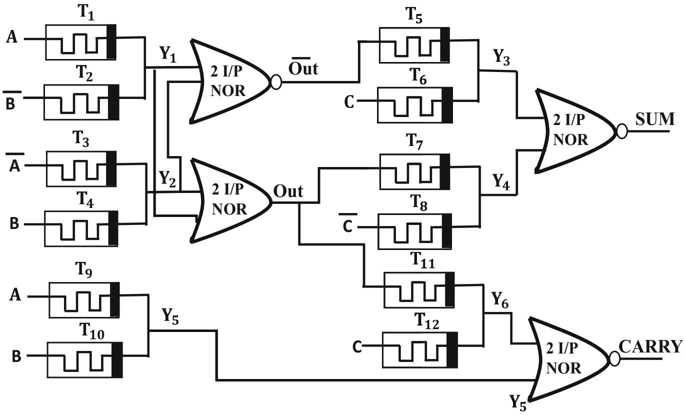


Fig. 12 PFSCL full adder design **a** NOR-based **b** FC-based **c** Mem-PA1-based **d** Mem-PA2-based architectures



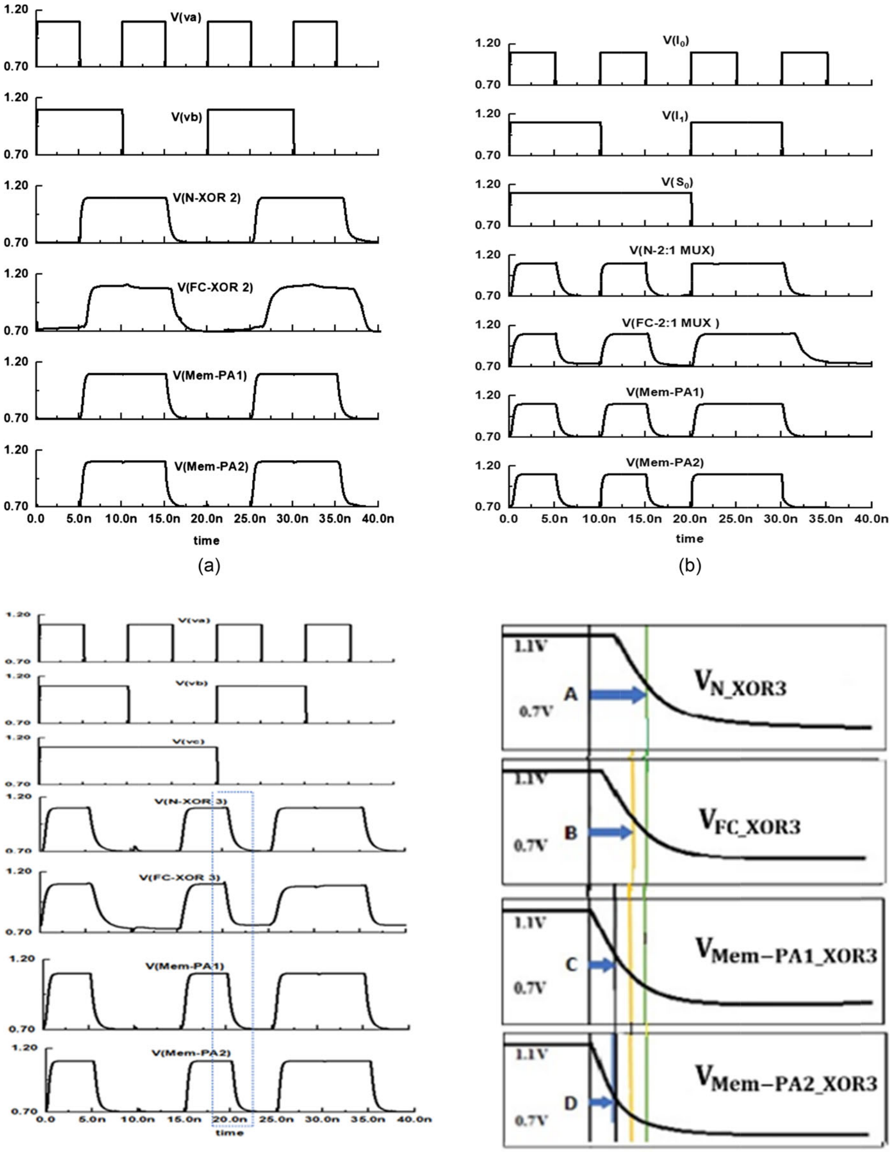
(d)

Fig. 12 continued

4.2 Performance Comparison

Various performance parameters such as power consumption, propagation delay ($C_L = 50$ fF and FO = 4), power delay product (PDP), gate count, area and noise margin are summarized in Table 4. The layouts of the proposed Mem-PA1- and Mem-PA2-based XOR2 gates are shown in Fig. 14. The following observations are made:

1. The conventional NOR/OR architecture is best suited to realize OR2 and AND2 gates as they outperform in comparison to the existing FC architecture as well as the proposed Mem-PA1 architecture. These gates cannot be realized using Mem-PA2 since there exists only one minterm.
2. For the functions expressed as sum of minterms, the results indicate that conventional NOR/OR architecture shows the largest propagation delay, power and PDP values which is a consequence of high gate count and multiple stages.
3. The proposed architectures-based gates show 12% to 33% improvement in propagation delay for a load capacitance, C_L, of 50fF with respect to their existing counterparts, thus giving an edge for the implementation of large fan-in gates. A similar trend is observed by simulating with a fan-out of 4.
4. Out of the two proposed architectures, it can be observed that Mem-PA2 gate shows 5.16% improvement in propagation delay over Mem-PA1 gate for same value of power consumption.
5. The proposed architecture-based gates show a maximum improvement of 80% and 86% in power and PDP values, respectively, over existing counterparts. This improvement is attributed to better power delay parameters in proposed architectures.
6. In terms of area, the proposed architecture-based gates have reduced area due to the fabrication by integrating memristors on top of a CMOS substrate using post-processing, e.g., nanoimprint lithography (NIL), a process that does not disturb the CMOS circuitry underneath [32, 40]. A maximum area reduction of 97% is



(a)

(b)

(c)

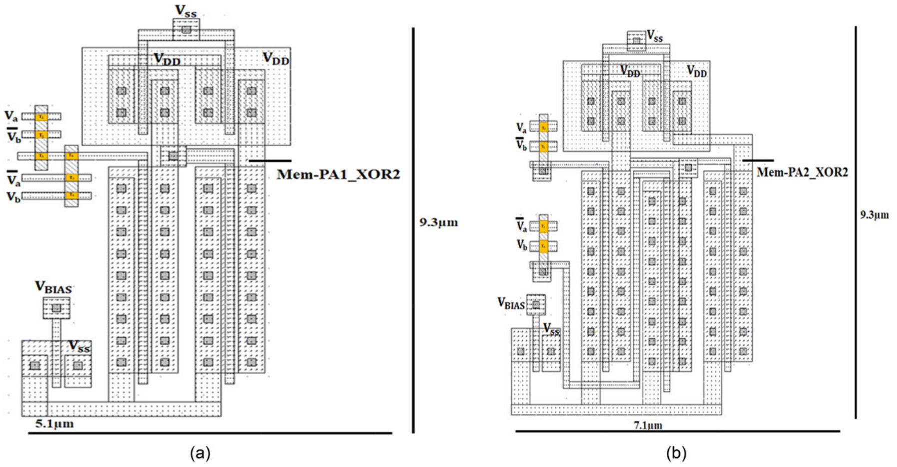
Fig. 13 Simulation waveforms **a** XOR2 **b** 2:1 MUX **c** XOR3

Table 4 Performance comparison

Function	Parameter	Architecture			
		NOR-based design	FC-based design	Mem-PA1-based design	Mem-PA2-based design
OR2	Power (μW)	110	110	110	–
	Propagation delay (ps) ($C_L = 50fF$)	494	595	496	
	Propagation delay (ps) (FO = 4)	328	795	300	
	PDP (fJ)	54.34	65.45	54.56	
	Gate count	1	1	1	
	Area (μm^2)	66.03	915	47.43	
	Noise margin (mV)	163	119	128	
AND2	Power (μW)	110	110	110	–
	Propagation delay (ps) ($C_L = 50fF$)	510	595	496	
	Propagation delay (ps) (FO = 4)	636	795	300	
	PDP (fJ)	56.1	65.45	54.56	
	Gate count	1	1	1	
	Area (μm^2)	66.03	915	47.43	
	Noise margin (mV)	163	119	128	
2:1 MUX	Power (μW)	330	110	110	110
	Propagation delay (ps) ($C_L = 50fF$)	730	595	506	496
	Propagation delay (ps) (FO = 4)	801	795	314	363
	PDP (fJ)	240.9	65.45	55.66	54.56
	Gate count	3	1	1	1
	Area (μm^2)	228	915	47.43	66.03
	Noise margin (mV)	163	119	128	164
XOR2	Power (μW)	330	110	110	110

Table 4 (continued)

Function	Parameter	Architecture			
		NOR-based design	FC-based design	Mem-PA1-based design	Mem-PA2-based design
	Propagation delay (ps) ($C_L = 50fF$)	730	595	506	496
	Propagation delay (ps) (FO = 4)	801	795	314	363
	PDP (fJ)	240.9	65.45	55.66	54.56
	Gate count	3	1	1	1
	Area (μm^2)	228	915	47.43	66.03
	Noise margin (mV)	163	119	128	164
XOR3	Power (μW)	550	220	110	110
	Propagation delay (ps) ($C_L = 50fF$)	800	787	523	496
	Propagation delay (ps) (FO = 4)	1979	1051	362	449
	PDP (fJ)	440	173.14	57.53	54.56
	Gate count	5	2	1	1
	Area (μm^2)	498	1830	47.43	125
	Noise Margin (mV)	182	119	128	188



(a) (b)

Fig. 14 Layout of the proposed XOR2 gate based on **a** Mem-PA1 architecture **b** Mem-PA2 XOR2 architecture

observed with Mem-PA1 architectures in comparison to FC-based design of XOR3 gate.

7. The FC-based design has the lowest noise margin as they are not able to steer the bias current completely in the outer transistors of the activated TTC.
8. The proposed Mem-PA1 gates have lower noise margin in comparison to existing NOR-based realizations and Mem-PA2 gates, since both of them connect transistors in parallel in their PDN and are able to make high-to-low transitions at much lower input voltages in comparison to Mem-PA1 that connects only one transistor in the PDN.

All designs of XOR2 gate based on existing and proposed architectures are laid out, and post-layout simulations are carried out. The observed propagation delay and PDP values are summarized in Table 5. It is found that the post-layout simulation results for all XOR2 gates show an increase in propagation delay and PDP values. The proposed Mem-PA2 XOR2 gate, however, shows minimum propagation delay and PDP among all designs. Further, a 1000-point Monte Carlo (MC) simulation on R_{off}, R_{on}, D and threshold voltage of MOS transistors is performed for the proposed XOR gates. It is observed that the propagation delays of Mem-PA1 XOR2 and Mem-PA2 XOR2 gates have a mean and standard deviation of (513 ps, 28 ps) and (497 ps, 28 ps), respectively. Similarly, the voltage swing of Mem-PA1 XOR2 and Mem-PA2 XOR2 has a mean and standard deviation of (394 mV, 13 mV) and (393 mV, 14 mV), respectively. Its effect on $\mathrm{V_{SWING}}$ and propagation delay is illustrated for Mem-PA2 XOR2 gates through timing waveform and histogram as shown in Fig. 15. It may be noted that both the gates show similar variations.

The simulation results of 4-bit RCA are listed in Table 6. The corresponding simulation results for the 4-bit RCA show a maximum reduction of 70% and 66% in propagation delay and power consumption values of the proposed architecture-based 4-bit RCA with respect to their existing counterparts, respectively. Also, out of the two proposed architectures, RCA based on Mem-PA2 gate shows 11% improvement in propagation delay over its Mem-PA1-based counterpart.

The impact of technology scaling on propagation delay and PDP values is studied on XOR2 gate at 180 nm, 130 nm and 90 nm CMOS technology nodes. The simulation results are summarized in Table 7. It may be observed that propagation delay and PDP values improve for all architectures with technology scaling. A maximum of 21% improvement in the delay as well as PDP is measured in Mem-PA2-based XOR2 gate.

Further, the performance is also examined at different design corners, and the findings for proposed architectures and existing counterparts-based XOR2 gate are summarized in Fig. 16 and Table 8. It is found that the propagation delay, power and PDP of the proposed Mem-PA1 XOR2 gate vary by a factor of 0.33, 0.81 and 0.33, respectively, between the best and the worst cases, whereas the corresponding variation for the proposed Mem-PA2 XOR2 gate is 0.31, 0.81 and 0.33, respectively. Thus, it is clear that the proposed Mem-PA1 shows approximately similar variations as the Mem-PA2 for different design corners. Overall, Mem-PA2 can be chosen as an optimum design option due to better PDP and propagation delay.

Table 5 Pre- and post-layout simulation results of XOR2 gate

Simulation	Parameter	NOR-based design	FC-based design	Mem-PA1-based design	Mem-PA2-based design
Pre-layout/Biolek model [7]	Propagation delay (ps)	730	595	506	496
	PDP (fJ)	240.9	65.45	55.66	54.56
Post-layout/memristor [38]	Propagation delay (ps)	897	742	779	607
	PDP (fJ)	296.0	81.620	85.690	66.77

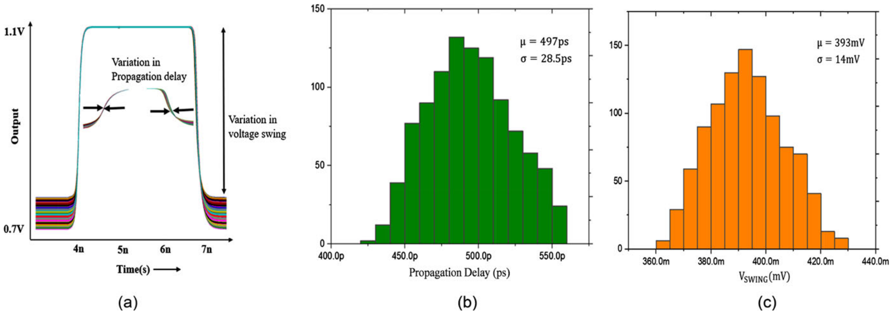


Fig. 15 1000-point MC simulation results **a** timing waveform; histogram for variation in **b** propagation delay **c** V_{SWING}

Table 6 Simulation results for 4-bit RCA

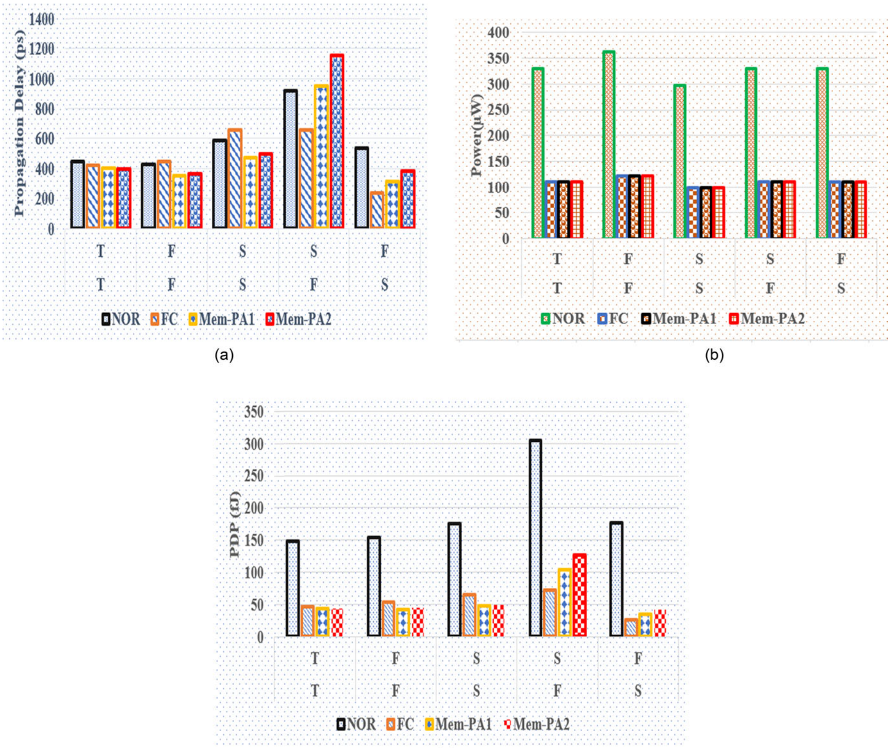
Parameter	Architecture			
	NOR-based design	FC-based design	Mem-PA1-based design	Mem-PA2-based design
Propagation delay (ns)	4.189	3.141	1.410	1.253
Power (μW)	3960	3080	1320	1760

Table 7 Performance comparison for XOR2 at different technology nodes

Architecture	Propagation Delay(ps)			PDP (fJ)		
	Technology node (nm)			Technology node (nm)		
	180	130	90	180	130	90
NOR	730	580	450	240.9	191.4	148
FC	595	570	420	65.4	60.5	46.2
Mem-PA1	506	450	400	55.6	49.5	44
Mem-PA2	496	444	396	54.5	48.8	43.5

5 Conclusion

This paper introduces memristors in PFSCL circuit design. Two architectures employing memristor network with a PFSCL inverter and a PFSCL NOR/OR gate are proposed. This paper addresses the limitation of existing PFSCL architectures in the realization of complex logic gates. Here memristor is introduced in PFSCL style and two architectures are suggested wherein MBN is embedded in PFSCL inverter and PFSCL NOR/OR gate. Different functions based on the proposed architectures are



(a)

(b)

(c)

Fig. 16 Performance at different design corners for XOR2 gate **a** Propagation delay variation **b** Power variation **c** PDP variation

Table 8 Process corner analysis for XOR2 gate realized in different architectures

Parameters	Architecture	T	F	S	F	S
		T	F	S	S	F
Power (µW)	NOR	330	363	297	330	330
	FC	110	121	99	110	110
	Mem-PA1	110	121	99	110	110
	Mem-PA2	110	121	99	110	110
Propagation delay (ps)	NOR	450	425	590	921	536
	FC	420	445	658	657	234
	Mem-PA1	400	350	475	948	314
	Mem-PA2	396	367	495	1151	384
PDP (fJ)	NOR	148.5	154.275	175.23	303.93	176.88
	FC	46.2	53.845	65.142	72.27	25.74
	Mem-PA1	44	42.35	47.025	104.28	34.54
	Mem-PA2	43.5	44.407	49.005	126.61	42.24

realized, and their performance is compared with the existing counterparts through simulations. A maximum improvement of 33%, 80% and 86% in propagation delay, power consumption and power delay product, respectively, is observed for Mem-PA1 exclusive-OR(XOR3) gate with respect to the existing architectures-based gates. Further, it is observed that the Mem-PA2-based XOR3 gate shows a 5.16% improvement in delay values over Mem-PA1 gate. The effect of parameter variations is also studied for functions designed using the proposed and existing architectures, and it is clear that the proposed Mem-PA1 shows approximately similar variations as the Mem-PA2 for different design corners. Thus, the use of Mem-PA2 offers an efficient design option for PFSCL designers.

Declarations

Conflict of interest The manuscript has no associated data.

References

1. A. Adamatzky, G. Chen (eds.), *Chaos, Cnn, Memristors and Beyond: A Festschrift For Leon Chua (With Dvd-rom, Composed By Eleonora Bilotta)* (World Scientific, Singapore, 2013)
2. K.A. Ali, M. Rizk, A. Baghdadi, J.-P. Diguet, J. Jomaah, MRL crossbar-based full adder design. in *2019 26th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*. 674–677 (2019). https://doi.org/10.1109/ICECS46596.2019.8964702
3. K.A. Ali, M. Rizk, A. Baghdadi, J.P. Diguet, J. Jomaah, Hybrid memristor–CMOS implementation of combinational logic based on X-MRL. Electronics **10**(9), 1018 (2021). https://doi.org/10.3390/electronics10091018
4. M. Alioto, L. Pancioni, S. Rocchi, V. Vignoli, Modeling and evaluation of positive-feedback source-coupled logic. IEEE Trans. Circuits Syst. I Regul. Pap. **51**(12), 2345–2355 (2004). https://doi.org/10.1109/TCSI.2004.838149
5. M. Alioto, L. Pancioni, S. Rocchi, V. Vignoli, Power–delay–area–noise margin tradeoffs in positive-feedback MOS current-mode logic. IEEE Trans. Circuits Syst. I Regul. Pap. **54**(9), 1916–1928 (2007). https://doi.org/10.1109/TCSI.2007.904685
6. M.W. Allam, M.I. Elmasry, Dynamic current mode logic (DyCML): A new low-power high-performance logic style. IEEE J. Solid-State Circ. **36**(3), 550–558 (2001). https://doi.org/10.1109/4.910495
7. D. Biolek, Z. Kolka, V. Biolkova, Z. Biolek, Memristor models for spice simulation of extremely large memristive networks. in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*. 389–392 (2016). https://doi.org/10.1109/ISCAS.2016.7527252
8. K. Cho, S.J. Lee, K. Eshraghian, Memristor-CMOS logic and digital computational components. Microelectron. J. **46**(3), 214–220 (2015). https://doi.org/10.1016/j.mejo.2014.12.006
9. L. Chua, Memristor-the missing circuit element. IEEE Trans. Circ. Theor. **18**(5), 507–519 (1971). https://doi.org/10.1109/TCT.1971.1083337
10. S. Duan, X. Hu, Z. Dong, L. Wang, P. Mazumder, Memristor-based cellular nonlinear/neural network: design, analysis, and applications. IEEE Trans. Neural Netw. Learn. Syst. **26**(6), 1202–1213 (2014). https://doi.org/10.1109/TNNLS.2014.2334701
11. Eshraghian K, Course notes on Memristive Circuits and Systems Technion, June (2011)
12. K. Gupta, U. Mittal, R. Baghla, P. Shukla, N. Pandey, On the implementation of PFSCL serializer. in *2016 3rd international conference on signal processing and integrated networks (SPIN)*. 436–440 (2016). https://doi.org/10.1109/SPIN.2016.7566734
13. K. Gupta, N. Pandey, M. Gupta, *Model and Design of Improved Current Mode Logic Gates* (Springer, Singapore, 2020)

14. K. Gupta, P. Shukla, N. Pandey, On the implementation of PFSCL adders. in 2016 *Second International Innovative Applications of Computational Intelligence on Power, Energy and Controls with their Impact on Humanity (CIPECH)*. 287–291 (2016). https://doi.org/10.1109/CIPECH.2016.7918784
15. K. Gupta, R. Sridhar, J. Chaudhary, N. Pandey, M. Gupta, Performance comparison of MCML and PFSCL gates in 0.18 µm CMOS technology. in *2011 2nd international conference on computer and communication technology (ICCCT-2011)*. 230–233 (2011). https://doi.org/10.1109/ICCCT.2011.6075165
16. K. Gupta, R. Sridhar, J. Chaudhary, N. Pandey, M. Gupta, New low-power tristate circuits in positive feedback source-coupled logic. J. Electr. Comput. Eng. (2011). https://doi.org/10.1155/2011/670508
17. H. Hassan, M. Anis, M. Elmasry, MOS current mode circuits: analysis, design, and variability. IEEE Trans. Very Large-Scale Integr. (VLSI) Syst. **13**(8), 885–898 (2005). https://doi.org/10.1109/TVLSI.2005.853609
18. Y. Ho, G.M. Huang, P. Li, Nonvolatile memristor memory: Device characteristics and design implications. in *Proceedings of the 2009 International Conference on Computer-Aided Design*. 485–490 (2009). https://doi.org/10.1145/1687399.1687491
19. J. Hu, H. Ni, Y. Xia, High-speed low-power MCML nanometer circuits with near-threshold computing. J. Comput. **8**(1), 129–135 (2013). https://doi.org/10.4304/jcp.8.1.129-135
20. X. Hu, M.J. Schultis, M. Kramer, A. Bagla, A. Shetty, J.S. Friedman, Overhead requirements for stateful memristor logic. IEEE Trans. Circ. Syst. I Regul. Pap. **66**(1), 263–273 (2018). https://doi.org/10.1109/TCSI.2018.2861463
21. P. Jin, G. Wang, H.H.C. Iu, T. Fernando, A locally active memristor and its application in a chaotic circuit. IEEE Trans. Circ. Syst. II Exp. Briefs **65**(2), 246–250 (2017). https://doi.org/10.1109/TCSII.2017.2735448
22. S. Kiaei, S.H. Chee, D. Allstot, CMOS source-coupled logic for mixed-mode VLSI. in *IEEE International Symposium on Circuits and Systems*. 1608–1611(1990). https://doi.org/10.1109/ISCAS.1990.112444
23. R. Kozma, R.E. Pino, G.E. Pazienza (eds.), *Advances in Neuromorphic Memristor Science and Applications* (Springer, Berlin, 2012)
24. O. Krestinskaya, A.P. James, L.O. Chua, Neuromemristive circuits for edge computing: a review. IEEE Trans. Neural Netw. Learn. Syst. **31**(1), 4–23 (2019). https://doi.org/10.1109/TNNLS.2019.2899262
25. S. Kvatinsky, D. Belousov, S. Liman, G. Satat, N. Wald, E.G. Friedman, U.C. Weiser, MAGIC—Memristor-aided logic. IEEE Trans. Circ. Syst. II: Exp. Briefs **61**(11), 895–899 (2014). https://doi.org/10.1109/TCSII.2014.2357292
26. S. Kvatinsky, G. Satat, N. Wald, E.G. Friedman, A. Kolodny, U.C. Weiser, Memristor-based material implication (IMPLY) logic: design principles and methodologies. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **22**(10), 2054–2066 (2013). https://doi.org/10.1109/TVLSI.2013.2282132
27. S. Kvatinsky, N. Wald, G. Satat, A. Kolodny, U.C. Weiser, E.G. Friedman, MRL—Memristor ratioed logic. in *2012 13th International Workshop on Cellular Nanoscale Networks and their Applications*. 1–6 (2012). https://doi.org/10.1109/CNNA.2012.6331426
28. G. Liu, S. Shen, P. Jin, G. Wang, Y. Liang, Design of memristor-based combinational logic circuits. Circ. Syst. Signal Process. **40**(12), 5825–5846 (2021). https://doi.org/10.1007/s00034-021-01770-1
29. B. Mohammad, D. Homouz, H. Elgabra, Robust hybrid memristor-CMOS memory: modeling and design. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **21**(11), 2069–2079 (2013). https://doi.org/10.1109/TVLSI.2012.2227519
30. J.M. Musicer, J. Rabaey, MOS current mode logic for low power, low noise CORDIC computation in mixed-signal environments. in *Proceedings of the 2000 international symposium on Low power electronics and design*. 102–107 (2000). https://doi.org/10.1145/344166.344532
31. Y.V. Pershin, M. Di Ventra, Experimental demonstration of associative memory with memristive neural networks. Neural Netw. **23**(7), 881–886 (2010). https://doi.org/10.1016/j.neunet.2010.05.001
32. J. Rofeh, A. Sodhi, M. Payvand, M.A. Lastras-Montaño, A. Ghofrani, A. Madhavan, L. Theogarajan, Vertical integration of memristors onto foundry CMOS dies using wafer-scale integration. in *2015 IEEE 65th Electronic Components and Technology Conference (ECTC)*. 957–962 (2015). https://doi.org/10.1109/ECTC.2015.7159710
33. V. Saminathan, K. Paramasivam, Design and analysis of low power hybrid memristor-CMOS based distinct binary logic nonvolatile SRAM cell. Circ. Syst. **7**(3), 119–127 (2016). https://doi.org/10.4236/cs.2016.73012

34. A. Singh, Design and analysis of memristor-based combinational circuits. IETE J. Res. **66**(2), 182–191 (2020). https://doi.org/10.1080/03772063.2018.1486741
35. D. Singh, K. Gupta, N. Pandey, A novel low-power nonvolatile 8T1M SRAM cell. Arab. J. Sci. Eng. **47**(3), 3163–3179 (2022). https://doi.org/10.1007/s13369-021-06035-2
36. Z. Toprak, Y. Leblebici, Low-power current mode logic for improved DPA-resistance in embedded systems. in *2005 IEEE International Symposium on Circuits and Systems*. 1059–1062 (2005). https://doi.org/10.1109/ISCAS.2005.1464774
37. A. Tyagi, N. Pandey, K. Gupta, PFSCL based linear feedback shift register. in *2016 international conference on computational techniques in information and communication technologies (ICCTICT)*. 580–585 (2016). https://doi.org/10.1109/ICCTICT.2016.7514646
38. W. Wang, C. Yakopcic, E. Shin, K. Leedy, T.M. Taha, G. Subramanyam, Fabrication, characterization, and modeling of memristor devices. in *NAECON 2014-IEEE National Aerospace and Electronics Conference*. 259–262 (2014). https://doi.org/10.1109/NAECON.2014.7045813
39. L. Xia, B. Li, T. Tang, P. Gu, P.Y. Chen, S. Yu, H. Yang, MNSIM: simulation platform for memristor-based neuromorphic computing system. IEEE Trans. Comput. Aided Des. Integr. Circ. Syst. **37**(5), 1009–1022 (2017). https://doi.org/10.1109/TCAD.2017.2729466
40. Q. Xia, W. Robinett, M.W. Cumbie, N. Banerjee, T.J. Cardinali, J.J. Yang, R.S. Williams, Memristor–CMOS hybrid integrated circuits for reconfigurable logic. Nano Lett. **9**(10), 3640–3645 (2009). https://doi.org/10.1021/nl901874j
41. L. Xie, H.A. Du Nguyen, J. Yu, A. Kaichouhi, M.Taouil, M. AlFailakawi, S. Hamdioui, Scouting logic: A novel memristor-based logic design for resistive computing. in *2017 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. 176–181 (2017). https://doi.org/10.1109/ISVLSI.2017.39
42. X. Xu, X. Cui, M. Luo, Q. Lin, Y. Luo, Y. Zhou, Design of hybrid memristor-MOS XOR and XNOR logic gates. in *2017 International Conference on Electron Devices and Solid-State Circuits (EDSSC)*. 1–2 (2017). https://doi.org/10.1109/EDSSC.2017.8126414
43. L. Yao, P. Liu, J. Wu, Y. Han, Y. Zhong, Z. You, Integrating two logics into one crossbar array for logic gate design. IEEE Trans. Circ. Syst. II Express Briefs **68**(8), 2987–2991 (2021). https://doi.org/10.1109/TCSII.2021.3071386
44. Y. Zhou, Y. Li, L. Xu, S. Zhong, R. Xu, X. Miao, A hybrid memristor-CMOS XOR gate for nonvolatile logic computation. Phys. Status Solidi (a). **213**(4), 1050–1054 (2016). https://doi.org/10.1002/pssa.201532872

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

Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Monitoring and sensing of glucose molecule by micropillar coated electrochemical biosensor via $CuO/[Fe(CN)_6]^{3-}$ and its applications

Purva Duhan [a], Deepak Kumar [b], Mukta Sharma [b], Deenan Santhiya [a], Vinod Singh [b],*

[a] *Department of Applied Chemistry, Delhi Technological University, Delhi 110 042, India*
[b] *Department of Applied Physics, Delhi Technological University, Delhi 110 042, India*

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Oxidation
Ferricyanide
Adsorption
Electrodes
Enzyme

ABSTRACT

In recent years, biosensing for the different types of substances affecting our day-to-day life has been evolving to a great extent. The sensing of the glucose level in food as well as the detection of blood sugar levels, are two essential steps for a healthy life. The glucose molecules, on oxidation in the presence of Ferricyanide, generate a current when connected to electrodes. In this paper, the method of current generation due to the oxidation of glucose molecules has been used and a sensor based on the principle of electrochemical sensing has been designed using COMSOL Multiphysics. Furthermore, the variation of current in the range $0 - 3\ \mu A$ with the concentration of the adsorbed glucose molecules in the range $0 - 100\ mgdl^{-1}$ on the sensing surface as well as time has been analyzed to achieve a sensitivity of $37.88\ \mu Amg^{-1}dl$ for the sensor. The calculated value of sensitivity for the designed sensor is $37.88\ \mu Amg^{-1}dl$. The high sensitivity of the sensor is the key factor for its wide range of applications in the field of biosensing.

Selection and peer-review under responsibility of the scientific committee of the Fourth International Conference on Recent Advances in Materials and Manufacturing 2022.

1. Introduction

Recently, significant advances are being made in the field of biosensors. The electrochemical redox reactions form the basic principle of the process for the detection of the substances or chemicals present in the human body, food, and all the things that impact a person's daily life. Biosensing technology has been developed to a great extent for the detection of protein, DNA, and numerous hurtful acids that affect the human body [1]. Glucose is a vital source of energy and is the end product of the digestion of carbohydrates. However, an excess amount of glucose in blood can cause severe health problems [2,3]. Blood glucose testing is a very serious issue and important for diabetics as well as non-diabetics to keep a check on their health and take steps for maintaining it. A precise detection of blood glucose levels is very important for the diabetes patients to regulate the dose of their medicines or injections. For non-diabetics, the detection of their blood sugar level is an important step to stay fit, maintain the balanced diets and also to prevent diabetes. The concentration of glucose in blood is in the range of $2 - 30\ mmol/L$. In fact in a human being's breath, about $21 - 0.5\ ppm$ of glucose is observed [4]. This makes glucose detection crucial to regulate the blood sugar level and to maintain an appropriate food intake.

Currently, in order to detect glucose level in blood, the most common detectors are the blood glucose test strips. These strips react with blood and oxidize the glucose present in the blood to produce gluconic acid [5]. The oxidation of glucose molecules leads to the production of ions that add to the current level of the sample. However, these glucose test strips do not give precise results and cannot be used for multiple times. Other than the test strips, there are semiconductor-based biosensors that can be used to detect glucose. However, semiconductor based sensors need the fabrication of a bio-electrode and the materials used for the synthesis lack stability. This makes the semiconductor based sensor less accurate and more complicated for day-to-day use and for essential applications [6–8]. To achieve accurate results, a flexible structured device is needed. Additionally, there is a need for a device that can detect the presence of glucose on microscale level such that even a tiny amount of glucose can be detected [9]. Electrochemical sensors turn out to be a potential candidate to meet these requirements. An electrochemical sensor works on the principle of oxidation or reduction of a target product. Following the oxidation or reduction reactions, the target product is detected

* Corresponding author.
E-mail address: vinodsingh@dtu.ac.in (V. Singh).

https://doi.org/10.1016/j.matpr.2023.03.059
2214-7853/Copyright © 2023 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the Fourth International Conference on Recent Advances in Materials and Manufacturing 2022.

Please cite this article as: P. Duhan, D. Kumar, M. Sharma et al., Monitoring and sensing of glucose molecule by micropillar coated electrochemical biosensor via $CuO/[Fe(CN)_6]^{3-}$ and its applications, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2023.03.059

and its concentration is measured. The results from current literature have shown electrochemical sensors with a higher electro catalytic activity, high sensitivity and excellent selectivity [1]. The maximum sensitivity of 59.16 mV/pH has been observed by the electrochemical sensor fabricated by Li et al. [10]. The study has been performed on pH dependent threshold voltages. In the current research done in the field of biosensors, it has been shown how these electrochemical sensors play a revolutionary role in glucose level monitoring [11,12].

In this paper, an electrochemical sensor array of micropillars for the adsorption of antigens in aqueous solutions has been designed using COMSOL Multiphysics. The presence of capillaries is necessary for letting the fluids or the foods in and an outlet that is further connected to the electrodes. The electrodes pick up the generated current due to the oxidation of glucose and the increase in current is detected which further gives precise information about the presence of the glucose in the tested materials. The surface for the deposition of the glucose analyte is coated with CuO which acts as the catalyst and a layer of ferricyanide acts as the glucose oxidase responsible for the redox reaction. On oxidation of glucose, the current is generated which is further supplied to the electrodes of the sensor. The current follows a linear relationship with the concentration of glucose. The graphs for the current and the concentration of the glucose have been plotted to show the detection of the glucose present in the analyte taken. Also, the variation in the surface coverage of the glucose molecules with time has been shown to denote the adsorption of the molecules on the catalytic surface. The maximum concentration of glucose level detected by the sensor is with the order of 10^2 $\mathrm{mg\,dl^{-1}}$. The calculated value of sensitivity for the designed sensor is 37.88 $\mu\mathrm{Amg^{-1}dl}$.

2. Methodology

2.1. Theoretical simulation

The 3-Dimensional model has been built for the electrochemical sensor with the electrochemical module under the laminar flow with time-dependent study in COMSOL Multiphysics. The geometry, parameters, geometrical non-linearity and materials for the pillar surfaces have been added. Under the physics module of the laminar flow, the transport of diluted species and surface reactions have been taken as the physics interface.

The geometry is set up for the sensors by taking the z- axis distance between the pillars as 0.002 m, the x-axis distance between the pillars as 0.0016 m and the dimensions of the cell as 0.012 m × 0.001 m × 0.0069 m. The maximum radius allowed for the pillar is 5.9031E−4 m. The following figure shows the geometry of the designed sensor (Fig. 1).

The designed sensor structure is further meshed up to note the finer details of the surface for the absorption of the molecules and the catalytic as well as the reactant layers on the surface. The mesh structure for the biosensor is shown in Fig. 2,

3. Mathematical operations

The oxidation of the glucose in the presence of ferricyanide is represented by the following reaction [11]:

$$Glucose + Ferricyanide \rightarrow Gluconic\ \ acid + Ferrocyanide \tag{1}$$

The above reaction is a complex reaction with the complex Ferrocyanide ($[Fe(CN)6]^{4-}$) being formed. In the above reaction, the oxidation of glucose is observed and the resulting product is gluconic acid ($C_6H_{12}O_7$). Also, the reduction of Ferricyanide to Ferrocyanide is observed in the reaction.

The rate of this reaction can be calculated from the Michaelis-Menten equation which has been shown in Eq. (2) [12],

$$r = \frac{C_{glucose} \times V_{max}}{1 + (K_m \times C_{glucose})} \tag{2}$$

where $C_{glucose}$ is the concentration of glucose, V_{max} is the maximum rate of the reaction and K_m is the Michaelis-Menten constant.

Inside the software, for the electroanalysis of the sensor, the battery and fuel cells module has been taken under the module of electrochemistry. The following boundary conditions have been applied for the analysis of the results [13],

$$\frac{\partial C_{glucose}}{\partial t} + \nabla \cdot \left(-D_{A_m} \nabla_{C_{glucose}}\right) + u \cdot \nabla_{C_{glucose}} = 0 \tag{3}$$

where $\frac{\partial C_{glucose}}{\partial t}$ is the rate of change of glucose concentration, D_{A_m} is the diffusion coefficient and u is the velocity vector of the glucose molecules.

The flux of the electric field is related to the absorption and desorption rates of the glucose molecules by the following equation [14],

$$\varnothing = -r_{abs} + r_{des} \tag{4}$$

where r_{abs} is the rate of absorption and r_{des} is the rate of desorption.

The current density for the electroanalytic sensor is calculated from the Butler–Volmer equation for oxidation,

$$J = j_e\left[exp(1-\alpha)\frac{F\eta}{RT}exp\left(\frac{-\alpha F}{RT}\right)\right] \tag{5}$$

where j_e is the equilibrium current density, F is the Faraday constant, α is the transfer coefficient, η is the over potential supplied at the working electrode, R is the gas constant and T is the temperature of the cell which is kept constant.

The current is further calculated from the current density as given in Eq. (6) and the curve is plotted between the current and concentration of glucose [15].

$$I = J \times A \tag{6}$$

where A is the area of the micro pillars cell.

The sensitivity of the designed sensor can further be calculated from Eq. (7),

$$S = \frac{\partial I}{\partial C_{Glucose}} \tag{7}$$

where ∂I is the change in the current and $\partial C_{Glucose}$ is the change in concentration of glucose molecules.

4. Results and discussion

The input for the sensor are shown in Table 1 which include the parameters for the time dependent study and the surface analysis of the glucose molecules.

The parameters are designed and the sensor study model is simulated for the velocity, pressure, concentration with the slices, concentration for the surface and the concentration adsorbed species. Figs. 3–5 show the electric field vector along with the 3-D variation time scale for the model and represent the streamline of the glucose molecules along the surface of the biosensor.

It is observed that the rate of the adsorption of the molecules of glucose increases as we move away from the centre of the cell toward the walls. The streamline velocity and the pressure of the molecules is increasing and the concentration is also further increasing with time and with the increase of the displacement of the position of the pillars of the array that provide the surface for the adsorption of the glucose molecules. Also, the molecules once adsorbed remain at the surface of the pillars along the walls

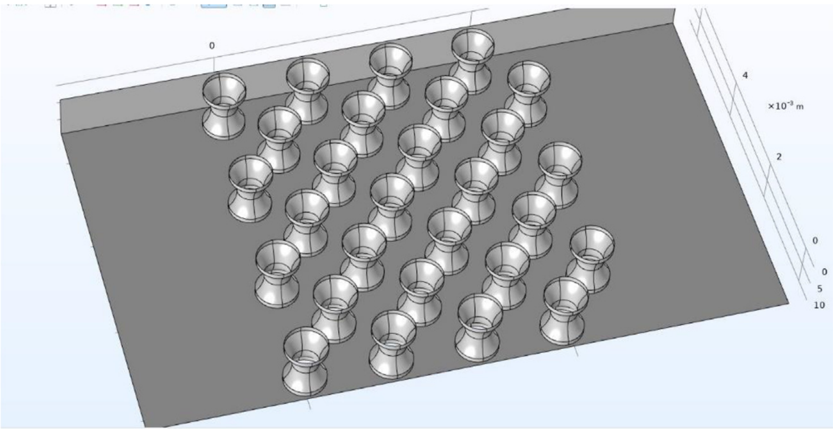


Fig. 1. The geometrical array of the micropillars coated with a layer of CuO and Ferricyanide for the absorption of the glucose molecules.

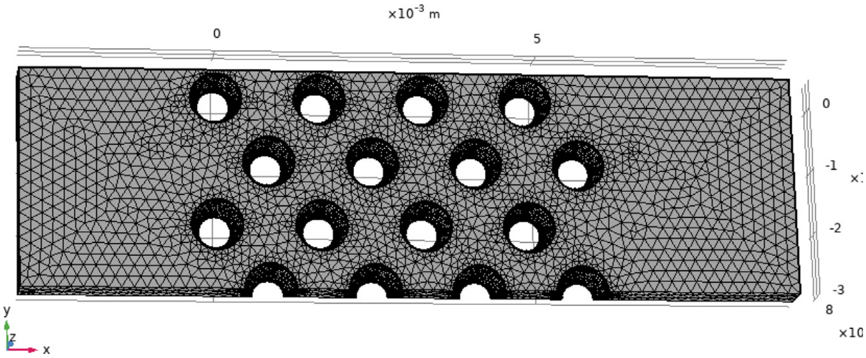


Fig. 2. The mesh structure of the array of the micropillars mounted inside a cell in the sensor.

for a longer time as compared to the rest of the pillars which makes the rate of desorption lower for them. The stream field consists of velocity distribution which creates a field such that the pillars that are present near the walls experience a positive change in their absorption such that the absorption level increases to a great extent. Also, the maximum adsorption level is affected by the rate of desorption of the pillars near the wall because they took longer for desorption.

The surface fraction of the absorbed glucose molecules firstly increases with time to the point of surface saturation after which the pillars start the process of desorption of the molecules which decreases the surface fraction of the glucose molecules up to a level at which the surface fraction of the molecules become constant and very low in value. The increase in the fraction of absorbed glucose molecules react to the maximum limit near the time $38 - 40$ s. This is the saturation point of the absorption because the maximum area of the surface is covered with the glucose molecules. After this point, the desorption of the glucose molecules starts and the surface fraction decreases with time. The distribution of surface fraction of the molecules varying with time in seconds in shown in Fig. 6.

The current through the sensor that is generated due to the oxidation of the glucose molecules, is proportional to the concentration of the glucose molecules. With the increase in the concentration of the glucose molecules, an increase in the current is seen as shown in Fig. 7.

When the time period increases from 0 to 10 s, 20 s and so on, there is a linear increase in the absorption of the glucose molecules on the surface of the pillars. With the increase in the absorption, the current also increases which gives a linear increment in the current generated by the sensor with time (Fig. 8).

Apart from the concentration of the glucose molecules, the current produced as a result of the oxidation of glucose also depends on the thickness of the layer of enzyme on the top of the pillars. As the thickness of the enzyme layer is increased, the process is catalysed more and the initial amount of the current produced is increased as shown in Fig. 9.

The sensitivity of the sensor can be calculated by considering the variation of current with respect to the concentration of the glucose by Eq. (7)

$$S = 37.88\ \mu\mathrm{Amg}^{-1}\mathrm{dl}$$

5. Conclusion

The electrochemical biosensor for the detection of glucose molecules has been designed with a structure with an array of pillars with a layer of Ferricyanide on its surface. The oxidation of the glucose molecules and the generation of the current has been represented. The 3-D modelling for the streamline velocity, pressure and concentration has also been shown. The fraction of surface molecules has been plotted with time. As the time increases, the surface fraction of glucose molecules first increases due to the absorption. Then, the saturation point of absorption is observed. After the saturation point, the desorption of the glucose molecules start and the surface fraction of the glucose molecules decreases [16]. With increase in concentration of the analyte with glucose, the best fit curve shows an increase in the current passing the electrode. The sensitivity of the designed sensor is $37.88\mu\mathrm{Amg}^{-1}\mathrm{dl}$. The sensor accurately detects the presence of glucose in food items, and also detects the glucose level of the sample. The designed sen-

Table 1
Parameters for the analysis and the designing of the sensor.

Name	Expression	Value	Description
k_{ads}	10^{-2} [m/s]	0.01 m/s	Forward rate constant
k_{des}	0.5 [mol/m^2/s]	0.5 mol/(m^2·s)	Backward rate constant
D	5×10^{-9}[m^2/s]	5E−9 m^2/s	Gas diffusivity
k_f	2×10^{-7} [mol/m^2/s]	2E−7 mol/(m^2·s)	Forward rate constant
k_r	4×10^{-8}[mol/m^2/s]	4E−8 mol/(m^2·s)	Reverse rate constant
u_{in}	2×10^{-4} [m/s]	2E−4 m/s	Inlet velocity
N_w	4	4	Number of pillars across
R_{pillar}	0.4 [mm]	4E−4 m	Radius of pillar
R_c	6×10^{-4} [m]	6E−4 m	Radius of carve-out
d_c	1.5×10^{-4} [m]	1.5E−4 m	Cut depth of carving
x_c	$R_{pillar} + R_c - d_c$	8.5E−4 m	x-position of carving circle
R_{c1}	6×10^{-4} [m]	6E−4 m	Radius of carve-out
d_{c1}	1.5×10^{-4} [m]	1.5E−4 m	Cut depth of carving
x_{c1}	$R_{pillar} + R_c - d_c$	8.5E−4 m	x-position of carving circle
W_{tot}	6.8×10^{-3} [m]	0.0068 m	Total width of pillar grid
L_{tot}	5.6×10^{-3}	0.0056	Total length of pillar grid (outer row)
d_{wall}	0.5×10^{-4} [m]	5E−5 m	Distance from pillar edge to cell side wall
d_z	$\frac{(W_{tot} - 2R_{pillar})}{(N_w - 1)}$	0.002 m	z-spacing between pillars
d_x	$\frac{(L_{tot} - 2R_{pillar})}{(N_w - 1)}$	0.0016 m	x-spacing between pillars
W_{box}	12×10^{-3} [m]	0.012 m	Width of cell
D_{box}	10^{-3} [m]	0.001 m	Depth of cell
H_{box}	6.9×10^{-3} [m]	0.0069 m	Height of cell
d_{pillar}	$\frac{\sqrt{d_z^2 + d_x^2}}{2} - 2R_{pillar}$	4.8062E−4 m	Current closest distance between two pillar edges
$d_{pillarallowed}$	0.1×10^{-3} [m]	1E−4 m	Allowed minimum distance between two pillar edges
$R_{\max allowed}$	$\frac{\sqrt{d_z^2 + d_x^2}}{4} - 2d_{pillarallowed}$	5.9031E−4 m	Allowed maximum pillar radius
c_{00}	400 [mol/m^3]	400 mol/m^3	Injection pulse amplitude
sol_{tol}	0.01	0.01	Relative tolerance of solvers
end_{time}	150	150	Simulation end time
$d_{timevalue}$	0.5	0.5	Dimensionless time for concentration plot
t_{value}	0	0	Time for time dependent plots

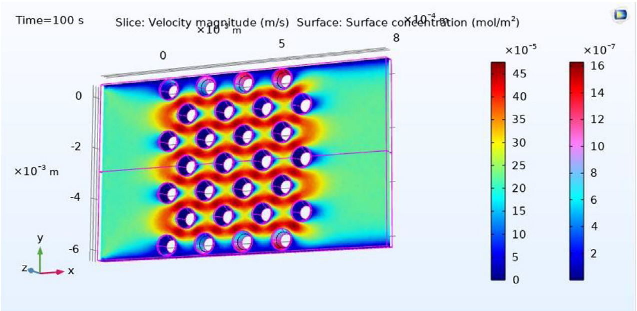


Fig. 3. Velocity of the glucose molecules and the surface concentration variation 3-D model with high velocity of the molecules near the pillars along the walls compared to the pillars located at the middle.

sor is equally efficient in comparison to other experimentally designed electrochemical glucose sensors and gives accurate results. Experimentally, in current literature, a sensitivity of

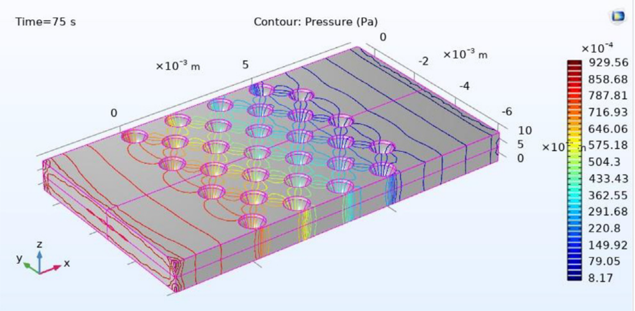


Fig. 4. Contour pressure model for the biosensor showing streamline for the pressure along the different pillars of the array at time interval of 75 s.

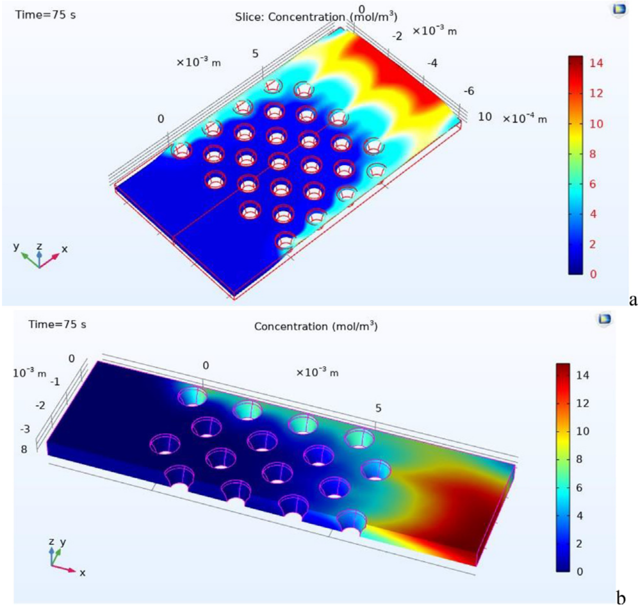


Fig. 5. The concentration in mole per metre cube is shown over the surface of the pillars with more concentration at the surfaces of the pillars that are along the walls of the cell. (a) The whole cell is seen for the variation of the concentration with a colour bar legend. (b) The half portion of the cell is shown for molar concentration with the half portions of the pillars at the middle.

1536.80 μA mM^{-1} cm^{-2} has been observed by the glucose biosensor.[17].

CRediT authorship contribution statement

Purva Duhan: Conceptualization, Data curation. **Deepak Kumar:** Conceptualization, Methodology, Resources, Data curation, Writing – original draft. **Mukta Sharma:** Conceptualization, Methodology, Resources, Data curation, Writing – original draft. **Deenan Santhiya:** Investigation, Supervision. **Vinod Singh:** Conceptualization, Writing – review & editing, Investigation, Supervision.

Data availability

No data was used for the research described in the article.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing

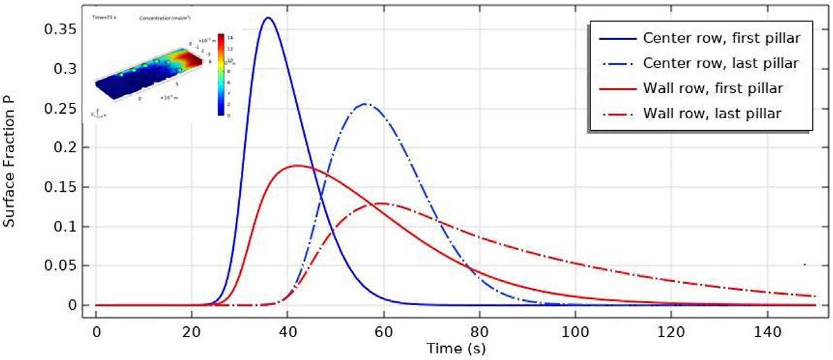


Fig. 6. The surface fraction of the absorbed glucose molecules for the pillars placed at the different positions varying with respect to time (s).

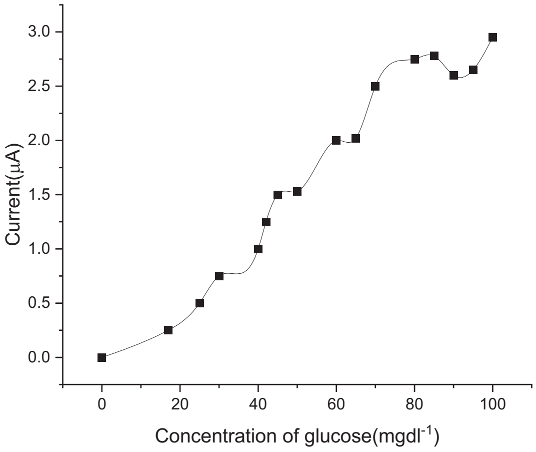


Fig. 7. The variation of current (μA) with the concentration of glucose (mgdl^{-1}).

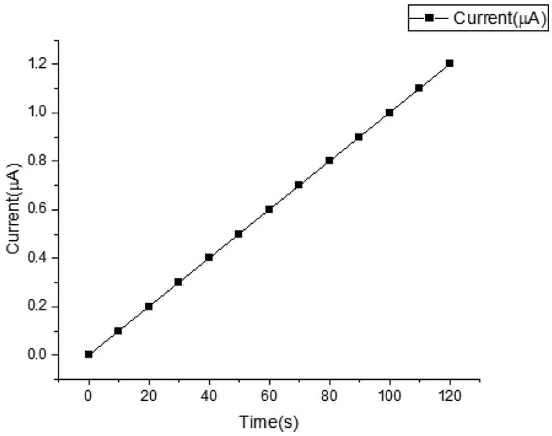


Fig. 8. Variation of current (μA) with the time taken for the adsorption of the glucose molecules.

interests: “Vinod Singh reports was provided by Delhi Technological University. Vinod Singh reports a relationship with Delhi Technological University that includes: employment and non-financial support”.

Acknowledgement

The authors, Purva Duhan, Deepak Kumar and Mukta Sharma would like to thank Prof. Vinod Singh and Dr. Deenan Santhiya for their support for this work.

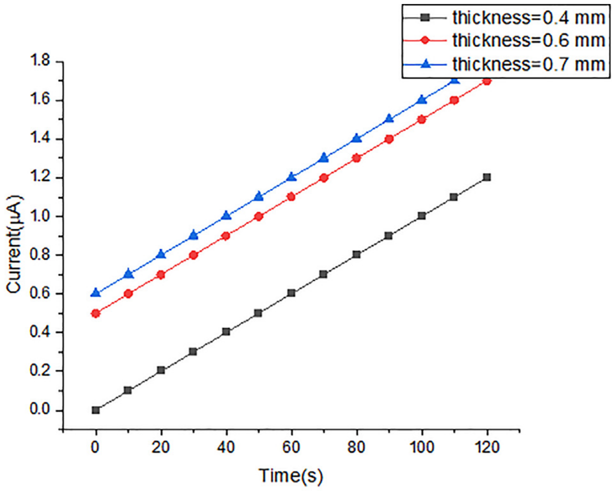


Fig. 9. The current versus time variation for different thicknesses of the enzyme layer over the pillars.

References

[1] Q. Li, Z. Shao, T. Han, M. Zheng, H. Pang, A High-Efficiency Electrocatalyst for Oxidizing Glucose: Ultrathin Nanosheet Co-Based Organic Framework Assemblies, ACS Sustainable Chem. Eng. 7 (9) (2019) 8986–8992, https://doi.org/10.1021/acssuschemeng.9b01148.

[2] Umesh Yadav, Ravindra Sarje, A.D. Shaligram, S.A. Gangal, Design, simulation, fabrication and testing of Electrochemical NO2 gas sensor, in: Proceedings of the 2015 2nd International Symposium on Physics and Technology of Sensors, 10.1109/ISPTS.2015.7220127.

[3] Y. Ao, J. Ao, L. Zhao, L. Hu, F. Qu, B. Guo, X. Liu, Hierarchical Structures Composed of Cu(OH)2 Nanograss within Directional Microporous Cu for Glucose Sensing, Langmuir 38 (45) (2022) 13659–13667, https://doi.org/10.1021/acs.langmuir.2c01300.

[4] G. Li, D. Wen, Sensing nanomaterials of wearable glucose sensors, Chin. Chem. Lett. 32 (1) (2021) 221–228, https://doi.org/10.1016/j.cclet.2020.10.028.

[5] E. Sehit, Z. Altintas, Significance of nanomaterials in electrochemical glucose sensors: An updated review (2016–2020), Biosens. Bioelectron. 159 (2020), https://doi.org/10.1016/j.bios.2020.112165.

[6] M.V. Varsha, G. Nageswaran, Review—2D Layered Metal Organic Framework Nanosheets as an Emerging Platform for Electrochemical Sensing, J. Electrochem. Soc. 167 (13) (2020), https://doi.org/10.1149/1945-7111/abb4f5.

[7] M. Yuan, X. Guo, Y. Liu, H. Pang, Si-based materials derived from biomass: synthesis and applications in electrochemical energy storage, J. Mater. Chem. A 7 (39) (2019) 22123–22147, https://doi.org/10.1039/C9TA06934H.

[8] Y. Wang, Y. Wang, L. Zhang, C.S. Liu, H. Pang, PBA@POM Hybrids as Efficient Electrocatalysts for the Oxygen Evolution Reaction, Chem. – Asian J. 110 (2019), https://doi.org/10.1002/asia.201900791.

[9] A. Nováková, L. Schreiberová, I. Schreiber, Study of dynamics of glucose-glucose oxidase-ferricyanide reaction, Russ. J. Phys. Chem. 85 (2011) 2305–2309, https://doi.org/10.1134/S003602441113019X.

[10] V. Singh, D. Kumar, M. Sharma, Gold/ZnO Interface-Based D-Shaped PCF Surface Plasmon Resonance Sensor with Micro-Openings, Analytic Designing, and Some Applications, in: K. Geetha, F.M. Gonzalez-Longatt, H.M. Wee (Eds.),

Recent Trends in Materials. Springer Proceedings in Materials, vol. 18, Springer, Singapore, https://doi.org/10.1007/978-981-19-5395-8_27.

[11] D. Kumar, M. Sharma, V. Singh, Surface Plasmon Resonance implemented Silver thin film PCF sensor with multiple-Hole microstructure for wide ranged refractive index detection, Mater. Today Proc. 62 (part 12) (2022) 6590–6595, https://doi.org/10.1016/j.matpr.2022.04.598.

[12] Deepak Kumar, Khurana Madhur, Mukta Sharma, Vinod Singh, Analogy of gold, silver, copper and aluminium based ultra-sensitive surface plasmon resonance photonic crystal fiber biosensors, Materials Today: Proceedings (2023), https://doi.org/10.1016/j.matpr.2023.02.319.

[13] H. Mazhab-Jafari, L. Soleymani, R. Genov, 16-channel CMOS impedance spectroscopy DNA analyzer with dual-slope multiplying ADCs, IEEE Trans. Biomed. Circuits Syst. 6 (5) (2012) 468–478, https://doi.org/10.3390/s17010074.

[14] Y. Shi, J. Wang, S. Li, B. Yan, H. Xu, K. Zhang, Y. Du, The Enhanced PhotoElectrochemical Detection of Uric Acid on Au Nanoparticles Modified Glassy Carbon Electrode, Nanoscale Res. Lett. (2017) 12–455, https://doi.org/10.1186/s11671-017-2225-3.

[15] S. Qi, B. Zhao, H. Tang, X. Jiang, Determination of ascorbic acid, dopamine, and uric acid by a novel electrochemical sensor based on pristine graphene, Electrochim. Acta 161 (2015) 395–402, https://doi.org/10.1016/j.electacta.2015.02.116.

[16] L.G. Gómez-Mascaraque, S.C. Pinho, Microstructural Analysis of Whey/Soy Protein Isolate Mixed Gels Using Confocal Raman Microscopy, Foods 10 (9) (2021) 2179, https://doi.org/10.3390/foods10092179.

[17] Z. Haghparas, Z. Kordrostami, M. Sorouri, et al., Highly sensitive non-enzymatic electrochemical glucose sensor based on dumbbell-shaped double-shelled hollow nanoporous CuO/ZnO microstructures, Sci. Rep. 11 (2021) 344, https://doi.org/10.1038/s41598-020-79460-2.

Wireless Personal Communications
https://doi.org/10.1007/s11277-023-10324-4

Multi-view Multi-modal Approach Based on 5S-CNN and BiLSTM Using Skeleton, Depth and RGB Data for Human Activity Recognition

Rahul Kumar[1] · Shailender Kumar[1]

Accepted: 25 February 2023

Abstract
Recognition of human activity is a challenging issue, especially in the presence of multiple actions and multiple scenarios. Therefore, in this paper, multi-view multi-modal based human action recognition (HAR) is proposed. Here, initially, motion representation of each image such as Depth motion maps, motion history images, and skeleton images are created from depth, RGB, and skeleton data of RGB-D sensor. After the motion representation, each motion is separately trained by using a 5-stack convolution neural network (5S-CNN). To enhance the recognition rate and accuracy, the skeleton representation is trained using a hybrid 5S-CNN and Bi-LSTM classifier. Then, decision-level fusion is applied to fuse the score value of three motions. Finally, based on the fusion value, the activity of humans is identified. To estimate the efficiency of the suggested 5S-CNN with the Bi-LSTM method, we conduct our experiments using UTD-MHAD. Results show that the suggested HAR method attained better than other existing approaches.

Keywords Recognition · Human activity · Multi-view · Motion history image · Depth motion maps · Skeleton · 5S-CNN · Bi-LSTM · Decision · Fusion

1 Introduction

HAR is a functioning exploration territory in computer vision. As a rule, it is generally utilized in human–computer interaction, clinical, video observation, and so on. For the current HAR process, video recordings are captured by the video camera. For recognition, spatial–temporal attributes are mainly used [1]. With the fast improvement of imaging innovation, nowadays depth cameras for example Microsoft Kinect cameras are used. By using this camera, we can capture, RGB, depth, and skeleton. One of the upsides of profound information and skeletal information contrasted with conventional RGB information is that they are less touchy to changes in lighting conditions.

✉ Rahul Kumar
rahuldtucs@gmail.com

[1] Computer Science and Engineering, Delhi Technological University, Delhi 110042, India

Published online: 12 March 2023

Earlier techniques for action recognition dependent on depth information utilize explicit carefully assembled include descriptors. The benefits of exclusively prepared CNNs on DMMs are introduced in [2, 3]. A DMM is an unequivocally shaped movement portrayal picture, which is built from crude depth outlines. This is like the development of MHI and optical stream portrayals made from RGB outlines. Autonomous utilization of skeletal information, for action recognition dependent on very much planned, handmade element descriptors is introduced in [4, 5]. In [6] an endeavor is made to develop surface pictures from skeleton joint groupings.

In the multistage processing of the visual cortex [7], deep convolution neural networks (ConvNets) can naturally take in separating highlights from information and are utilized in assignments identified with picture characterization, recognition, division, discovery, and recovery [8, 9]. ConvNets utilizes strategies to scale organizations to a great many boundaries and takes in boundaries from an enormous named information base. These days, there is a critical improvement in the HAR dependent on the individual utilization of RGB, depth, and skeletal information. Nonetheless, joining these interesting viewpoints in various approaches to additional upgrade recognition stays a test for specialists. In [10], human action recognition was made utilizing RGB and depth information. In [11], skeletal information and idleness information are intertwined at both the component level and choice level. In [12], depth movement maps with skeleton information are joined together to zero in on HAR. To improve the performance of the HAR system, in this paper multi-view multi-modal HAR system is proposed.

2 Literature Survey

Many of the researchers focus on human action recognition using multi-modal. Among them few research works are discussed here;

Verma et al. [13] developed HAR using a deep learning algorithm based on multi-modals such as RGB and Skeleton. Here, initially, MHI and motion energy image (MEI) are extracted from the input RGB video. Then, three views of the skeleton image were extracted using skeleton intensity. After the motion extraction process, the extracted features are trained using LSTM. Finally, based on the softmax score value, the decision is carried out. This approach was tested using three famous datasets namely, UTD-MHAD, CAD-60, and NTU-RGB + D120. Moreover, Wang et al. [14] aimed to develop HAR. For this, weighted hierarchical DMM (WHDMM) and three-channel deep CNN (3ConvNets) were utilized. An initially different viewpoint of depth can be captured using a 3D point. This is used to train the ConvNets. Then, temporal scales are constructed to form an AHDMM. For simulation, they utilized three types of datasets.

In [15], Chen et al. aimed to develop skeleton, RGB, and depth fusion-based HAR. Similarly, in [16], Escobedo et al. had developed HAR based on skeleton and depth data. In [17], Gaglio et al. they had aimed to develop HAR based on machine learning. Here, they utilized skeleton joint data. Khaire et al. [18] had developed a combined CNN stream-based HAR system. Here, three types of motion representation images namely, MHI, DMMs, and skeleton images are designed from input videos. After the extraction process, each data was individually trained using CNN classifier. Finally, from the classification results, one score value was obtained. Based on the score value recognition has been done. The efficiency of the suggested approach was analyzed by using three different datasets.

Moreover, Guo et al. [19] had created human activity acknowledgment by mutually abusing video and Wi-Fi pieces of information. They influence the way that Wi-Fi signals convey discriminative data of human activities, which was strong to optical limits. To approve this imaginative idea, they consider a down-to-earth system for HAR and arrangement a dataset containing both video clasps and Wi-Fi Channel State Information of human activities. The 3D convolutional neural organization was utilized to remove the video highlights and the measurable calculations were utilized to extricate radio highlights. An old-style direct help vector machine was utilized as the classifier after the video and radio element combination.

Similarly, Tran et al. [20] had aimed to develop hand gesture recognition. To attain the recognition process, here, they utilized multi-modal streams. Three types of stream depth, RGB, and optical flow are used for the recognition process. These streams are fed to the feature extraction process. Then these features are fed to the classifier to classify a different activity. For simulation different gesture dataset was used. In [21], Nie et al. had aimed to develop emotion recognition using a multi-layer LSTM classifier. Using this classifier, they can easily recognize the activity. But due to the gap between video frames, this method can't able to accurately find out the activity. Khowaja et al. [12] had aimed to develop HAR based on deep cross-modal learning. Here, they utilized RGB and optical flow was used. For experimental analysis, UCF101 and HMDB51 datasets are utilized.

3 Problem Statement and Contributions

In [22], Pratishtha et al. had used the advantages of the skeleton joint and RGB video. In the approach, the authors trained the CNN using RGB data as well as the skeleton data was processed using CNN and LSTM network. Although the authors had achieved better accuracy of recognition, the computation complexity is heavy as they have presented four convolutional layers and three fully connected layers after combining the features. So, to reduce computational complexity and maximize the learning process, we present the following contributions in this paper.

- Initially, we construct DMMs, MHI, and skeleton images from depth, RGB, and skeletal data of the RGB-D sensor.
- The extracted images are trained on individual CNN with 5-stacked convolutional layers. Skeleton and DMM images are trained on 5S-CNN for multi-view such as top, front, and side.
- As CNN is used to learn spatial information, BiLSTM is used for temporal dependency in the training process of skeleton images. After extracting features from each view of Skelton images using 5S-CNN, these features are combined and given as input to the BiLSTM
- The output score from each model is fusion using a weighted product model (WPM). From the fusion output, the action of humans is recognized.
- The efficiency of the suggested scheme is evaluated based on accuracy, F-score, precision, and recall.

4 Proposed Action Recognition System Model

Recognizing human activity from video footage is a challenging mission due to, partial opacity, background clutter, vision, size changes, appearance, and lighting. Numerous functions require Recognition systems, such as human–computer communication, video surveillance systems, and robotics for human behavior. Previous approaches often use only the Motion History image and depth map for HAR. It does not provide maximum accuracy for the Recognition system. To avoid the problem, in this paper multi-view and multi-modal-based automatic human action recognition is proposed. Here three modalities are used for recognition systems such as MHI, DMMs, and skeleton images. The overall concept of the proposed approach is given in Fig. 1. In the proposed approach, motion representation images, namely, DMMs, MHI, and skeleton images are extracted from depth, RGB, and skeletal data of RGB-D sensor. The DMMs and MHI are separately trained by a 5S-CNN. Similarly, the skeleton multi-view images are trained by 5S-CNN, and this trained output is given to the input of BiLSTM. Finally, the output score from each model is fusion using WPM. From the fusion output, the action of humans is recognized.

4.1 Constructing Motion Representation Images

The main objective of this section is to select the motion representation images (features) from each input image for HAR. In this paper, three types of motions are constructed namely, MHI, skeleton images, and DMMs, from RGB, skeletal data, and depth of RGB-D sensor.

4.1.1 Constructing MHI

The MHI approach is simple and robust and is mainly used to represent movements in videos. MHI was first created in [1]. It provides temporary information about the movement in the form of an image in the videos. The MHI pixel intensity is a function of the kinetic density at that location. The MHI representation is that it can be encrypted multiple times over a range, and in this way, MHI spreads the time scale of human gestures. If the brightness of the pixels in an image is high, it means that the movement occurred very recently and if the intensity is low, the motion happened before. The MHI is evaluated by using Eq. (1).

$$M\tau_{(a,b,k)} = \begin{cases} \tau, & if\ \Theta_{(a,b,k)} = 1 \\ \max\left(0,\ M\tau_{(a,b,k-1)} - \delta\right) & otherwise \end{cases} \tag{1}$$

where $\Theta_{(a,b,k)}$ → Occurrence of motion or object in the current frame in a video; k → time; (a,b) → pixel position; δ → decay parameter.

The parameter τ controls the temporary movement. Typically, MHI is generated from a binary image, which is extracted from the frame, using a threshold ξ.

$$\Theta(q,b,k) = \begin{cases} 1 \ \ if\ D(a,b,k) \ \ \geq \ \ \ \xi \\ 0 \ \ otherwise \end{cases} \tag{2}$$

$$D(a,b,k) = |I(k) - I(a,b,k \pm \Delta)| \tag{3}$$

where $I(a,b,k)$ represent the intensity value.

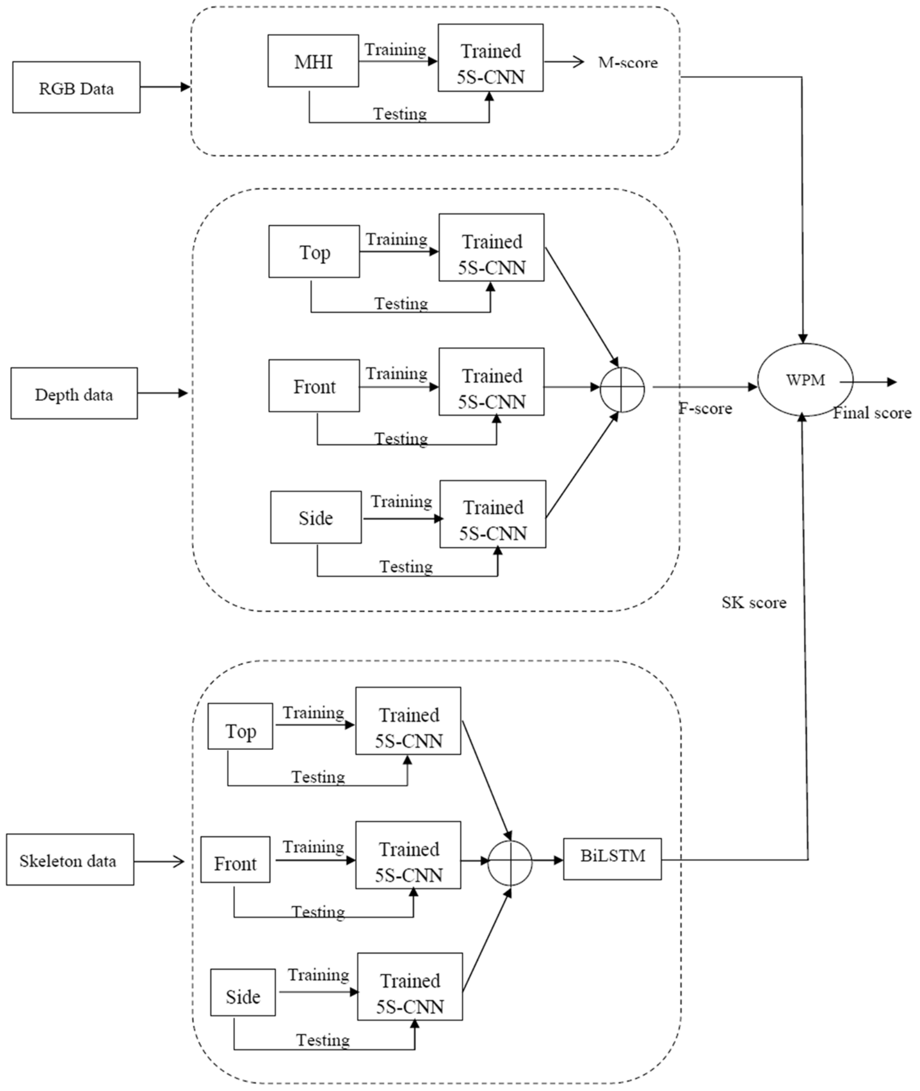


Fig. 1 Overall structure of the proposed approach

In this paper, the threshold value Δ is set as 1 for all the experiments. The final output of MHI is obtained from $M\tau_{(a,b,t)}$.

4.1.2 Constructing Depth Motion Map (DMM)

The depth map of the image is captured using an RGB-D sensor. In this 3D structure and shape of the structure is captured. To extract extra body shape, we project the depth frame into three orthogonal Cartesian planes. At that point, we set the ROI of each extended map as the jumping box of a closer view (for example non-zero) region, which is additionally standardized to a fixed size. This standardization can decrease intra-class varieties, for example, Material heights and operating lengths of various materials when performing a similar action. So, each frame creates three views such as front, side, and top view, i.e. Map^F , Map^S , *and* Map^T.Then, we evaluate the kinetic energy of each frame by evaluating the difference between the two consecutive maps and the threshold. The kinetic energy of the binary map refers to the moving parts or locations where motion occurs at each temporal interval. This gives a strong clue as to what kind of action is being taken. The DMM of the entire video can be calculated as follows;

$$DMM_{iu} = \sum_{i=1}^{N-1} \left(\left| Map_u^{j+1} - Map_u^{j} \right| > \xi \right) \tag{4}$$

where $v \in \{f, s, t\}$→Projection view; Map_u^i→Projected map of the ith frame; N→Number of frames; $\left| Map_u^{j+1} - Map_u^{j} \right| > \xi$→Binary map of motion energy.

4.1.3 Constructing Skeleton Image

The skeletal images are mainly used for the activity recognition process. The number of N-number frames and K-joints is available in the bone data. The function to function, the number of frames for action will vary. The number of frames is differing for the different subjects. The number of joints all over the activity is often unchanged. In each frame, x, y, and z is a combination of three-dimensional coordinate values. Let us consider, each joint in the skeleton as J_i, where, i represent the joints for activity and J_i is a three-dimensional vector. Since the person can be seen anywhere in the coverage area of the sensor, it is essential to centralize the integration space to fit any joint. In this paper, we consider the hip center point as the first frame for normalization. The following Eqs. (5)–(9) is used to normalize the joint coordinates.

$$D_i(S) = L_i(S) - L_0(1) \tag{5}$$

$$D_{\max} = \max\left(D_i(S)\right) \qquad \forall i \in joints \qquad s \in frames \tag{6}$$

$$D_{\min} = \min\left(D_i(S)\right) \qquad \forall i \in joints \qquad s \in frames \tag{7}$$

$$d_i(s) = \left(D_i(S) - D_{\min}\right) / \left(D_{\max} - D_{\min}\right) \tag{8}$$

where $L_i(S)$→Coordinate the ith joint in the Sth frame; $D_i(S)$→Distance between the vectors $L_i(S)$ and $J_0(1)$; $d_i(p)$→Distance between the vector $L_i(S)$ and $J_0(1)$ normalized; $L(1)$→Coordinates of the hip center in the first frame.

To create skeletal images from skeletal shots, the skeleton images are divided into five segments such as the trunk, left and right arm, and left and right leg. The size of each segment is 160×160×5, each corresponding to each view (top, side, and bottom). The image's corresponding dimension is 160×160. Let $A\left(x, y, v_i\right)_{st}$ be the image of the size x, y corresponding to the given v_i^{th} view and Sth body part and the tth frame.

$$P_x(top)_{st} = h_1 + k_1 + d_i(t)_x \tag{9}$$

$$P_x(side)_{st} = h_3 + k_3 + d_i(t)_y \tag{10}$$

$$P_x(front)_{pq} = h_1 + k_5 + d_i(q)_x \tag{11}$$

$$P_y(top)_{st} = h_2 + k_2 + d_i(t)_z \tag{12}$$

$$P_y(side)_{st} = h_2 + k_4 + d_i(t)_z \tag{13}$$

$$P_y(front)_{st} = h_4 + k_6 + d_i(t)_y \tag{14}$$

The value of h1, h2, h3, h4, h5, h6, k1, k2, k3, k4, k5, k6 are constants. The value of h1, h2, h3, h4, h5, and h6 are used to skeleton image center align. The value of k1, k2, k3, k4, k5, and k6 are calculated spacing between the joints and stretch of the given joint feature. The values selected for normalization may vary depending on the joints.

4.2 RGB and Depth Data Training Using 5S-CNN Classifier

After the RGB, depth, and skeleton data construction, the data are trained using a 5S-CNN classifier. In this paper, the MHI and DMM are trained with a 5S-CNN classifier. But the skeleton data is trained with a combination of 5S-CNN and Bi-LSTM classifiers. Here, all the images are separately trained using a 5S-CNN classifier. Figure 2 clearly shows the training process. The proposed CNN consists of four layers namely, the input layer, five convolutional layers, and pooling layers followed by a fully connected layer. These layers are arranged according to their functionality. An elaborate working of each layer in the proposed CNN architecture is explained below;

Convolutional layer The initial layer of CNN is the convolution layer. This layer is used to generate the feature map of the input image. For generating a feature map, in this paper, we utilize a 5×5 filter. The mathematical form of the convolution layer is given in Eq. (15).

$$B_i^b = \sum_{j\in F_i} B_j^{b-1} \otimes \xi_{ij}^b + L_i^b \tag{15}$$

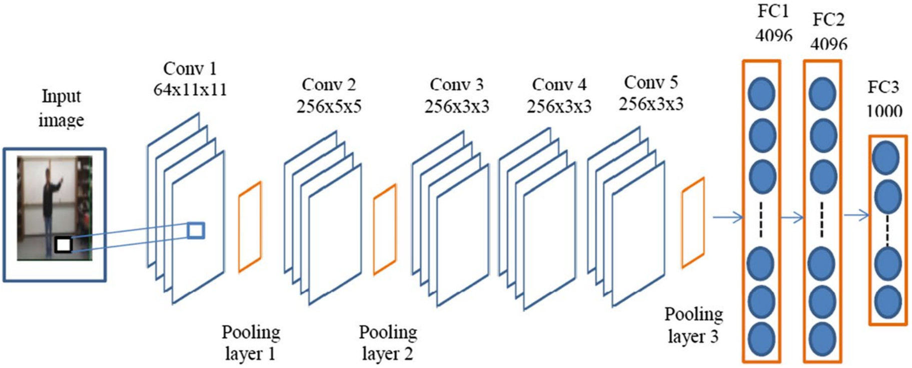


Fig. 2 Structure of CNN

where the symbol '⊗' is represented the Convolution operator, ξ_{ij}^{b} represent the weight value of *i*th the filter of the *b*th convolutional layer, L_{i}^{b} is represent the bias of the *i*th filter of the *b*th convolutional layer and the activation map is represented as B_{i}^{b}.

Pooling layer The convolution layer is followed by the pooling layer. This layer reduces the size of the feature map, thus reducing the computational effort of the network. In each feature map, a pooling process is applied. Maximum pooling is used in this paper. The pooled map is represented by the below equation.

$$P_i = \underset{j \in R_j}{Max} F_i \tag{16}$$

Fully connected layer The reduced feature map is given to the input of the fully connected layer. This layer acts as the classifier. Here, the feed-forward neural network is used for the fully connected layer.

4.3 Skeleton Data Training Using 5S-CNN with Bi-LSTM

After the skeleton data extraction, the three-view data are given to the 5S-CNN followed by the Bi-LSTM classifier. The 5S-CNN concept is already explained in the above section. The output of 5S-CNN is given to the input of the Bi-LSTM classifier. A Bi-LSTM is a sequential processing model consisting of two LSTMs. The first one is used in the forward direction and the other in the backward direction. Bi-LSTMs efficiently maximize the quantity of data available to the network. The structure of Bi-LSTM is given in Fig. 3.

The LSTM is a modified structure of RNN. The LSTM overcomes the difficulties present in the RNN. The LSTM consist of three control gates such as input, forget and output gates. In this memory, the cell state is used to store and update information. The mathematical expression of LSTM is given below;

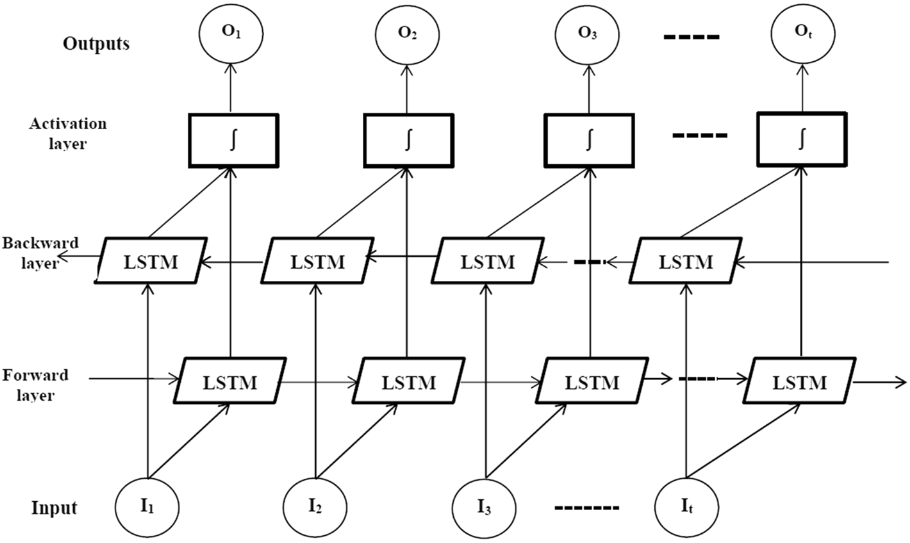


Fig. 3 Structure of Bi-LSTM

$$I_t = \left(\left[m_{t-1},\, v_t\right] W_I + d_I\right)\, \sigma \tag{17}$$

$$F_t = \left(\left[m_{t-1},\, v_t\right] W_F + d_F\right)\, \sigma \tag{18}$$

$$O_t = \left(\left[m_{t-1},\, v_t\right] W_O + d_O\right)\, \sigma \tag{19}$$

$$\tilde{U}_t = \tanh\left(\left[m_{t-1}, v_t\right] W_U + d_U\right) \tag{20}$$

$$U_t = \tilde{U}_t * I_t + U_{t-1} * F_t \tag{21}$$

$$m_t = \tanh\left(U_t\right) * O_t \tag{22}$$

where Tanh→hyperbolic tangent function; σ→sigmoid activation function; v_t→input vector; I_t→output of the input; F_t→output of forget gates; O_t→output of output gates at time t; d→bias value; W→weight of the control gates; $\tilde{U}_t$→input's current state; U_t and h_t→update state and output at time t.

Typically, in an LSTM, the image is encrypted from one direction only. However, two LSTMs can be used as bidirectional encoders, also known as two-way LSTMs (B-LSTMs). Utilizing this Bi-LSTM network can solve the long-distance dependence problem of conventional RN by improving the state transfer system by presenting the CAT mechanism. For an input image, a sequence of hidden states is produced by the Bi-LSTM. The output hidden state of the corresponding input I_t is calculated as follows:

$$m_t = LSTM\left(m_{t-1},\, I_t\right) \tag{23}$$

$$m_t' = LSTM\left(m_{t-1},\, I_t\right) \tag{24}$$

$$M_t = \left[m_t,\, m_t'\right] \tag{25}$$

where m_t and m'_t denote the vector of the hidden layer in the direction of positive and the direction of negative at time t respectively. M_t denotes the vector of final output at time t. Finally, we obtained the score value. This score value is used for further processing.

4.4 Decision-Level Fusion

After the training process, each input data have given one output score value. By using WPM the obtained score values are fused. The proposed WPM is mainly used for the decision-making process. Let MHI_s, DMM_F, DMM_T, DMM_S, SK_T, SK_F, SK_S are the score values obtained from MHI, DMMs, and skeleton joints. These scores serve as the decision-making criteria for WPM. In this case, the number of classes will vary. Based on the criteria (scores) of the results, the best alternative (class) is selected and classified. The score value is calculated using Eq. (26).

$$WPM^S = Max\left[MHI_s^1 \ * \ DMM_F^2 \ * \ DMM_T^3 \ * \ DMM_S^4 \ * \ SK_T^5 \ * \ SK_F^6 \ * \ SK_F^7\right] \tag{26}$$

From the score value, we can identify the correct action of humans.

5 Result and Discussion

The results obtained by proposed human activity recognition are analyzed in this section. The proposed system is implemented using python.3.7 with Windows 7 Operating system at a 2 GHz dual-core PC machine with 4 GB main memory running a 64-bit version of Windows 2007. The proposed methodology is evaluated using the UTD-MHAD dataset.

5.1 Dataset Description

UTD-MHAD is the latest functional database. There are twenty-seven human activities in this database. These database videos are collected using a deep camera and a wearable passive sensor. This database contains RGB videos, deep videos, skeleton positions, and transition signals. This database is used for both the training and testing processes. Odd models are used for the training process and samples are also used for the testing process. Some of the sample activities are listed in Fig. 4.

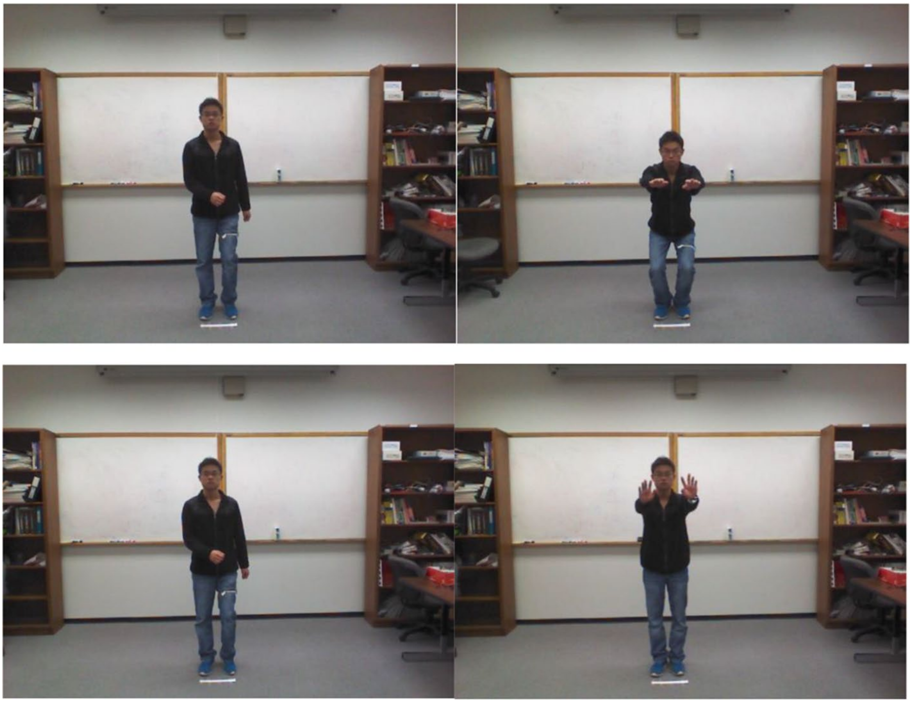
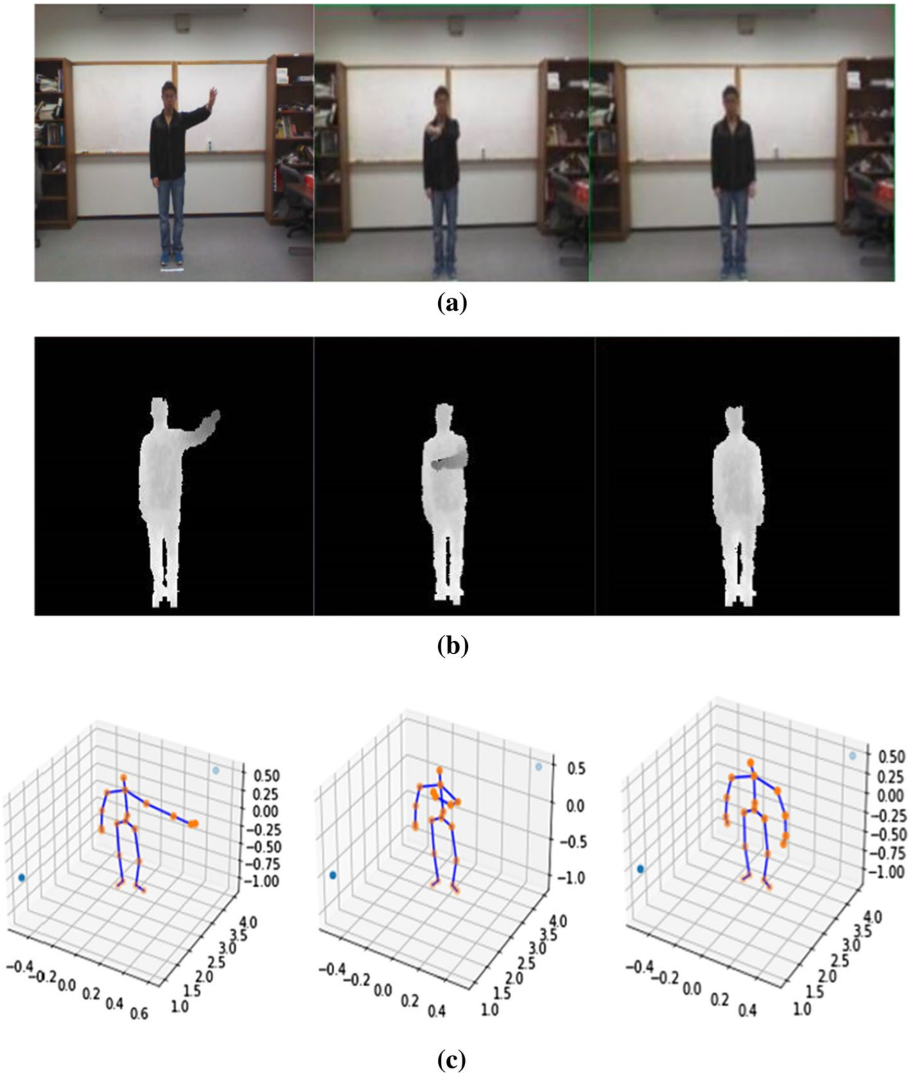


Fig. 4 Experimental used sample images

5.2 Experimental Results

The main objective of the proposed methodology is human activity recognition from video or images. In this paper, for activity recognition three modalities such as MHI, DMMs, and skeletons are utilized. The dataset sequence is presented in Fig. 5 and the confusion matrix is given in Fig. 6.

The AUC for three different actions is presented in Fig. 7. The curve is drawn between the false positive rate and the true positive rate. Here, for boxing area under the ROC curve is 0.984, for bowing 0.9231, and for the tennis swing 0.9573.



(a)

(b)

(c)

Fig. 5 UTD-MHAD dataset sequence, **a** RGB frames, **b** depth frames and **c** skeleton sequences

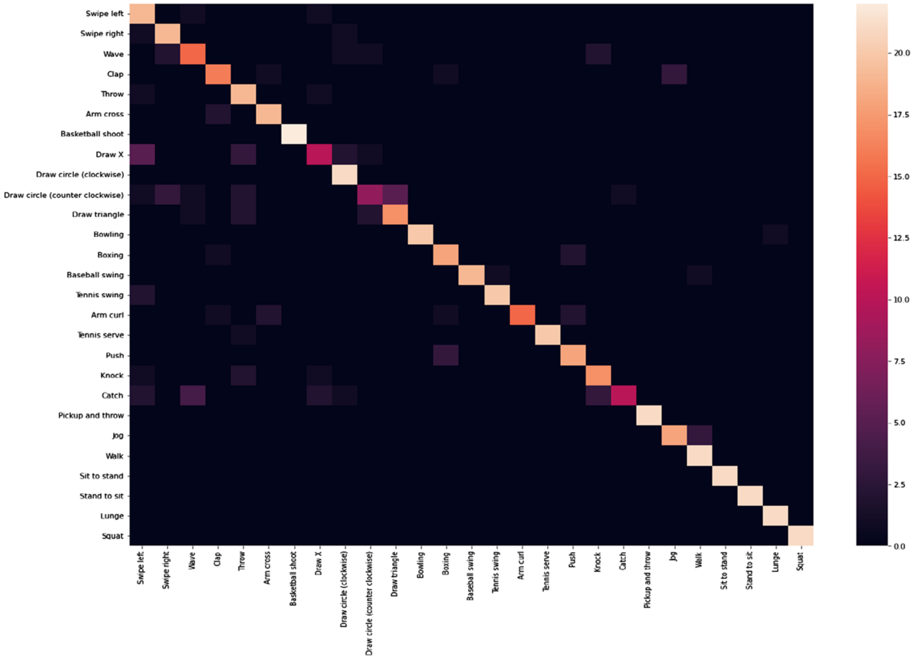


Fig. 6 Confusion matrix of the proposed methodology

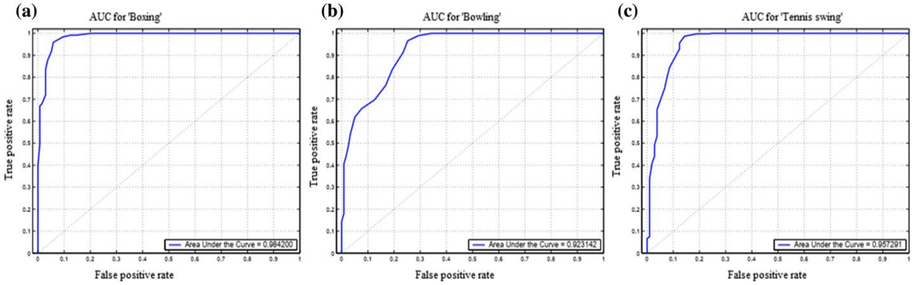


Fig. 7 AUC for three actions **a** boxing, **b** bowling, and **c** tennis swing

The accuracy and loss of the training dataset and validation data for 80 epochs are shown in Figs. 8 and 9. The graph shows, the validation process attained the maximum accuracy of 96.2% and loss of 0.45%.

5.3 Comparative Analysis Results

To prove the efficiency of the proposed approach, we compare our work performance with different classifiers namely, SVM-based human action recognition, KNN-based human action recognition, and CNN-based human action recognition. in this paper, we used novel

Fig. 8 Accuracy versus Epochs

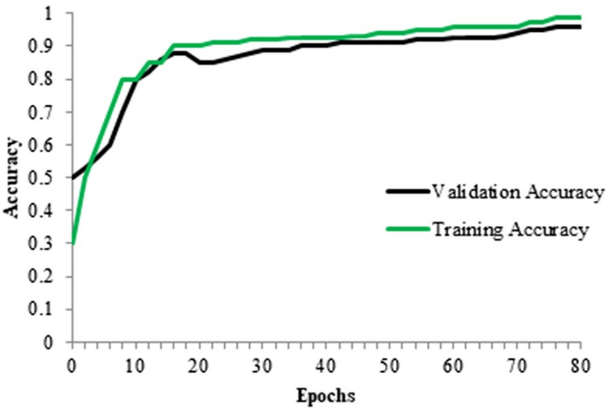
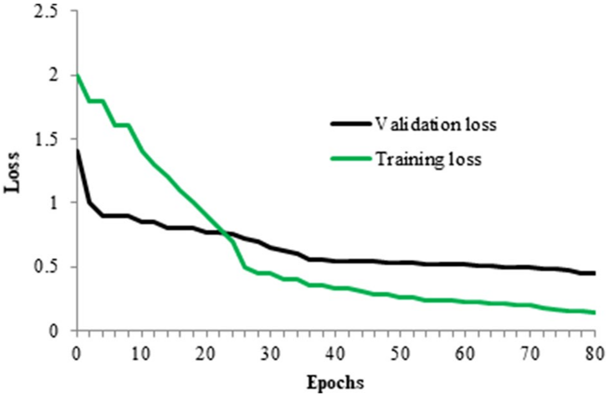


Fig. 9 Loss versus Epochs



5S-CNN+Bi-LSTM classifier for prectiction. The performance analysed based on precision, accuracy, recall and recognition rate.

In Fig. 10, the performance of the recommended approach is discussed in terms of accuracy by varying training and testing data size. In the recommended approach, three methods such as MHI, DMMs, and skeletons are used for the recognition process. From each method, features are extracted. These features are trained using different classifiers. The 5S-CNN classifier is used to train MHI and DMM. Similarly, the 5S-CNN and Bi-LSTM classifiers are used to train the skeleton image. Recognition of this hybrid approach increases the accuracy rate. To demonstrate the effectiveness of the recommended approach, we compare our algorithm with different classifiers, i.e. K-Nearby Neighbor Classifier (KNN), Support Vector Machine (SVM), and CNN Classifier. In this program, the methods are trained using the same classifier. According to Fig. 10, the recommended method reached a maximum accuracy of 96.2%, which is 79% for KNN-based Recognition, 83% for CNN-based Recognition, and 86% for CNN-based Recognition. Furthermore, in Fig. 11, the performance of the recommended method is analyzed based on accuracy. According to Fig. 11, our recommended method reached a maximum accuracy of 96.5%, which is higher than KNN-based Recognition, SVM-based Recognition, and CNN-based Recognition. This is due to the hybrid 5S-CNN and dual-LSTM-based skeletal training

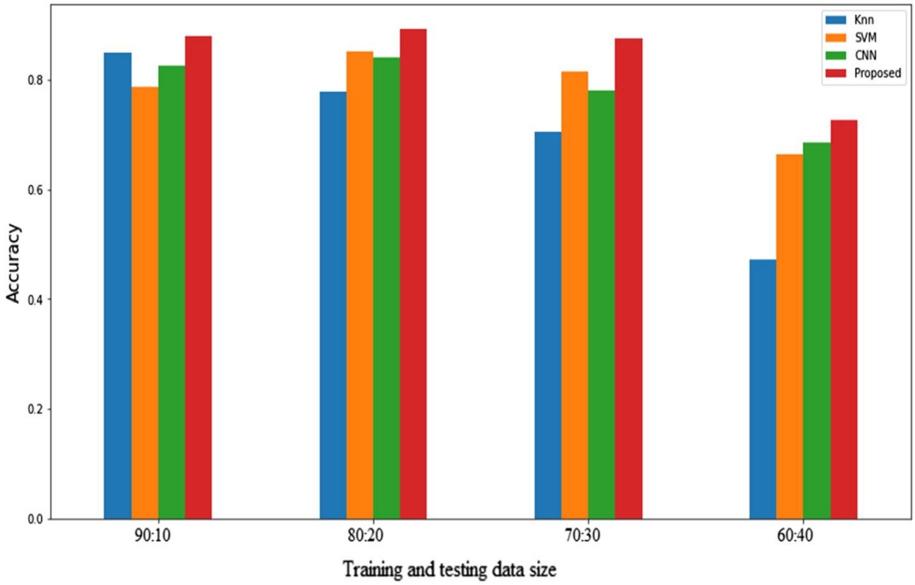


Fig. 10 Performance analysis based on accuracy

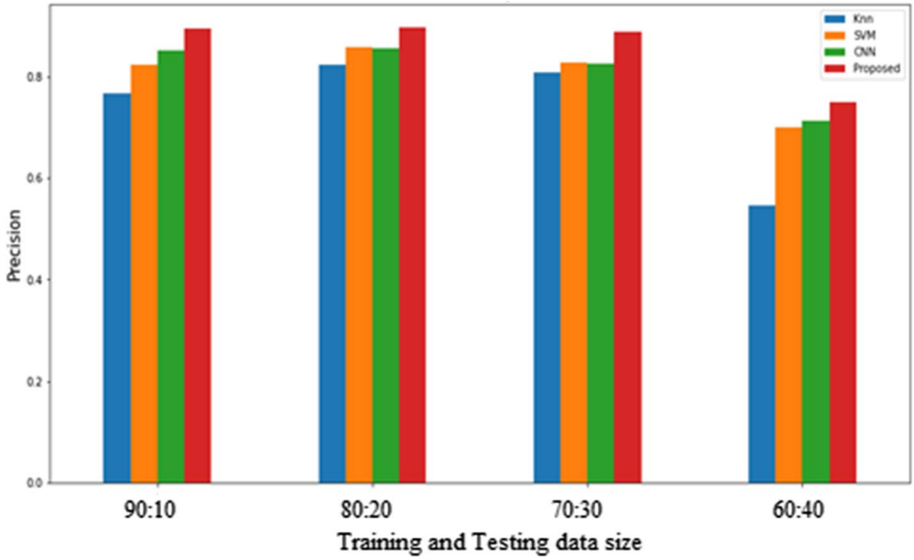


Fig. 11 Performance analysis based on precision

and three-pronged fusion process. In Fig. 12, the performance of the recommended method is discussed in terms of recall size. A good organization has the highest recall value. According to Fig. 12, the recommended method withdraws a maximum of 95%, which is higher compared to other methods. Multi-View Multi-Model Based Human Performance

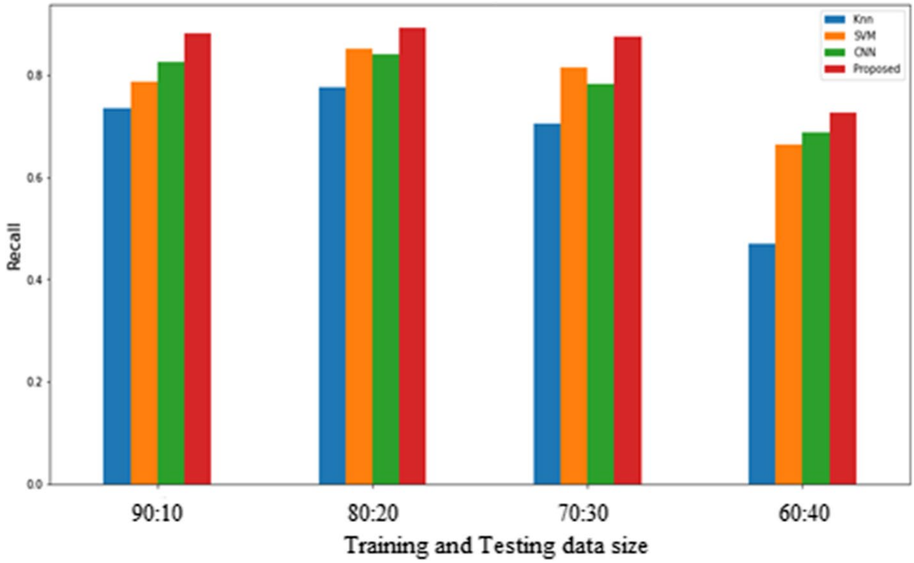


Fig. 12 Performance analysis based on recall

Recognition Performance Based on Authorization Rate is given in Fig. 13. The approval ratio is measured based on the fusion score value. Fusion is done between MHIs, DMMs, and skeletons. According to Fig. 13, our proposed approach takes the maximum recognition rate to be 0.6, which is 0.53 for SVM-based Recognition, 0.52 for CNN-based recognition, and 0.46 for CNN-based Recognition. Due to these three methods, 5S-CNN and

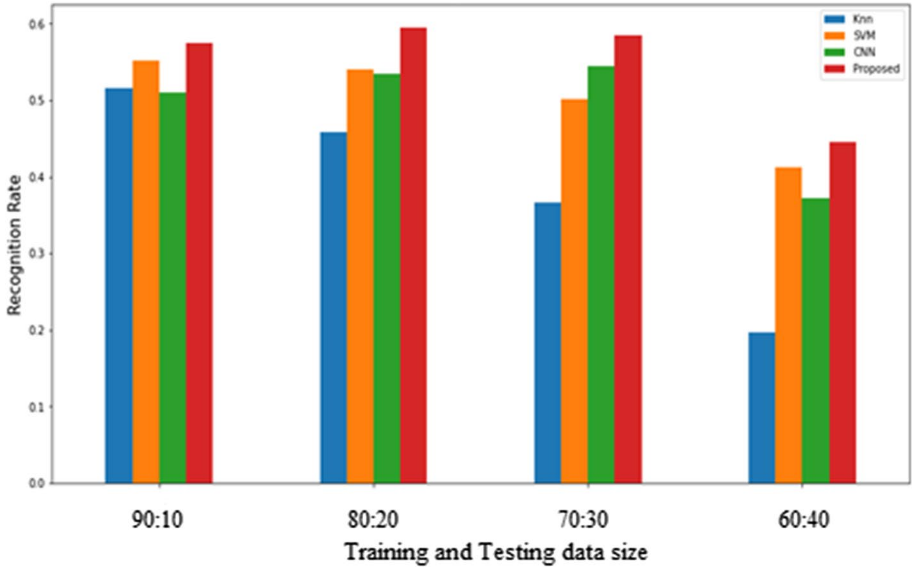


Fig. 13 Performance analysis based on recognition rate

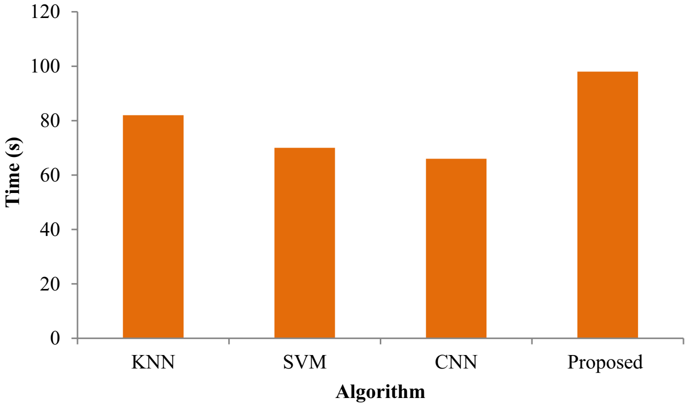


Fig. 14 Complexity analysis based on time

Table 1 Comparative analysis based on the accuracy

Approaches	Accuracy (%)
Chen et al., [22] (Depth + Inertia)	79.1
Bulbul et al. [23] (Depth)	88.4
Annadani et al. [24] (Skeleton)	86.12
Escobedo and Camara[25] (RGB + Depth + Skeleton)	84.4
Proposed (RGB + Depth + Skeleton)	**96.2**

Bi-LSTM-based training and integration processes offer better accreditation rates compared to other methods. Moreover, in Fig. 14, the time complexity is analyzed. The time complexity of an algorithm quantifies the amount of time taken by an algorithm to run as a function of the length of the input. When analysing Fig. 14, compared to other techniques, proposed method takes more time to execute. This is because of the hybrid approach. This processing time does not affect the overall recognition accuracy.

5.4 Comparison with Published Work

To prove the efficiency of the suggested method, the suggested algorithm is compared with different methods as shown in Table 1. In this section, we compare our work with already published works such as Chen et al. [22], Bulbul et al. [23], Annadani et al. [24], and Escobedo and Camara [25]. As shown in the table, the accuracy of the proposed approach is 96.2%. In [22], human action recognition is carried out using depth and inertial sensors. Here, for recognition collaborative representation classifier (CRC) has been utilized. Here, the UTD-MHAD dataset is utilized. This method is useful for multi-modality human action recognition. In [23], they fuse three different kinds of features to tackle the action recognition problem. Here, Three DMMs such as front, side, and top projection views are extracted from input images or videos. Here, two decision-level fusions were developed. In [24], skeleton-based activity recognition was proposed. In [25], intensity, depth, and skeleton joints are used for the recognition process.

Methods proposed by Chen et al. [22], Bulbul et al. [23], Annadani et al. [24], and Escobedo and Camara [25] are among the best current programs for multimodal human activity recognition. Furthermore, they used MHI, DMMs, and skeletal representation. Therefore, we have chosen to compare the performance of our proposed method with these. In Table 1, we compare our proposed approach with the method mentioned above. When analyzing Table 1, our proposed approach reached a maximum accuracy of 96.2%, which is 79.1% for [22], 88.4% for [23], 86.12% for [24], and 84.4% for [25]. This is because our proposed approach considers three methods and a hybrid approach to training. The current approach used only one or two methods. From the results, it is clear that the proposed approach achieved better results compared to other approaches.

6 Conclusion

The main purpose of multi-view multi-model-based human activity recognition is proposed in this study. The problem with human activity Recognition is solved by joining the RGB-D sensor's multiple view notes. Here, initially, the features of the input image are extracted and each feature is trained in different classifiers. The mathematical expression of each phase is clearly illustrated. Experimental results show that the recommended method is useful in using different data methods and achieves better results with other existing works. In the future, we will focus on optimization and different classification for the Recognition process.

Author Contributions All authors read and approved the final manuscript.

Funding The authors declare that they have competing interests and funding.

Data Availability Data sharing is not applicable to this article because of proprietary nature.

Code Availability Code sharing is not applicable to this article because of proprietary nature.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Aggarwal, J. K., & Ryoo, M. S. (2011). Human activity analysis: A review. *ACM Computing Surveys (CSUR), 43*(3), 16.
2. Wang, P., Li, W., Gao, Z., Tang, C., Zhang, J., & Ogunbona, P. (2015). Convnets-based action recognition from depth maps through virtual cameras and pseudocoloring. In *Proceedings of the 23rd ACM international conference on Multimedia*, ACM, pp. 1119–1122.
3. Wang, P., Li, W., Gao, Z., Zhang, J., Tang, C., & Ogunbona, P. O. (2016). Action recognition from depth maps using deep convolutional neural networks. *IEEE Transactions on Human-Machine Systems, 46*(4), 498–509.
4. Wang, P., Li, W., Ogunbona, P., Gao, Z., & Zhang, H. (2014). Mining mid-level features for action recognition based on effective skeleton representation. In *Digital lmage computing: techniques and applications (DlCTA), 2014 international conference on, IEEE*, pp. 1–8.

5. Vemulapalli, R., Arrate, F., & Chellappa, R. (2014). Human action recognition by representing 3d skeletons as points in a lie group. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 588–595.
6. Simonyan, K., & Zisserman, A. (2015). *Very deep convolutional networks for largescale image recognition*, arXiv preprint arXiv:1409.1556, ICLR, pp. 1–10.
7. Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1725–1732.
8. Simonyan, K., & Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. In *Advances in neural information processing systems*, pp. 568–576.
9. Luo, J., Wang, W., & Qi, H. (2014). Spatio-temporal feature extraction and representation for RGB-D human action recognition. *Pattern Recognition Letters, 50*, 139–148.
10. Chen, C., Jafari, R., & Kehtarnavaz, N. (2015). Improving human action recognition using fusion of depth camera and inertial sensors. *IEEE Transactions on Human-Machine Systems, 45*(1), 51–61.
11. El Madany, N. E. D., He, Y., & Guan, L. (2016). Human action recognition via Multiview discriminative analysis of canonical correlations. In: *Image processing (ICIP), 2016 IEEE international conference on, IEEE*, pp. 4170–4174.
12. Verma, P., Sah, A., & Srivastava, R. (2020). Deep learning-based multi-modal approach using RGB and skeleton sequences for human activity recognition. *Multimedia Systems, 26*(6), 671–685.
13. Wang, P., Li, W., Gao, Z., Zhang, J., Tang, C., & Ogunbona, P. O. (2015). Action recognition from depth maps using deep convolutional neural networks. *IEEE Transactions on Human-Machine Systems, 46*(4), 498–509.
14. Chen, C., Jafari, R., & Kehtarnavaz, N. (2016). Fusion of depth, skeleton, and inertial data for human action recognition. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 2712–2716. IEEE.
15. Escobedo, E., & Camara, G. (2016). A new approach for dynamic gesture recognition using skeleton trajectory representation and histograms of cumulative magnitudes. In *2016 29th SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)*, pp. 209–216. IEEE.
16. Gaglio, S., Re, G. L., & Morana, M. (2014). Human activity recognition process using 3-D posture data. *IEEE Transactions on Human-Machine Systems, 45*(5), 586–597.
17. Khaire, P., Kumar, P., & Imran, J. (2018). Combining CNN streams of RGB-D and skeletal data for human activity recognition. *Pattern Recognition Letters, 115*, 107–116.
18. Guo, J., Bai, H., Tang, Z., Xu, P., Gan, D., & Liu, B. (2020). Multi modal human action recognition for video content matching. *Multimedia Tools and Applications, 79*, 34665–34683.
19. Tran, T.-H., Tran, H.-N., & Doan, H.-G. (2019). Dynamic hand gesture recognition from multi-modal streams using deep neural network. In *International conference on multi-disciplinary trends in artificial intelligence*. Springer, Cham, pp. 156–167.
20. Nie, W., Yan, Y., Song, D., & Wang, K. (2020). Multi-modal feature fusion based on multi-layers LSTM for video emotion recognition. *Multimedia Tools and Applications, 80*, 1–10.
21. Khowaja, S. A., & Lee, S.-L. (2020). Hybrid and hierarchical fusion networks: a deep cross-modal learning architecture for action recognition. *Neural Computing and Applications, 32*(14), 10423–10434.
22. Chen, C., Jafari, R., & Kehtarnavaz, N. (2015). UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. In *Image processing (ICIP), 2015 IEEE international conference on, IEEE*, pp. 168–172.
23. Bulbul, M. F., Jiang, Y., & Ma, J. (2015). Dmms-based multiple features fusion for human action recognition. *International Journal of Multimedia Data Engineering and Management (IJMDEM), 6*(4), 23–39.
24. Annadani, Y., Rakshith, D., & Biswas, S. (2016). Sliding dictionary based sparse representation for action recognition, arXiv preprint arXiv:1611.00218, 1–7.
25. Escobedo, E., & Camara, G. (2016). A new approach for dynamic gesture recognition using skeleton trajectory representation and histograms of cumulative magnitudes. In *Graphics, patterns and images (SIBGRAPI), 2016 29th SIBGRAPI conference on, IEEE*, pp. 209–216.

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

Rahul Kumar is currently a Ph.D. Candidate in the Department of Computer Science & Engineering under the supervision of Dr. Shailender Kumar, Professor at Delhi Technological University, India. He has completed his Master's degree in Computer Science & Engineering from Delhi Technological University, Delhi. His research interest includes Computer Vision, Activity Recognition, Machine Learning and Deep Learning.

Shailender Kumar is working as a professor in the Department of Computer Science & Engineering at Delhi Technological University, India. He has more than 15 years of experience in teaching and research. He is having specialization in Database Systems, Data mining, Big Data Analytics, Machine Learning and Information security.

Journal of Computational Electronics
https://doi.org/10.1007/s10825-023-02008-w

Numerical modeling of a dielectric modulated surrounding-triple-gate germanium-source MOSFET (DM-STGGS-MOSFET)-based biosensor

Amit Das[1] · Sonam Rewari[2] · Binod Kumar Kanaujia[1] · S. S. Deswal[3] · R. S. Gupta[4]

Received: 19 May 2022 / Accepted: 1 January 2023

Abstract

This paper presents for the first time an analytical model of a dielectric modulated surrounding-triple-gate MOSFET with a germanium source-based biosensor, which shows excellent improvement in sensitivity when compared to a silicon source. The mathematical analysis is based on the center-channel potential which is obtained by solving Poisson's equation in the cylindrical coordinate system using a parabolic approximation. The channel potential profile, threshold voltage, drain current, and subthreshold swing are obtained mathematically. Biosensing performance is investigated for different charged and neutral biomolecules by varying different device metrics including cavity thickness, channel thickness, cavity length, channel doping, and drain voltage. The analytical results are validated and verified with numerical simulations conducted with the ATLAS TCAD simulator and show outstanding agreement with the simulated results.

Keywords Analytical model · Biosensing · Gate-all-around MOSFET · Surrounding-gate MOSFET · TCAD · Threshold voltage · Subthreshold swing

✉ Sonam Rewari
rewarisonam@gmail.com

Amit Das
amitofficial7492@gmail.com

Binod Kumar Kanaujia
bkkanaujia@yahoo.co.in

S. S. Deswal
satvirdeswal@hotmail.com

R. S. Gupta
rsgupta1943@gmail.com

[1] School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi 110067, India

[2] Department of Electronics and Communication Engineering, Delhi Technological University, New Delhi 110042, India

[3] Department of Electrical and Electronics Engineering, Maharaja Agrasen Institute of Technology, New Delhi 110086, India

[4] Department of Electronics and Communication Engineering, Maharaja Agrasen Institute of Technology, New Delhi 110086, India

1 Introduction

Continuous development and improvement in science and technology have led to the evolution of different electronic devices [1–3]. The metal–oxide–semiconductor field-effect transistor (MOSFET), one of the most prominent devices, has been widely used for different applications [4, 5]. MOSFET-based biosensors are becoming popular nowadays given their cost-effectiveness, high sensitivity, label-free detection, scalability, and ability to detect and sense biomolecules [6]. The sensing ability of a surrounding-gate MOSFET biosensor (SGMB) is due to a change in the value of the metrics used [7]. Different device metrics that are commonly used to study biosensing performance are threshold voltage, transconductance, drain current, subthreshold slope, gate capacitance, etc. Detection of biomolecules using a MOSFET has efficacy and utility in the biomedical field. An embedded cavity is created inside the gate oxide to immobilize and hold different biomolecules.

It has been reported in the literature that gate oxide stacking and gate engineering can significantly improve sensitivity. Gautam et al. reported and discussed a numeric model of a surrounding-gate MOSFET-based biosensor [8]. Recently, Pratap et al. reported a highly sensitive surrounding-gate junction-less MOSFET biosensor with a one-sided cavity

Published online: 27 February 2023

for the detection of different neutral biomolecules [9]. Goel et al. reported a triple-gate surrounding-gate MOSFET biosensor [10] and a dielectric-modulated biotube FET biosensor [11]. Chakraborty et al. reported a junction-less surrounding-gate biosensor with a two-sided cavity and a gate stacking technique [6]. In 2016, Padmanaban et al. reported a triple-gate MOSFET biosensor with gate oxide stacking and a two-sided cavity [12]. Das et al. reported a surrounding-gate MOSFET biosensor based on a charge plasma concept [13]. Recently, Li et al. investigated a vertically stacked nanosheet surrounding-gate FET-based biosensor [14].

Literature review reveals that an optimization in device parameters is needed in every biosensor to obtain high sensitivity. The effect of fringing fields is less in case of high-κ dielectrics, but depositing it on silicon requires extra processing time and steps during fabrication. In the current study, gate oxide stacking of SiO_2 and HfO_2 has been used, and a symmetric cavity is created completely inside the oxide layer. The advantage of an I-shaped layer over the conventional one-sided cavity is that the former offers a better fill-in factor probability, and all the simulations have been performed assuming a fill-in factor of 1. A dual-sided cavity is much more efficacious when compared to the conventional one-sided cavity in terms of power consumption and current sensitivity [15, 16]. Source and drain material also influence the sensitivity of the device. Wu et al. [17] and Brunco et al. [18] discussed and investigated the germanium MOSFET in the past. Saha et al. [19] recently discussed a FET-based biosensor with a germanium source.

In our investigation, the proposed biosensor has shown the highest sensitivity for a germanium source when compared to a silicon source. Using a highly doped germanium source increases the tunneling probability [20] and improves charge carrier transport (due to the lower bandgap energy of germanium) which changes the potential barrier at the source–channel junction (bias- and electric field-dependent). The dependence of threshold voltage, drain current, and subthreshold swing on the source–channel potential barrier makes the germanium source MOSFET more sensitive to biomolecules than a silicon source MOSFET. It is worth noting that gate stacking and material engineering have improved the sensitivity of the device by a considerable amount. To date, analytical modeling of various surrounding-gate MOSFET-based biosensors has been discussed, but surrounding-gate MOSFET-based biosensors with a germanium source and gate oxide stacking have not yet been reported. Germanium-based biosensors have excellent biosensing performance but are still not commonly used in practical applications due to some inherent disadvantages [21–23] including incompatibility with existing technology processes (industrial drawback), limited fabrication technology availability which makes them a bit costlier than silicon-based biosensors, an operating temperature range lower than that of silicon which can limit the biosensing performance (due to which moving beyond a certain temperature range makes the sensitivity highly temperature-dependent), and high leakage current that may cause unintentional changes in the sensitivity that could result in unrealistic results.

Investigating the biosensing performance of a biosensor only requires changing the value of the metrics. In this investigation, the vertical electrical field is changed due to the localization of biomolecules inside the cavity which ultimately changes the gate oxide capacitance [24–26]; hence, sensitivity is calculated for the biosensor as given by the relative change in the value of the sensing metric. The horizontal electric field (doping-dependent) is almost constant since the doping of the source, channel, and drain is constant [27, 28], so it has a negligible effect on the sensitivity.

In the present work, an analytical model of a dielectric modulated surrounding-triple-gate germanium-source MOSFET-based biosensor (DM-STGGS-MOSFET) with a double-sided cavity of 25 × 6 nm is developed. In the era of digitalization and nanotechnology, the DM-STGGS-MOSFET can act as a versatile biosensor that not only is capable of detecting biomolecules but can also sense different environmental and physical parameters, offering the advantage of a cost-efficient device that can detect nanoparticles or biochemical species with high sensitivity. The DM-STGGS-MOSFET can be utilized in the monitoring and detection of hazardous chemicals [29–31] in production warehouses or reactors where there is a threat to human health and life. The analytical results so obtained show close agreement with the simulation results obtained on a technology computer-aided design (TCAD) simulator [6, 32, 33].

The paper is divided into five sections. The device structure, fabrication process, and simulator specifications are discussed in Sect. 2. Section 3 presents the analytical model of the biosensor. The results and discussion are given in Sect. 4. Section 5 highlights the important conclusions of this work.

2 Device structure, fabrication process, and simulation

Figure 1 shows the cylindrical structure and cross-sectional view of the DM-STGGS-MOSFET. Table 1 shows all the structural parameters of the DM-STGGS-MOSFET used during the simulation. The fabrication starts with basic fabrication steps [14, 34, 35] including wafer cleaning, growth of epitaxial silicon, sacrificial layer deposition, and deposition of a lightly doped silicon body over a heavily doped drain to form a vertical pillar-like annular structure [36]. The rest of the fabrication steps (using a vertically stacked nanowire structure) are shown in a flowchart in Fig. 2a. The basic detection principle of the DM-STGGS-MOSFET is

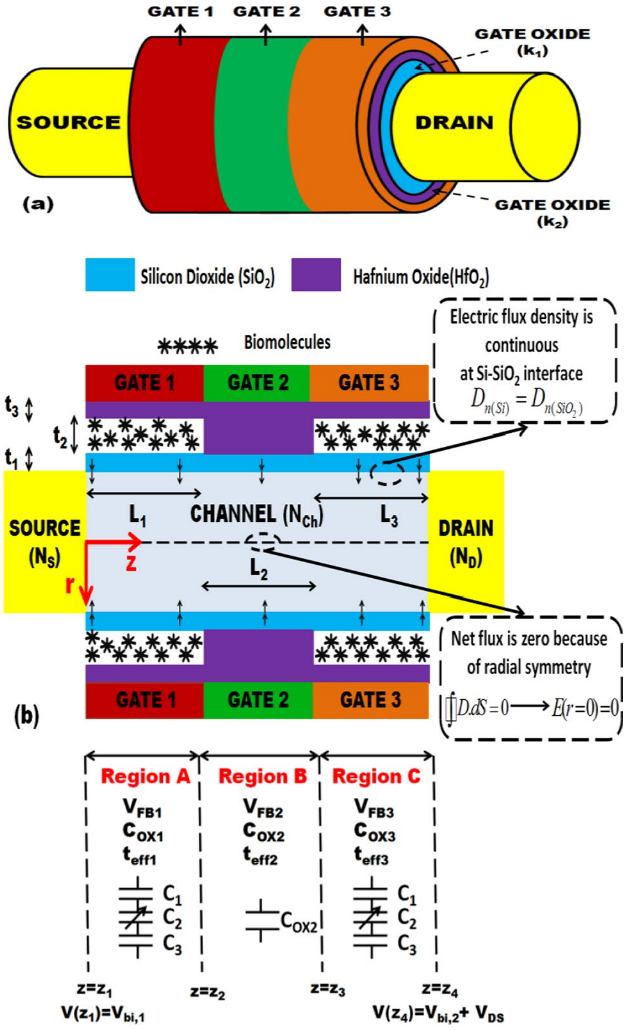


Fig. 1 **a** Cylindrical and **b** cross-sectional view of the DM-STGGS-MOSFET

based upon localization of different biomolecules inside the cavity, which changes the potential and field profile across the channel, and hence the device becomes sensitive to biomolecules, as illustrated in Fig. 2b. Figure 2c and 2d shows the calibration with experimental data [37] and a block diagram depicting the step-by-step fabrication process for the DM-STGGS-MOSFET. The proposed biosensor has shown excellent improvement in sensitivity when compared with a similar and prominent variant of a MOSFET-based biosensor. Figure 2e shows a comparison of the threshold voltage sensitivity between the DM-STGGS-MOSFET and an already existing biosensor. The main disadvantage of the proposed biosensor is regarding selectivity and a non-unity fill-in factor. Although the proposed biosensor shows high sensitivity, selectivity could be one of its limitations that have not been explored in this work. Another disadvantage that can affect the sensitivity is the fill-in factor of biomolecules. In a practical scenario, the whole cavity might not be

Table 1 Structural parameters

Parameter		Value	Parameter		Value
Doping	Source	10^{19}/cm^3	Material	Source	Germanium (Ge)
	Channel	10^{10}/cm^3		Channel	Silicon (Si)
	Drain	10^{19}/cm^3		Drain	Silicon (Si)
Radius (a)	Source	10 nm	Work function (ϕ_M)	ϕ_{M1}	5.10 eV (Au) [13]
	Channel	10 nm		ϕ_{M2}	4.53 eV (Mo) [13]
	Drain	10 nm		ϕ_{M3}	4.10 eV (Al) [38]
Oxide layer thickness (t_{ox})	$t_1/t_2/t_3$	1/6/1 nm	Channel length		60 nm
Cavity thickness (t_{cav})		6 nm	Cavity length	L_1/L_3	25 nm
Source/drain length		20 nm	Oxide length	L_{OX}/L_2	60/10 nm
Gate oxide	SiO$_2$	k_1 (3.9)	n_i [33] (T = 300 K)	Silicon	$1*10^{10}$/cm^3
	HfO$_2$	k_2 (25)		Germanium	$2.5*10^{13}$/cm^3

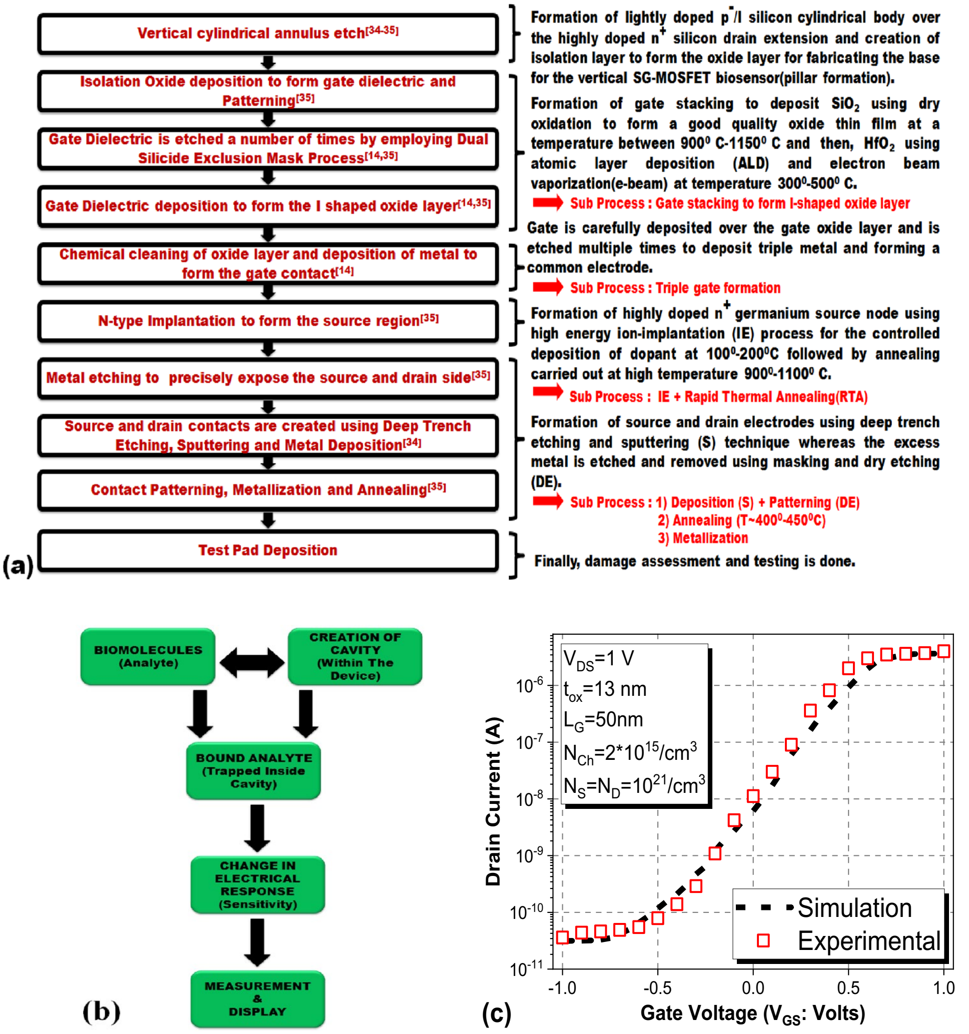

Vertical cylindrical annulus etch[34-35]
Isolation Oxide deposition to form gate dielectric and Patterning[35]
Gate Dielectric is etched a number of times by employing Dual Silicide Exclusion Mask Process[14,35]
Gate Dielectric deposition to form the I shaped oxide layer[14,35]
Chemical cleaning of oxide layer and deposition of metal to form the gate contact[14]
N-type Implantation to form the source region[35]
Metal etching to precisely expose the source and drain side[35]
Source and drain contacts are created using Deep Trench Etching, Sputtering and Metal Deposition[34]
Contact Patterning, Metallization and Annealing[35]
Test Pad Deposition
(a)
Formation of lightly doped p⁻/I silicon cylindrical body over the highly doped n⁺ silicon drain extension and creation of isolation layer to form the oxide layer for fabricating the base for the vertical SG-MOSFET biosensor(pillar formation).
Formation of gate stacking to deposit SiO₂ using dry oxidation to form a good quality oxide thin film at a temperature between 900⁰ C-1150⁰ C and then, HfO₂ using atomic layer deposition (ALD) and electron beam vaporization(e-beam) at temperature 300⁰-500⁰ C.
Sub Process : Gate stacking to form I-shaped oxide layer
Gate is carefully deposited over the gate oxide layer and is etched multiple times to deposit triple metal and forming a common electrode.
Sub Process : Triple gate formation
Formation of highly doped n⁺ germanium source node using high energy ion-implantation (IE) process for the controlled deposition of dopant at 100⁰-200⁰C followed by annealing carried out at high temperature 900⁰-1100⁰ C.
Sub Process : IE + Rapid Thermal Annealing(RTA)
Formation of source and drain electrodes using deep trench etching and sputtering (S) technique whereas the excess metal is etched and removed using masking and dry etching (DE).
Sub Process : 1) Deposition (S) + Patterning (DE)
2) Annealing (T~400⁰-450⁰C)
3) Metallization
Finally, damage assessment and testing is done.
BIOMOLECULES (Analyte)
CREATION OF CAVITY (Within The Device)
BOUND ANALYTE (Trapped Inside Cavity)
CHANGE IN ELECTRICAL RESPONSE (Sensitivity)
MEASUREMENT & DISPLAY
(b)
V_DS=1 V
t_ox=13 nm
L_G=50nm
N_Ch=2*10^15/cm^3
N_S=N_D=10^21/cm^3
Simulation
Experimental
Drain Current (A)
10^-6
10^-7
10^-8
10^-9
10^-10
10^-11
-1.0
-0.5
0.0
0.5
1.0
Gate Voltage (V_GS: Volts)
(c)

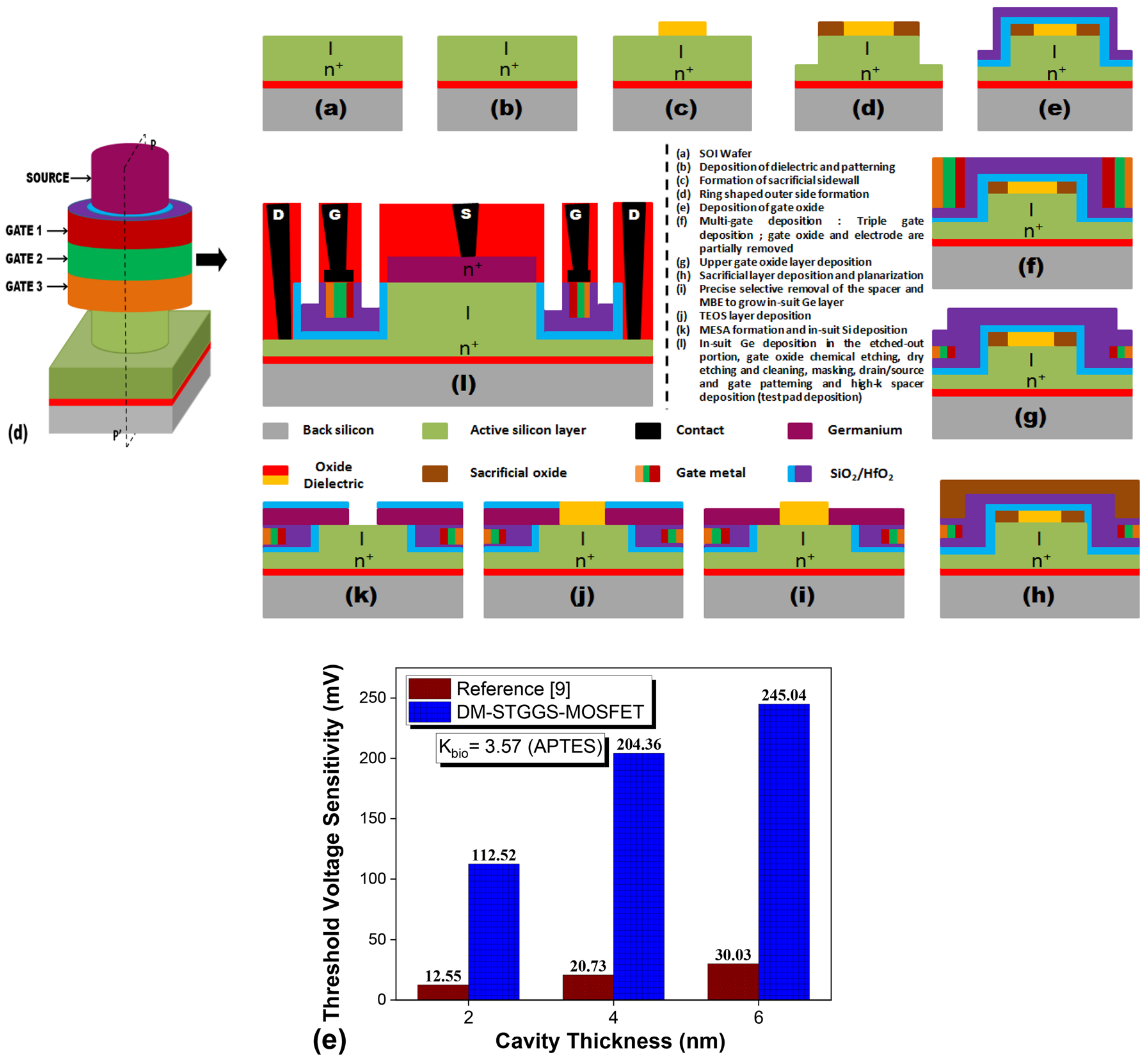
Fig. 2 **a** Fabrication process (vertically stacked nanowire architecture) flowchart [39], **b** biosensing principle flowchart [7], **c** calibration with experimental data [37] **d** block diagram showing different fabrication steps (conventional method [39]) across the cross section P–P′, and **e** threshold voltage sensitivity comparison with the reference biosensor [9] for the DM-STGGS-MOSFET-based biosensor

filled with biomolecules, which may cause a deviation in the sensitivity from the ideal values. But again, the problem of selectivity and a non-unity fill-in factor is common in every MOSFET-based biosensor [7].

The different analytical results were verified using simulations performed on ATLAS TCAD software. Different models were incorporated in the simulation: the Auger model (recombination model to calculate high charge densities), the Schottky–Read–Hall (SRH) model (to calculate carrier lifetimes of majority and minority charge carriers at high doping), the CONMOB model (concentration-dependent mobility model to calculate concentration-dependent mobility), the Boltzmann model (carrier statistic model), the FLDMOB model (electric field-dependent mobility model to calculate any type of velocity saturation at high doping), and the Lombardi CVT model (mobility model used in nonplanar MOSFETs). The method used to solve the complex differential and integral meshing of the biosensor is the Newton–Gummel method, which is a slow and robust method.

Table 2 shows the threshold voltage sensitivity (S_{Vt}), subthreshold swing sensitivity (S_{SS}), and off-current sensitivity (S_{IOFF}) of a surrounding-gate MOSFET with different combinations of gates and source materials. It has been observed that the biosensor shows better results in terms of threshold voltage change and off-current change when the source material is germanium. A germanium source offers approximately twofold higher mobility for conducting electrons and fourfold higher mobility for holes as compared to a silicon source [40]. This is due to the lower effective mass of electrons and holes that results in high mobility in the case of germanium. This mobility enhancement leads to the improvement in the subthreshold characteristics of the biosensor. Thus, a germanium source-based biosensor will show a comparatively larger swing in current for the same range of gate voltage than a silicon-based biosensor. This means that a DM-STGGS-MOSFET with a germanium source will show a larger variation in different sensing metrics such as threshold voltage and subthreshold swing that will make it more sensitive to biomolecules as compared to

Fig. 2 (continued)


SOURCE
GATE 1
GATE 2
GATE 3
P
P'
(d)
I
n^+
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
D
G
S
G
D
(a) SOI Wafer
(b) Deposition of dielectric and patterning
(c) Formation of sacrificial sidewall
(d) Ring shaped outer side formation
(e) Deposition of gate oxide
(f) Multi-gate deposition : Triple gate deposition ; gate oxide and electrode are partially removed
(g) Upper gate oxide layer deposition
(h) Sacrificial layer deposition and planarization
(i) Precise selective removal of the spacer and MBE to grow in-suit Ge layer
(j) TEOS layer deposition
(k) MESA formation and in-suit Si deposition
(l) In-suit Ge deposition in the etched-out portion, gate oxide chemical etching, dry etching and cleaning, masking, drain/source and gate patteming and high-k spacer deposition (test pad deposition)
Back silicon
Active silicon layer
Contact
Germanium
Oxide Dielectric
Sacrificial oxide
Gate metal
SiO_2/HfO_2
Threshold Voltage Sensitivity (mV)
250
200
150
100
50
0
2
4
6
Cavity Thickness (nm)
Reference [9]
DM-STGGS-MOSFET
K_{bio}= 3.57 (APTES)
12.55
112.52
20.73
204.36
30.03
245.04
(e)

a silicon source biosensor. It should be noted that increasing the number of gates can increase the sensitivity of the device, but it also increases the complexity during fabrication. For single-gate, double-gate, and triple-gate architecture, the total length and area have been kept the same. Only the length of the individual gates has been varied, and the total gate length has been kept constant at 60 nm (the rest of the parameters are the same in each case). The performance of the triple-gate device is superior to that of the single-/double-gate architecture [10, 41, 42] in terms of sensitivity improvement. The improvement and enhancement of sensitivity in the triple-gate architecture is due to impact ionization, because of which short-channel effects decrease and the electron net velocity increases. This improves the biosensing ability of the MOSFET [10]. Also, using three gates instead of a single gate increases the steepness of the potential profile, which varies with the work function difference. A larger work function difference between the adjacent gates improves the transport efficiency and the step profile of the channel's potential [44]. Due to this, a large number of charge carriers flow in a more controlled manner in a triple-gate architecture. Hence, the value of different sensing parameters (threshold voltage, subthreshold swing, off-current) changes relatively more in the presence of biomolecules. The size of streptavidin and biotin is roughly 5 nm [45], whereas that of protein lies in the micro–femtometer range [44, 46]. Kim et al. [47] reported an experimental demonstration of entrapping streptavidin.

In the absence of biomolecules, air ($K=1$ and $N_f=0$) is present inside the cavity. The presence of biomolecules is simulated by considering that the cavity is filled with different materials (neutral biomolecules: $K_{bio} \neq 1$ and $N_f=0$; for charged biomolecules, $K_{bio} \neq 1$ and $N_f \neq 0$). Biomolecules inside the cavity change the gate oxide capacitance and lateral electric field which changes different device metrics, and this change in the device metric is responsible for its sensing ability. This dielectric modulation in the presence of biomolecules enables the MOSFET to act as a biosensing

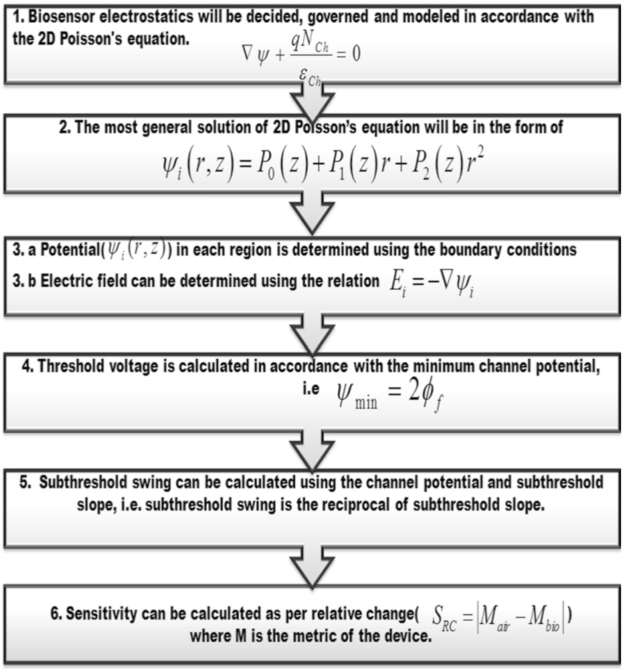
Table 2 S_{Vt}, S_{SS}, and S_{IOFF} of a DM-STGGS-MOSFET for different combinations of gates and source materials

Material		Biomolecule					
		Threshold voltage sensitivity (S_{Vt}: mV) [9] & S_{Vt} $= \lvert V_t$(no biomolecules) $- V_t$(with biomolecules)$\rvert$					
Source	Drain	Number of gates	Streptavidin	Protein	Biotin	APTES	Hydroprotein
Si	Si	SG	22.468	26.515	27.654	38.779	61.107
		DG	55.283	68.476	72.063	94.165	115.701
		TG	130.872	162.727	171.927	220.03	260.52
Ge	Si	SG	26.12	32.514	34.292	43.743	69.63
		DG	59.85	69.44	73.05	96.03	117.54
		TG	157.529	191.217	200.044	245.036	281.09

Material		Off-current sensitivity (S_{IOFF}: fA) and subthreshold swing sensitivity (S_{SS}: mV/decade) $S_{\mathrm{IOFF}} = \lvert I_{\mathrm{OFF}}$(no biomolecules) $- I_{\mathrm{OFF}}$ (with biomolecules)$\rvert$ and $S_{\mathrm{SS}} = \lvert$SS (no biomolecules) $-$ SS (withbiomolecules)$\rvert$										
Source	Drain Drain	No. of gates	Streptavidin		Protein		Biotin		APTES		Hydroprotein	
			S_{IOFF}	S_{SS}	S_{IOFF}	S_{SS}	S_{IOFF}	S_{SS}	S_{IOFF}	S_{SS}	S_{IOFF}	S_{SS}
Si	Si	SG	46.2	3.009	51.1	3.62	52.2	3.788	57.4	4.695	60.7	5.489
		DG	2990	3.374	3310	3.778	3380	3.893	3700	4.702	3870	5.686
		TG	31,200	6.088	31,600	7.835	31,610	8.308	31,700	10.776	31,710	13.35
Ge	Si	SG	62.3	3.216	68.9	3.899	70.5	4.088	77.4	5.13	81.8	6.064
		DG	4230	3.74	4700	3.94	4810	4.38	5320	5.19	5610	6.12
		TG	91,900	6.106	93,300	7.875	93,500	8.447	94,000	11.97	94,100	15.097
			(SG)Singlegate				(DG)Doublegate				(TG)Triplegate	
			ϕ_{M} 4.6 eV[43]				ϕ_{M1} 4.6 eV[43] ϕ_{M2} 4.53 eV[13]				ϕ_{M1} 5.10 eV(Au) [13] ϕ_{M2} 4.53 eV(Mo) [13] ϕ_{M3} 4.1 eV(Al) [38]	
			L_{G} 30 nm				$L_{\mathrm{G1}}/L_{\mathrm{G2}}$ 30 nm				$L_{\mathrm{G1}}/L_{\mathrm{G3}}$ 25 nm and L_{G2} 10 nm	

device. Different neutral biomolecules such as streptavidin ($K_{\mathrm{bio}} = 2.1$) [10], protein ($K_{\mathrm{bio}} = 2.5$) [9], biotin ($K_{\mathrm{bio}} = 2.63$) [48], ChOx ($K_{\mathrm{bio}} = 3.3$) [9], APTES ($K_{\mathrm{bio}} = 3.57$) [6], and hydroprotein ($K_{\mathrm{bio}} = 5$) [11] have been considered in this work. Charged biomolecules such as a single strand of DNA [49, 50], which is non-hybridized ($K_{\mathrm{bio}} = 5$: $N_f = -5 \times 10^{10}$ to -5×10^{12} cm^{-2}) has been considered.

An optimization in threshold voltage (V_t) has been done by altering the work function (Φ_{M}) of three gates so that the turn-on time remains the same for the SG-MOSFET. A similar optimization of work function was carried out by Kumari et al. [44]. This is how we aimed at $\Phi_{\mathrm{M1}} = 5.1$ eV, $\Phi_{\mathrm{M2}} = 4.53$ eV, and $\Phi_{\mathrm{M3}} = 4.1$ eV. All the gates used are compatible with complementary metal–oxide–semiconductor (CMOS) technology [13, 38]. Employing the threshold voltage optimization so that the three MOSFETs (SG/DG/TG) are at the same turn-on point of V_t and keeping the channel doping low are the two structural optimizations which have been considered in this work.



1. Biosensor electrostatics will be decided, governed and modeled in accordance with the 2D Poisson's equation. $\nabla \psi + \frac{qN_{Ch}}{\varepsilon_{Ch}} = 0$

2. The most general solution of 2D Poisson's equation will be in the form of $\psi_i(r,z) = P_0(z) + P_1(z)r + P_2(z)r^2$

3. a Potential($\psi_i(r,z)$) in each region is determined using the boundary conditions

3. b Electric field can be determined using the relation $E_i = -\nabla \psi_i$

4. Threshold voltage is calculated in accordance with the minimum channel potential, i.e $\psi_{\min} = 2\phi_f$

5. Subthreshold swing can be calculated using the channel potential and subthreshold slope, i.e. subthreshold swing is the reciprocal of subthreshold slope.

6. Sensitivity can be calculated as per relative change($S_{RC} = \lvert M_{air} - M_{bio} \rvert$) where M is the metric of the device.

Fig. 3 Flowchart of the process illustrating the analytical modeling approach

3 Analytical model

The potential and field distribution in the DM-STGGS-MOSFET can be easily studied by solving Poisson's equation, which helps to analyze and understands its electrostatics and behavior. The analytical model of channel potential, electric field, threshold voltage, and subthreshold swing is based on center-channel potential method. Figure 3 shows the flowchart of the process used in developing the analytical model.

3.1 Surface potential

The two-dimensional (2D) Poisson equation is given as

$$\frac{1}{r}\frac{\partial}{\partial r}\left[r\frac{\partial}{\partial r}\{\psi(r,z)\}\right]+\frac{\partial^2}{\partial z^2}\psi(r,z)=-\frac{qN_{Ch}}{\varepsilon_{Ch}} \tag{1}$$

where ψ(r,z) is the surface potential across the channel, r denotes the radius of channel, q is the electronic charge, N_{Ch} is the doping of channel, and ϵ_{Ch} is the permittivity of the channel material. The diameter of the channel is $2a$, so $r=0$ denotes the center of the channel with the potential expressed as ψ_{C}(z) or ψ(0,z), and $r=$a denotes the surface interface between the channel and oxide layer with potential expressed as $\psi_{\mathrm{I}}(z)$ or ψ(a,z). ψ(r,z) is an implicit function of r and z, so its general solution can be expressed in accordance with a parabolic potential profile.

$$\psi(r,z)=P_0(z)+P_1(z)r+P_2(z)r^2 \tag{2}$$

Different boundary conditions used are as follows:

(i) The potential at the source–channel boundary is equal to the built-in potential developed across it,

$$\psi\left(r,z_1\right)=V_{bi,1}=V_1=\frac{k_BT}{q}\left\{\ln\left(\frac{N_SN_{\mathrm{Ch}}}{n_{i(\mathrm{Si})}n_{\mathrm{i(Ge)}}}\right)\right\} \tag{3}$$

where $V_{\mathrm{bi,1}}$ is the built-in potential at the source–channel interface, n_{i}(Ge) and n_{i}(Si) are the intrinsic carrier concentration of germanium and silicon, respectively, k_{B} is Boltzmann constant and is equal to 1.38066×10^{-23} J/K, T is the temperature in Kelvin, and N_{S} is the source doping.

(ii) The potential at the drain–channel interface is equal to the sum of the built-in potential and drain-to-source voltage applied:

$$\psi\left(r,z_4\right)=V_{\mathrm{bi,2}}=V_4=\frac{k_BT}{q}\ln\left(\frac{N_{\mathrm{Ch}}N_D}{n^2_{\mathrm{i(Si)}}}\right) \tag{4}$$

where $V_{\mathrm{bi,2}}$ is the built-in potential at the source–channel interface, and N_{D} is the drain doping.

(iii) $\psi\left(r,z_2\right)=V_2$ and $\psi\left(r,z_3\right)=V_3$ (5)

(iv) The electric field is continuous at $z=z_2$ and $z=z_3$:

$$\left.\frac{\partial\{\psi(r,z)\}}{\partial z}\right|_{z=z_2^-}=\left.\frac{\partial\{\psi(r,z)\}}{\partial z}\right|_{z=z_2^+} \tag{6a}$$

$$\left.\frac{\partial\{\psi(r,z)\}}{\partial z}\right|_{z=z_3^-}=\left.\frac{\partial\{\psi(r,z)\}}{\partial z}\right|_{z=z_3^+} \tag{6b}$$

(v) The net electric field at the center of the channel is zero because of the radial symmetry:

$$\left.\frac{\partial\psi(r,z)}{\partial r}\right|_{r=0}=0 \tag{7}$$

(vi) The electric field is continuous at the Si–SiO$_2$ interface since the electric flux density is the same just above and below the Si–SiO$_2$ interface in the absence of any interface trap charges:

$$\varepsilon_{\mathrm{Ch}}\left.\frac{\partial\{\psi(r,z)\}}{\delta r}\right|_{r=a}=C_{\mathrm{ox}}\left[V_{\mathrm{GS}}-V_{\mathrm{FB}}-\psi(r,z)\right]\Big|_{r=a} \tag{8a}$$

$$\varepsilon_{\mathrm{Ch}}\left.\frac{\partial\{\psi(r,z)\}}{\delta r}\right|_{r=a}=\frac{C_{\mathrm{ox}}}{\varepsilon_{\mathrm{Ch}}}\left[V^*_{\mathrm{GS}}-\psi_I(z)\right] \tag{8b}$$

Using the boundary conditions (7)–(8), ψ(r,z) can be expressed in terms of $\psi_{\mathrm{I}}(z)$:

$$\psi(r,z)=\frac{1}{2\varepsilon_{\mathrm{Ch}}}\left[\psi_I(z)\left\{C_{\mathrm{ox}}a+2\varepsilon_{\mathrm{Ch}}\right\}-V^*_{\mathrm{GS}}C_{\mathrm{ox}}a\right]+\frac{C_{\mathrm{ox}}}{2\varepsilon_{Ch}}\left[V^*_{\mathrm{GS}}-\psi_I(z)\right] \tag{9}$$

where $V^*_{\mathrm{GS}}=V_{\mathrm{GS}}-V_{\mathrm{FB}}$ is the effective gate-to-source voltage, C_{OX} is the gate oxide capacitance, V_{GS} is the gate-to-source voltage, and V_{FB} is the flatband voltage.

Since $\psi(0,z)=\psi_{\mathrm{C}}(z)$, a relationship can be derived between $\psi_{\mathrm{C}}(z)$ and $\psi_{\mathrm{I}}(z)$ and can be expressed as

$$\psi_I(z)=\frac{\left[C_{\mathrm{ox}}aV^*_{\mathrm{GS}}+2\psi_C(z)\varepsilon_{\mathrm{Ch}}\right]}{2\varepsilon_{\mathrm{Ch}}+C_{\mathrm{ox}}a} \tag{10}$$

Putting (9) in (1) and using (10), it is possible to form a differential equation in terms of $\psi_{\mathrm{C}}(z)$ which can be expressed as

$$\mu^2\frac{\partial^2\psi_C(z)}{\partial z^2}+\theta=\psi_C(z) \tag{11}$$

where

$$\theta=V^*_{\mathrm{GS}}+\frac{qN_{\mathrm{Ch}}a}{2C_{\mathrm{ox}}}+\frac{qN_{\mathrm{Ch}}a^2}{4\varepsilon_{\mathrm{Ch}}} \tag{12a}$$

and

$$\mu^2 = \frac{2\varepsilon_{\text{Ch}}a + a^2C_{\text{ox}}}{4C_{\text{ox}}} \tag{12b}$$

The most general solution of (11) is [32]

$$\psi_C(z) = pe^{\frac{z}{\mu}} + qe^{\frac{-z}{\mu}} + \theta \tag{13}$$

Since the parameters μ and θ are material-dependent, its value will be different in different regions. Thus, $\psi_C(z)$ is expressed and specified differently in three regions:

Region A:$z_1 \leq z \leq z_2$

$$\psi_{C1}(z) = p_1e^{\frac{z}{\mu_1}} + q_1e^{\frac{-z}{\mu_1}} + \theta_1 \tag{14}$$

$$\mu_1 = \sqrt{\frac{2\varepsilon_{\text{Ch}}a + a^2C_{\text{ox1}}}{4C_{\text{ox1}}}} \tag{15a}$$

$$\theta_1 = V_{\text{GS}} - V_{\text{FB1}} + \frac{qN_{\text{Ch}}a}{2C_{\text{ox1}}} + \frac{qN_{\text{Ch}}a^2}{4\varepsilon_{\text{Ch}}} \tag{15b}$$

$$\frac{1}{C_{\text{ox1}}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} \tag{15c}$$

$$C_1 = \frac{\varepsilon_{\text{ox1}}}{a\ln\left(1 + \frac{t_1}{a}\right)}, C_2 = \frac{\varepsilon_{\text{bio}}}{a\ln\left(1 + \frac{t_2}{a}\right)} \tag{15d}$$

$$C_3 = \frac{\varepsilon_{\text{ox2}}}{a\ln\left(1 + \frac{t_3}{a}\right)}, V_{\text{FB1}} = \phi_{M1} - \phi_{\text{Si}} - \frac{qN_f}{C_2} \tag{15e}$$

where φ_{M1} and φ_{Si} are the work function of gate 1 and silicon, respectively, N_{f} is the charge density of biomolecules, and ϵ_{ox1}/ ϵ_{ox2}/ ϵ_{bio} are the dielectric constants of SiO_2, HfO_2, and biomolecules, respectively. Flatband voltage changes significantly in the presence of charged biomolecules.

Region B:$z_2 \leq z \leq z_3$.

$$\psi_{C2}(z) = p_2e^{\frac{z}{\mu_2}} + q_1e^{\frac{-z}{\mu_2}} + \theta_2 \tag{16}$$

$$\mu_2 = \sqrt{\frac{2\varepsilon_{\text{Ch}}a + a^2C_{\text{ox2}}}{4C_{\text{ox2}}}} \tag{17a}$$

$$\theta_2 = V_{\text{GS}} - V_{FB2} + \frac{qN_{\text{Ch}}a}{2C_{\text{ox2}}} + \frac{qN_{\text{Ch}}a^2}{4\varepsilon_{\text{Ch}}} \tag{17b}$$

$$C_{\text{ox2}} = \frac{\varepsilon_{\text{ox1}}}{a\ln\left(1 + \frac{t_{\text{eff}}}{a}\right)}, t_{\text{eff}} = t_1 + t_{23}\left(\frac{\varepsilon_{\text{SiO}_2}}{\varepsilon_{\text{HfO}_2}}\right) \tag{17c}$$

$$V_{\text{FB2}} = \phi_{M2} - \phi_{\text{Si}}, t_{23} = t_2 + t_3 \tag{17d}$$

where φ_{M2} is the work function of gate 2.

Region C:$z_3 \leq z \leq z_4$

$$\psi_{C3}(z) = p_3e^{\frac{z}{\mu_1}} + q_3e^{\frac{-z}{\mu_1}} + \theta_3 \tag{18}$$

$$\mu_3 = \sqrt{\frac{2\varepsilon_{\text{Ch}}a + a^2C_{\text{ox3}}}{4C_{\text{ox3}}}} \tag{19a}$$

$$\theta_3 = V_{\text{GS}} - V_{\text{FB3}} + \frac{qN_{\text{Ch}}a}{2C_{\text{ox3}}} + \frac{qN_{\text{Ch}}a^2}{4\varepsilon_{\text{Ch}}} \tag{19b}$$

$$C_{ox3} = C_{ox1}, V_{FB3} = \phi_{M3} - \phi_{Si} - \frac{qN_f}{C_2} \tag{19c}$$

where φ_{M3} is the work function of gate 3.

Using the boundary conditions (3)–(8), coefficients p_1, q_1, p_2, q_2, p_3, and q_3 can be determined.

$$p_1 = \alpha\left[\theta_1\left(e^{\frac{-z_1}{\mu_1}} - e^{\frac{-z_2}{\mu_1}}\right) + V_{bi,1}e^{\frac{-z_2}{\mu_1}} - V_2e^{\frac{-z_1}{\mu_1}}\right] \tag{20}$$

$$q_1 = \alpha\left[\theta_1\left(e^{\frac{z_2}{\mu_1}} - e^{\frac{z_1}{\mu_1}}\right) + V_2e^{\frac{z_1}{\mu_1}} - V_{bi,1}e^{\frac{z_2}{\mu_1}}\right] \tag{21}$$

$$p_2 = \beta\left[\theta_2\left(e^{\frac{-z_2}{\mu_2}} - e^{\frac{-z_3}{\mu_2}}\right) + V_2e^{\frac{-z_3}{\mu_2}} - V_3e^{\frac{-z_2}{\mu_2}}\right] \tag{22}$$

$$q_2 = \beta\left[\theta_2\left(e^{\frac{z_3}{\mu_2}} - e^{\frac{z_2}{\mu_2}}\right) + V_3e^{\frac{z_2}{\mu_2}} - V_2e^{\frac{z_3}{\mu_2}}\right] \tag{23}$$

$$p_3 = \gamma\left[\theta_3\left(e^{\frac{-z_3}{\mu_3}} - e^{\frac{-z_4}{\mu_3}}\right) - V_{DS}e^{\frac{-z_3}{\mu_3}} + V_3e^{\frac{-z_4}{\mu_3}} - V_{bi,2}e^{\frac{-z_3}{\mu_3}}\right] \tag{24}$$

$$q_3 = \gamma\left[\theta_3\left(e^{\frac{z_4}{\mu_3}} - e^{\frac{z_3}{\mu_3}}\right) + V_{DS}e^{\frac{z_3}{\mu_3}} - V_3e^{\frac{z_4}{\mu_3}} + V_{bi,2}e^{\frac{z_3}{\mu_3}}\right] \tag{25}$$

where

$$\alpha = \frac{\text{csch}\left(\frac{z_1 - z_2}{\mu_1}\right)}{2} \tag{26a}$$

$$\beta = \frac{\text{csch}\left(\frac{z_2 - z_3}{\mu_2}\right)}{2}, \gamma = \frac{\text{csch}\left(\frac{z_3 - z_4}{\mu_3}\right)}{2} \tag{26b}$$

Since $p_i = f(V_2, V_3)$ and $q_i = f(V_2, V_3)$, the two unknown potentials V_2 and V_3 can be determined using the boundary conditions (10)–(11) and can be expressed as [6]

$$V_2 = e_4(se_3 - te_2), V_3 = e_4(se_2 - te_1) \tag{27}$$

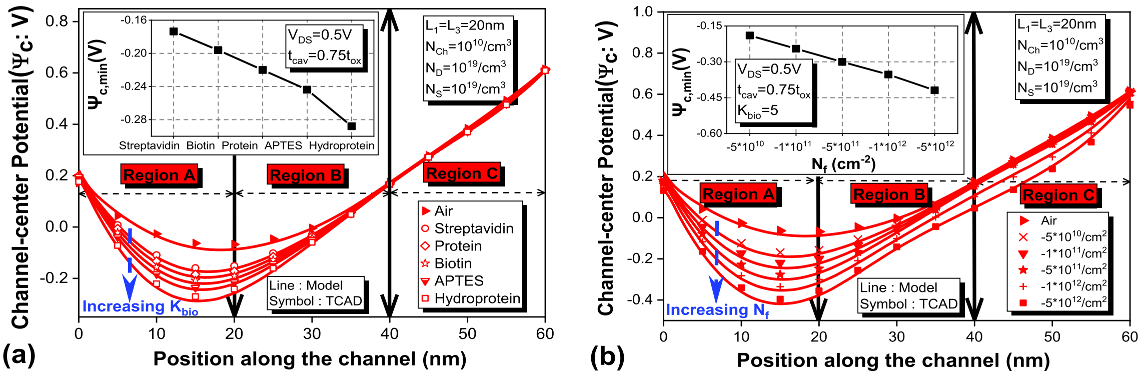


Fig. 4 Potential variation along the channel in the DM-STGGS-MOSFET for **a** neutral biomolecules and **b** charged biomolecules. (Inset) Minimum channel-center potential for different biomolecules

The values of e_1, e_2, e_3, e_4, s, and t are given in Appendix A.

Since $E = -\nabla\psi$, the electric field distribution in each region can be written as [6]

$$E_{Z_i} = -\frac{1}{\mu_i}\left(p_i e^{\frac{z}{\mu_i}} - q_i e^{\frac{-z}{\mu_i}}\right) \tag{28}$$

3.2 Threshold voltage

Threshold voltage can be obtained graphically or analytically. Conventionally, threshold voltage is defined as the minimum gate voltage where charge density inversion occurs in the substrate [9, 38]. Minimum potential can be also be determined separately for three different regions, and $\psi_{\min}$ can subsequently be defined as $\psi_{\min} = \min(\psi_{C1}, \psi_{C2}, \psi_{C3})$. Analytically, the minimum channel center potential occurs in region I, which is due to the drain voltage that continues to decrease while moving towards the source from the drain. Differentiating $\psi_{C1}(z)$ will give the precise location of the minimum center-channel potential.

$$\left.\frac{d\psi_{C_1}(z)}{dz}\right|_{z=z_{\text{Minima}}} = 0 \rightarrow z_{\text{Minima}} = \mu_1 \ln\left(\sqrt{\frac{q_1}{p_1}}\right) \tag{29}$$

Substituting (29) in (14) and replacing V_{GS} by V_t in the expression will yield the analytical expression for threshold voltage.

$$V_t = \frac{-n_{12} + \sqrt{n_{12}^2 - n_{11}n_{13}}}{2n_{12}} \tag{30}$$

Values of n_{11}, n_{12}, and n_{13} are given in Appendix B.

3.3 Drain current and subthreshold swing

Using the minimum surface potential calculated above, it is possible to develop the analytical model for the drain current (I_{DS}) and subthreshold swing (SS). The drain current (I_{DS}) in the linear region is calculated separately in the three different regions as [51, 52]

$$I_{\text{Lin},A} = \frac{aC_{\text{ox},1}\mu_n}{L_1'}\left[\left(V_{\text{GS}} - V_{\text{Th}}\right)\left(V_2 - V_1\right) - \frac{\left(V_2 - V_1\right)^2}{2}\right] \tag{31}$$

$$I_{\text{Lin},B} = \frac{aC_{\text{ox},2}\mu_n}{L_2'}\left[\left(V_{\text{GS}} - V_{\text{Th}}\right)\left(V_3 - V_2\right) - \frac{\left(V_3 - V_2\right)^2}{2}\right] \tag{32}$$

$$I_{\text{Lin},C} = \frac{aC_{\text{ox},3}\mu_n}{L_3'}\left[\left(V_{\text{GS}} - V_{\text{Th}}\right)\left(V_4 - V_3\right) - \frac{\left(V_4 - V_3\right)^2}{2}\right] \tag{33}$$

For calculating the drain current in the saturation region, replace V_{DS} by $V_{\text{DS,Sat}}$ which is given by [51, 52]

$$V_{\text{DS,Sat}} = \frac{\left(V_{\text{GS}} - V_{\text{Th}}\right)}{1 + \frac{\mu_{\text{efld}}\left(V_{\text{GS}} - V_{\text{Th}}\right)}{\left(L_1 + L_2 + L_3\right)v_{\text{Sat}}}} \tag{34}$$

$$\mu_{\text{efld}} = \frac{\mu_n}{\left\{1 - \zeta(V_{\text{GS}} - V_{\text{Th}}\right\}\left\{1 + \Omega\frac{V_{\text{DS}}\mu_n}{\left(L_1 + L_2 + L_3\right)v_{\text{Sat}}}\right\}} \tag{35}$$

$$\Omega = \left[\frac{V_{\text{DS}}\mu_n}{\left(L_1 + L_2 + L_3\right)v_{\text{Sat}}}\right]\left[1.5 + \left\{\frac{V_{\text{DS}}\mu_n}{\left(L_1 + L_2 + L_3\right)v_{\text{Sat}}}\right\}\right]^{-1} \tag{36}$$

where μ_{efld} and v_{Sat} are the maximum low-field mobility and saturation velocity of electrons ($v_{\text{Sat}} = 1 \times 10^7$ cm/s),

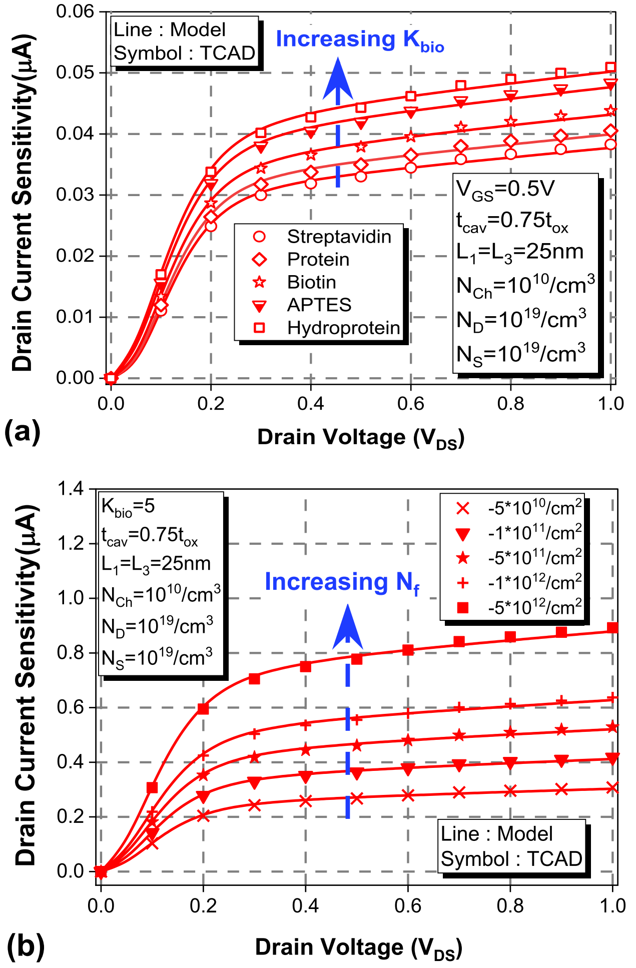


Fig. 5 Drain current sensitivity (S_{IDS}) as a function of drain-to-source voltage (V_{DS}) in the DM-STGGS-MOSFET for **a** neutral biomolecules and **b** charged biomolecules

respectively, μ_n is the mobility of electrons, and ζ is the fitting parameter whose value taken here is 0.43.

Subthreshold swing is another critical analog parameter which is the reciprocal of subthreshold slope and is calculated by the formula [53]

$$\mathrm{SS} = V_T \ln 10 \left\{ \left(\frac{\partial \psi_C(a,z)}{\partial V_{\mathrm{GS}}} \right)^{-1} \right\} \Bigg|_{Z_{\min}} \tag{37}$$

3.4 Sensitivity

Sensitivity of the biosensor is calculated according to the following formula:

$$S_M = \left| M_{\mathrm{Air}} - M_{\mathrm{Bio}} \right| \tag{38}$$

where SM is the sensitivity of the device for metric M, and M can be any metric of the biosensor such as threshold voltage, subthreshold swing, on-current, or off-current. $\mathrm{M}_{\mathrm{Air}}$ and $\mathrm{M}_{\mathrm{Bio}}$ are the values of metric M in the absence and presence of biomolecules, respectively.

4 Results and discussion

The variation in the channel-center potential (ψ_c) along the channel in the DM-STGGS-MOSFET is shown in Fig. 4. Figure 4a shows the potential variation for neutral biomolecules, whereas Fig. 4b shows the potential variation for charged biomolecules. The relative change in the potential is important in determining the sensitivity of the biosensor. It can be clearly observed that the potential profile along the channel changes in the presence of biomolecules. When the cavity is filled with different biomolecules ($K_{\mathrm{bio}} > 1$), the effective gate oxide capacitance increases which increases the coupling between the gate and the charge carriers flowing across the channel. Hence, the channel-center potential decreases, which mean that more gate-to-source voltage is needed to deplete the channel completely; therefore, threshold voltage increases in the presence of biomolecules, which increases the overall threshold voltage sensitivity. High coupling between the gate and the channel [25, 54] decreases the channel potential, and this coupling continues to increase with the increasing K_{bio}. The presence of negatively charged biomolecules further increases the coupling because of its higher binding capability [55, 56] than the neutral biomolecules, which results in the decrease in the channel potential. This variation in the channel potential can be clearly seen in Fig. 4. A good agreement is noted between the developed analytical model and TCAD simulation. The inset in Fig. 4 shows the value of minimum channel-center potential for different biomolecules in three different regions.

Figure 5a and b shows the drain current sensitivity (S_{IDS}) as a function of drain voltage (V_{DS}) in the DM-STGGS-MOSFET for different neutral and charged biomolecules, respectively, at $V_{\mathrm{GS}} = 0.5$ V. The binding capability of negatively charged biomolecules is larger than that of positively charged and neutral biomolecules. At a constant gate voltage, drain current decreases with the increasing dielectric constant of the biomolecule due to the increased effective gate oxide capacitance. But the relative change in the drain current continues to increase with the increasing K_{bio}. The presence of negatively charged biomolecules further decreases the drain current due to more control over the flow of charge carriers across the channel.

The germanium source [20] has a larger number of intrinsic charge carriers than silicon, which participates in the conduction process and is also responsible for the biosensing action in the DM-STGGS-MOSFET. Germanium has nearly 1650 times more intrinsic charge carriers at room temperature than silicon. When the numbers of

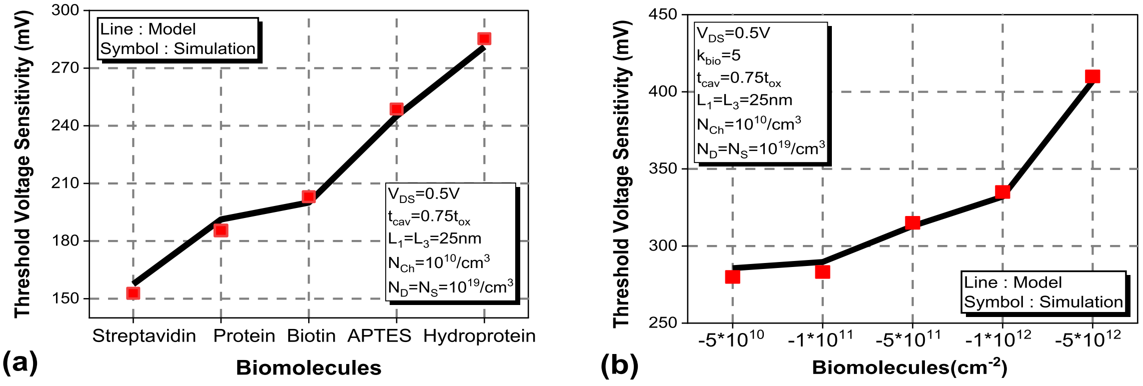


Fig. 6 Threshold voltage sensitivity (S_{Vt}) variation in the DM-STGGS-MOSFET for **a** neutral biomolecules and **b** charged biomolecules

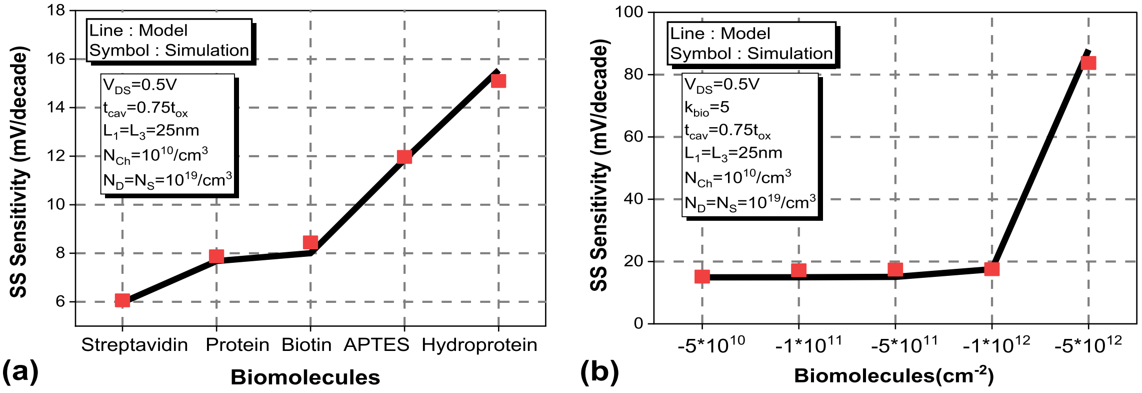


Fig. 7 Subthreshold swing sensitivity (S_{SS}) variation in the DM-STGGS-MOSFET for **a** neutral biomolecules and **b** charged biomolecules

charge carriers increase, the change in the different device metrics (threshold voltage and subthreshold swing) will be greater, since the flow of a large number of charge carriers is affected (which changes the threshold voltage and subthreshold swing) when the biomolecules are localized inside the cavity [7]. Hence, this will increase the sensitivity of the proposed biosensor.

Figures 6 and 7 show variation of the threshold voltage sensitivity (S_{Vt}) and subthreshold swing sensitivity (S_{SS}) in the DM-STGGS-MOSFET for different biomolecules. Threshold voltage (V_{t}) increases while subthreshold swing (SS) decreases with the increase in the dielectric constant because the subthreshold swing is inversely proportional to the effective dielectric constant of the gate oxide. The gate oxide capacitance increases with the increasing K_{bio} [6, 54], which is the primary reason for a significant increase in the relative change in threshold voltage and subthreshold swing. Hence, S_{Vt} and S_{SS} increase with the increasing dielectric constant of the biomolecule. The presence of negatively charged biomolecules inside the cavity further increases the threshold voltage because more gate-to-source voltage is needed to turn on the device. At constant gate voltage, high on-current is obtained at low charge density of negatively charged biomolecules [57]. This implies a steeper slope in the log (I_{DS})–V_{GS} curve which indicates high subthreshold slope but low subthreshold swing at low charge density of biomolecules. High subthreshold slope indicates high current variation for a given operating range of gate voltage [58]. However, high current variation will ultimately result in low subthreshold swing. The subthreshold swing decreases with respect to the unfilled cavity case, but the relative change in subthreshold swing increases with the increasing charge density of negatively charged biomolecules.

Table 3 shows the threshold voltage sensitivity (S_{Vt}) and subthreshold swing sensitivity (S_{SS}) in the DM-STGGS-MOSFET for different parameter variation. With the increase in the length or thickness of the cavity, the dimension of the cavity increases, which can entrap more biomolecules inside it, and hence, the sensitivity of the biosensor increases significantly. Thus, a significant increase in threshold voltage sensitivity can be seen when the length or thickness of the cavity increases. It is quite interesting to note that the biosensor under consideration is slightly more sensitive

Table 3 S_{Vt} and S_{SS} variation in the DM-STGGS-MOSFET for different parameter variation

Biomolecules	S_{Vt} and S_{SS}	Different parameters											
		Length of cavity (L_{CAV}: $L_1=L_3=\mathrm{L}_{\mathrm{Cav}}$)			Cavity thickness ($t_{\mathrm{CAV}}=t_2$)			Radius of channel (R_{CH}: a)			Ambient temperature (T)		
		25 nm	20 nm	15 nm	6 nm	4 nm	2 nm	10 nm	11 nm	12 nm	250 K	300 K	350 K
Neutral biomolecules													
Streptavidin	S_{Vt}	157.53	94.797	36.605	157.53	146.03	82.669	157.53	145.92	138.95	152.77	157.53	161.94
	S_{SS}	6.06	1.66	1.514	6.06	6.051	6.004	6.06	4.358	4.005	4.094	6.06	9.573
Protein	S_{Vt}	191.22	115.44	45.309	191.22	168.7	94.461	191.22	177.79	166.56	182.15	191.22	193.83
	S_{SS}	7.875	2.132	2.1	7.875	7.860	7.822	7.875	5.896	4.837	5.579	7.875	11.919
Biotin	S_{Vt}	200.05	121.08	47.807	200.05	174.6	97.487	200.05	186.23	176.5	189.83	200.05	202.2
	S_{SS}	8.447	2.512	2.316	8.447	8.399	8.109	8.447	6.405	5.423	6.027	8.447	12.648
APTES	S_{Vt}	245.04	151.45	62.334	245.04	204.36	112.52	245.04	229.91	217.99	228.82	245.04	254.83
	S_{SS}	11.968	3.358	3.323	11.968	11.919	9.497	11.968	9.555	8.867	8.956	11.968	16.78
Hydroprotein	S_{Vt}	281.1	178.14	76.852	281.1	227.94	124.1	281.1	265.85	250.3	259.64	281.1	297.8
	S_{SS}	15.097	4.851	4.324	15.097	14.225	10.405	15.097	12.734	11.392	11.51	15.097	20.582
Charged biomolecules													
−5e10/cm^2	S_{Vt}	*285.62*	182.32	79.723	285.62	230.51	125.28	285.62	270.1	255.07	263.17	285.62	305.16
	S_{SS}	*15.146*	4.858	4.351	15.146	14.235	10.388	15.146	12.785	11.452	11.559	15.146	20.649
−1e11/cm^2	S_{Vt}	*289.65*	186	82.327	289.65	232.9	126.4	289.65	273.96	259.37	266.41	289.65	311.09
	S_{SS}	*15.195*	4.868	4.37	15.195	14.239	10.444	15.195	12.834	11.513	15.152	15.195	20.714
−5e11/cm^2	S_{Vt}	*312.96*	206.88	97.748	312.96	247.86	133.56	312.96	297.9	265.32	285.91	312.96	340.56
	S_{SS}	*15.423*	4.956	4.46	15.423	14.289	10.569	15.423	13.04	11.931	14.621	15.423	21.004
−1e12/cm^2	S_{Vt}	*331.99*	223.7	109.97	331.99	259.61	139.9	331.99	317.8	268.13	302.07	331.99	362.66
	S_{SS}	*15.591*	4.9908	4.812	15.591	14.90	10.969	15.591	14.148	12.403	14.985	15.591	21.105
−5e12/cm^2	S_{Vt}	*406.84*	293.35	170.85	406.84	321.17	492.8	406.84	350.02	297.14	361.5	406.84	452.16
	S_{SS}	*83.727*	20.34	5.7642	83.727	80.852	75.232	83.727	61.986	53.953	74.249	83.727	86.979
S_{Vt}: [mV]	S_{SS}: [mV/dec]												

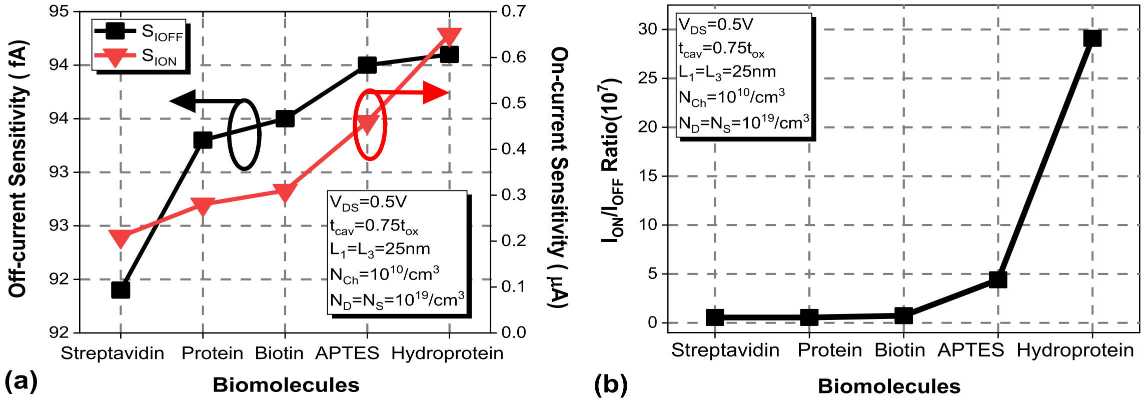


Fig. 8 **a** Off-current sensitivity (S_{IOFF}) and on-current sensitivity (S_{ION}) variation and **b** the $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ ratio variation for different biomolecules in the DM-STGGS-MOSFET

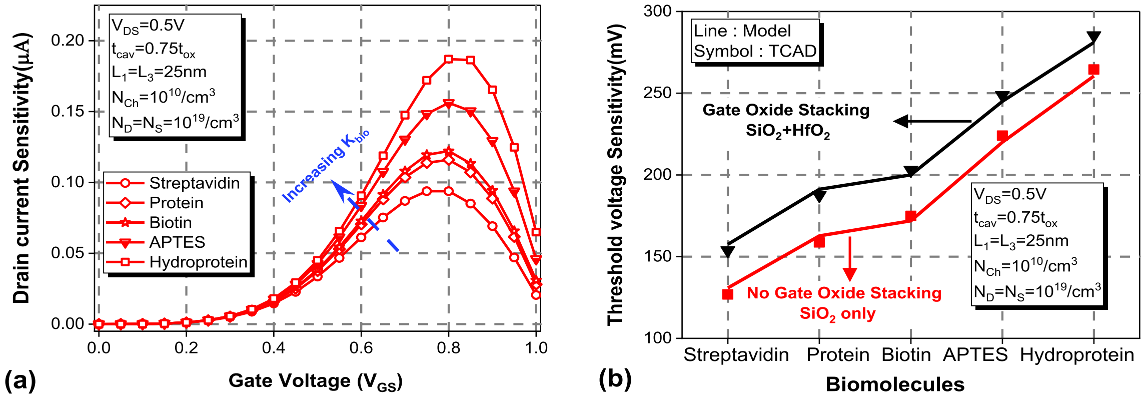


Fig. 9 **a** Drain current sensitivity (S_{IDS}) as a function of gate-to-source voltage (V_{GS}) and **b** threshold voltage sensitivity (S_{Vt}) in the presence and absence of gate oxide stacking in the DM-STGGS-MOSFET

to the length of the cavity than the thickness of the cavity, which is because the gate tends to lose control over the flow of charge carriers with the increasing gate oxide thickness (thickness of the cavity is increased). Increasing the radius or thickness of the channel will decrease the threshold voltage because the same amount of drain current can be obtained at a lower V_{GS} due to an increase in the cross-sectional area of the channel, and hence, the threshold voltage sensitivity decreases with the increasing radius of the channel. We can obtain a larger S_{Vt} at a lower channel radius, but decreasing the channel radius below 10 nm will make the quantum effects predominant [59], due to which S_{Vt} might decrease. Increasing the radius of the channel increases the channel cross-sectional area, but interestingly, subthreshold swing sensitivity decreases, making the device less sensitive to biomolecules with a larger channel radius. Increasing the temperature will decrease the threshold voltage due to the generation of a large number of charge carriers at high temperature. Thus, the same amount of drain current can be obtained at a lower V_{GS}, which decreases the threshold voltage of the device, but interestingly, S_{Vt} increases with the increasing temperature due to the increase in the relative change in threshold voltage. When more biomolecules are entrapped inside the cavity, a larger variation in subthreshold swing can be observed, and hence the subthreshold swing sensitivity increases with the increasing length/thickness of the cavity. A similar pattern in subthreshold swing sensitivity can be observed when temperature is increased due to generation of large number of charge carriers, and hence a relative change in subthreshold swing is more at higher temperatures. An increase in the ambient temperature will increase both on-current (I_{ON}) and off-current (I_{OFF}) in the biosensor, but $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ decreases, which indicates a substantial decrease in subthreshold slope but an increase in the subthreshold swing. A decrease in the value of subthreshold swing is a good indicator of improved switching performance. The maximum S_{Vt} obtained here increases to 203% for $K_{\mathrm{bio}} = 5$ and $N_f = -5 \times 10^{12}/\mathrm{cm}^2$.

Figure 8a shows the off-current sensitivity (S_{IOFF}) and on-current sensitivity (S_{ION}) variation, whereas Fig. 8b

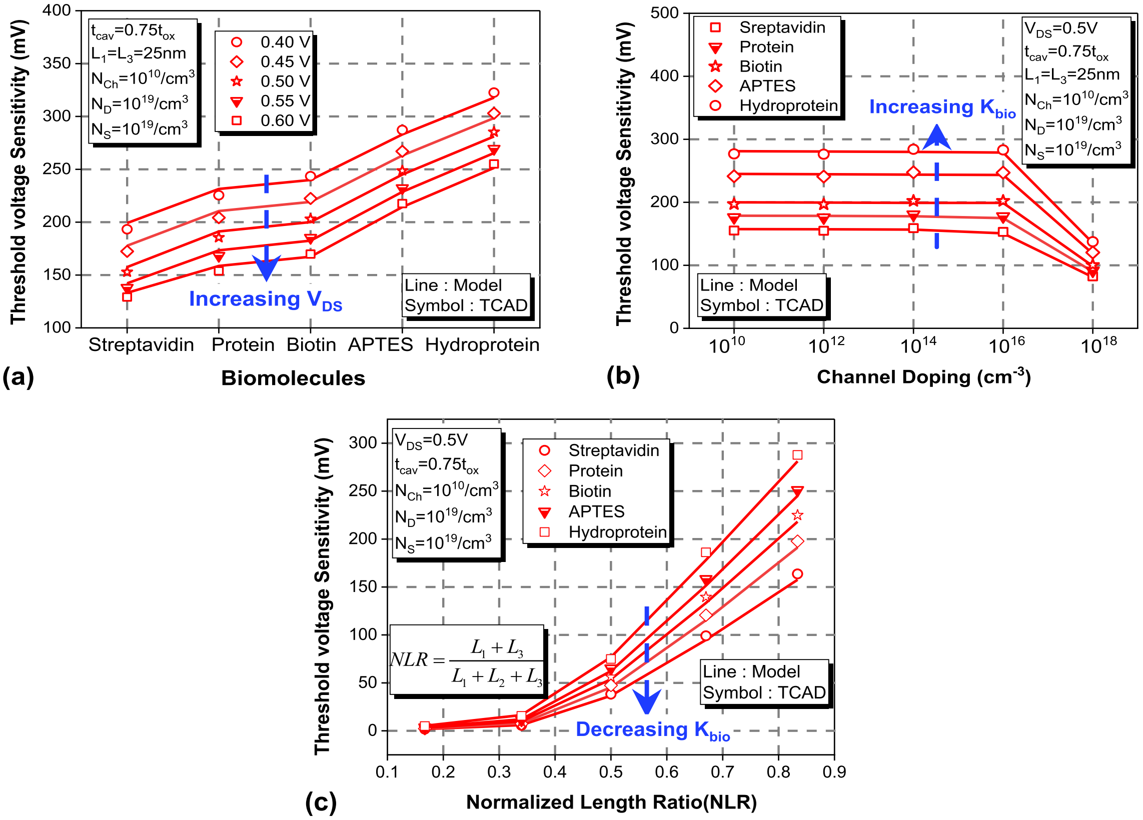


Fig. 10 Effect of **a** drain voltage (V_{DS}), **b** channel doping (N_{Ch}), and **c** normalized cavity length ratio (NLR) on threshold voltage sensitivity (S_{Vt}) in the DM-STGGS-MOSFET

shows the plot of the I_{ON}/I_{OFF} ratio for different biomolecules in the DM-STGGS-MOSFET. The increasing K_{bio} of the biomolecule results in the enhanced gate oxide capacitance, and hence the gate has more control over the flow of charge carriers, and the on-current decreases. The current in the biosensor is not zero at $V_{GS} = V_{Th}$ because of the subthreshold current. Due to an increase in effective oxide capacitance, the channel is weakly inverted even before reaching the threshold voltage, which decreases the off-current [39]. Thus, both the on-current and off-current decrease, but the relative change in I_{ON} and I_{OFF} increases. However, S_{ION} increases significantly more when compared to S_{IOFF}. The I_{ON}/I_{OFF} ratio also increases with the increasing K_{bio} (decrease in off-current is dominant, which increases the I_{ON}/I_{OFF} ratio), which is desirable if the biosensor is to be tested with high-frequency test signals. S_{ION} increases to 100% for $K_{bio} = 5$, and the I_{ON}/I_{OFF} ratio changes roughly around 5800 times when the cavity is filled with biomolecule having a dielectric constant of 5. This indicates that the I_{ON}/I_{OFF} ratio can be used as a sensing parameter to analyze the sensitivity of a biosensor along with the other device metrics, which can increase the reliability of the biosensor.

Figure 9a shows the plot of drain current sensitivity (S_{IDS}) as a function of gate-to-source voltage (V_{GS}) in the DM-STGGS-MOSFET. Drain current flowing in the biosensor is highly sensitive to biomolecules when V_{GS} is roughly around 0.8 V. S_{IDS} further increases with the increase in dielectric constant of the biomolecule. The maximum drain current sensitivity obtained is roughly 67% obtained for $K_{bio} = 5$. Figure 9b shows the plot of threshold voltage sensitivity (S_{Vt}) in the DM-STGGS-MOSFET with and without gate stacking. S_{Vt} significantly increases when the biosensor is operated with a gate stack, which shows the dominance of gate oxide stacking over a single-gate oxide layer. With the gate stack, effective gate oxide capacitance increases, which changes the potential and field distribution across the channel. The increase in the effective gate oxide due to gate stacking increases the coupling between the gate and the charge carriers. Hence, a larger relative change in threshold voltage increases the overall threshold voltage sensitivity with the use of gate oxide stacking. The results obtained are in accordance with the previously reported findings [6].

Figure 10 shows the effect of various parameters on threshold voltage sensitivity (S_{Vt}) in the DM-STGGS-MOSFET. Figure 10a shows the effect of drain voltage on threshold voltage sensitivity. With the increase in drain voltage, less gate voltage will be required to turn on the device, which decreases the extrinsic threshold voltage, and hence S_{Vt} decreases with the increasing drain voltage. Figure 10b shows the effect of channel doping on threshold voltage sensitivity. Increasing the channel doping makes it difficult to turn on the device at low V_{GS} because more gate voltage will be needed for the complete depletion of the channel [6]. Hence, threshold voltage increases but S_{Vt} decreases with the increase in channel doping in the presence of biomolecules. Figure 10c shows the effect of normalized cavity length ratio (NLR) on threshold voltage sensitivity. It can be clearly seen that increasing the NLR increases the threshold voltage sensitivity, because with the increasing NLR, more biomolecules can be accommodated inside the cavity, and thus the device becomes more sensitive to biomolecules. The biosensor shows maximum sensitivity at an NLR of roughly 0.83. This means that a high NLR can essentially increase the biosensor sensitivity to biomolecules.

5 Conclusion

An analytical model of a biosensor designed with a surrounding-triple-gate MOSFET (DM-STGGS-MOSFET) incorporating gate stacking and a germanium source has been proposed in this paper. The analytical model which is based upon the center-potential method and parabolic approximation has shown excellent agreement with the simulated results. Results shows a significant improvement in the sensitivity when the silicon source is replaced with a germanium source. The achieved sensitivity exceeded 200% in some cases, which solidifies its biosensing performance. The biosensor has been tested with different neutral and charged biomolecules, and the sensitivity of the device has been studied by varying different parameters such as cavity length, cavity thickness, channel doping, channel radius, drain voltage, and operating temperature. The excellent results and enhanced sensitivity make the device a potential candidate in the field of bioelectronics, since a small structural variation in the DM-STGGS-MOSFET can be done to detect hazardous or inert gases, and the device can also be used to detect different enzymes, chemicals, or potential biological substances needed for crop damage pre-assessment.

Appendix A

$$e_1 = 2\left[\frac{\alpha}{\mu_1}\cosh\left(\frac{z_1 - z_2}{\mu_1}\right) + \frac{\beta}{\mu_2}\cosh\left(\frac{z_2 - z_3}{\mu_2}\right)\right],$$

$$e_2 = 2\frac{\beta}{\mu_2},$$

$$e_3 = 2\left[\frac{\beta}{\mu_2}\cosh\left(\frac{z_2 - z_3}{\mu_2}\right) + \frac{\gamma}{\mu_3}\cosh\left(\frac{z_3 - z_4}{\mu_3}\right)\right],$$

$$e_4 = \left(e_1 e_3 - e_{2}^{2}\right)^{-1}$$

$$s = \frac{\alpha}{\mu_1}s_1 - \frac{\beta}{\mu_2}s_2, t = \frac{\gamma}{\mu_3}t_1 - \frac{\beta}{\mu_2}t_2$$

$$s_1 = s_{1a} - s_{1b}, s_2 = s_{2a} - s_{2b}$$

$$s_{1a} = \theta_1 e^{\frac{z_2}{\mu_1}}\left(e^{\frac{-z_1}{\mu_1}} - e^{\frac{-z_2}{\mu_1}}\right)$$

$$s_{1b} = \theta_1 e^{\frac{-z_2}{\mu_1}}\left(e^{\frac{z_2}{\mu_1}} - e^{\frac{z_1}{\mu_1}}\right) - 2V_{bi,1}$$

$$s_{2a} = \theta_2 e^{\frac{z_2}{\mu_2}}\left(e^{\frac{-z_2}{\mu_2}} - e^{\frac{-z_3}{\mu_2}}\right), s_{2b} = \theta_2 e^{\frac{-z_2}{\mu_2}}\left(e^{\frac{z_3}{\mu_2}} - e^{\frac{z_2}{\mu_2}}\right)$$

$$t_1 = t_{1a} - t_{1b}, t_2 = t_{2a} - t_{2b}, t_{1a} = \theta_3 e^{\frac{z_3}{\mu_3}}\left(e^{\frac{-z_3}{\mu_3}} - e^{\frac{-z_4}{\mu_3}}\right) - 2V_{bi,2}$$

$$t_{1b} = \theta_3 e^{\frac{-z_3}{\mu_3}}\left(e^{\frac{z_4}{\mu_3}} - e^{\frac{z_3}{\mu_3}}\right) + 2V_{DS},$$

$$t_{2a} = \theta_2 e^{\frac{z_3}{\mu_2}}\left(e^{\frac{-z_2}{\mu_2}} - e^{\frac{-z_3}{\mu_2}}\right)$$

$$t_{2b} = \theta_2 e^{\frac{-z_3}{\mu_2}}\left(e^{\frac{z_3}{\mu_2}} - e^{\frac{z_2}{\mu_2}}\right)$$

Appendix B

$$n_{11} = 4d_{11} - 1, n_{12} = 4d_{22} + 2\xi_1 + 4\phi_f,$$

$$n_{13} = 4d_3 - 4\phi_f^2 - \xi_1^2 - 4\phi_f\xi_1,$$

$$d_{11} = d_{33} = d_1 d_3, d_{22} = d_1 + d_3$$

$$d_2 = \alpha n_1, \quad d_4 = \alpha n_2, \quad \phi_f = V_T \ln\left(\frac{N_{Ch}}{n_{i(Si)}}\right)$$

$$d_1 = \alpha\left[\left(e^{\frac{-z_1}{\mu_1}} - e^{\frac{-z_2}{\mu_1}}\right) - e^{\frac{-z_1}{\mu_1}} e_3 e_4\left(m_1 - m_2\right) - e^{\frac{-z_1}{\mu_1}} e_2 e_4\left(m_4 - m_5\right)\right]$$

$$d_3 = \alpha\left[\left(e^{\frac{z_2}{\mu_1}} - e^{\frac{z_1}{\mu_1}}\right) + e^{\frac{z_1}{\mu_1}} e_3 e_4\left(m_1 - m_2\right) - e^{\frac{z_1}{\mu_1}} e_2 e_4\left(m_4 - m_5\right)\right], V_T = \frac{k_B T}{q},$$

$$n_1 = \left[-\xi_1\left(e^{\frac{-z_1}{\mu_1}} - e^{\frac{-z_2}{\mu_1}}\right) + V_{bi,1}e^{\frac{-z_2}{\mu_1}} - e^{\frac{-z_1}{\mu_1}} e_3e_4\left(-m_1\xi_1 + m_2\xi_2 - m_3\right)\right.$$
$$\left. + e^{\frac{-z_1}{\mu_1}} e_2e_4\left(-m_4\xi_3 + m_5\xi_2 - m_6\right)\right],$$
$$n_2 = \left[-\xi_1\left(e^{\frac{z_2}{\mu_1}} - e^{\frac{z_1}{\mu_1}}\right) - V_{bi,1}e^{\frac{z_2}{\mu_1}} + e^{\frac{z_1}{\mu_1}} e_3e_4\left(-m_1\xi_1 + m_2\xi_2 - m_3\right)\right.$$
$$\left. - e^{\frac{z_1}{\mu_1}} e_2e_4\left(-m_4\xi_3 + m_5\xi_2 - m_6\right)\right],$$
$$\xi_1 = \frac{qN_{Ch}a}{2C_{ox_1}} + \frac{qN_{Ch}a^2}{4\varepsilon_{Ch}} - V_{FB_1},$$
$$\xi_2 = \frac{qN_{Ch}a}{2C_{ox_2}} + \frac{qN_{Ch}a^2}{4\varepsilon_{Ch}} - V_{FB_2},$$
$$\xi_3 = \frac{qN_{Ch}a}{2C_{ox_3}} + \frac{qN_{Ch}a^2}{4\varepsilon_{Ch}} - V_{FB_3},$$
$$m_3 = \frac{-2\alpha}{\mu_1}V_{bi_1}, m_6 = \frac{2\gamma}{\mu_3}\left[V_{bi_2} + V_{DS}\right]$$

$$m_1 = \frac{\alpha}{\mu_1}\left[e^{\frac{z_2}{\mu_1}}\left(e^{\frac{-z_1}{\mu_1}} - e^{\frac{-z_2}{\mu_1}}\right) - e^{\frac{-z_2}{\mu_1}}\left(e^{\frac{z_2}{\mu_1}} - e^{\frac{z_1}{\mu_1}}\right)\right]$$

$$m_2 = \frac{\beta}{\mu_2}\left[e^{\frac{z_2}{\mu_2}}\left(e^{\frac{-z_2}{\mu_2}} - e^{\frac{-z_3}{\mu_2}}\right) - e^{\frac{-z_2}{\mu_2}}\left(e^{\frac{z_3}{\mu_2}} - e^{\frac{z_2}{\mu_2}}\right)\right]$$

$$m_4 = \frac{\gamma}{\mu_3}\left[e^{\frac{z_3}{\mu_3}}\left(e^{\frac{-z_3}{\mu_3}} - e^{\frac{-z_4}{\mu_3}}\right) - e^{\frac{-z_3}{\mu_3}}\left(e^{\frac{z_4}{\mu_3}} - e^{\frac{z_3}{\mu_3}}\right)\right]$$

$$m_5 = \frac{\beta}{\mu_2}\left[e^{\frac{z_3}{\mu_2}}\left(e^{\frac{-z_2}{\mu_2}} - e^{\frac{-z_3}{\mu_2}}\right) - e^{\frac{-z_3}{\mu_2}}\left(e^{\frac{z_3}{\mu_2}} - e^{\frac{z_2}{\mu_2}}\right)\right]$$

Acknowledgements The authors are grateful to the Director, Maharaja Agrasen Institute of Technology, Delhi, for providing the facilities to carry out this research work. One of the authors (Amit Das) would like to thank UGC, Government of India, for financial assistance to carry out this research work.

Funding Not Applicable.

Data availability Not Applicable.

References

1. Sood, H., Srivastava, V.M., Singh, G.: Advanced MOSFET technologies for next generation communication systems - perspective and challenges: A review. J. Eng. Sci. Technol. Rev. **11**(3), 180–195 (2018)
2. Das, A., Kanaujia, B.K., Nath, V., Rewari, S., Gupta, R.S.: Impact of reverse gate oxide stacking on gate all around tunnel fet for high frequency analog and RF applications. In: 2020 IEEE 17th India Council International Conference INDICON 2020, pp. 1–6 (2020)
3. Goyal, P., Srivastava, G., Rewari, S., Gupta, R.S.: Controlling ambipolarity and rising ion in TFETs for enhanced reliability: a review. In: 2020 5th IEEE International Conference. Recent Adv. Innov. Eng. ICRAIE 2020 - Proceeding, pp. 1–6 (2020)
4. Sharma, S., Rewari, S., Nath, V., Deswal, S. S., Gupta, R. S.: Schottky barrier double surrounding gate MOSFET for high-frequency implementation. In: 2020 5th IEEE International Conference on. Recent Advance Innovation Engineering, pp. 5–8 (2020). https://doi.org/10.1109/ICRAIE51050.2020.9358359.
5. Sharma, S., Goel, A., Rewari, S., Vandana, N., Gupta, R.S.: Enhanced analog performance and high-frequency applications of dielectric engineered high-K Schottky nanowire FET. In: Silicon (2022)
6. Chakraborty, A., Sarkar, A.: Analytical modeling and sensitivity analysis of dielectric-modulated junctionless gate stack surrounding gate MOSFET (JLGSSRG) for application as biosensor. J. Comput. Electron. **16**(3), 556–567 (2017)
7. Das, A., Rewari, S., Kanaujia, B.K., Gupta, R.S.: Recent technological advancement in surrounding gate MOSFET for biosensing applications - a synoptic study. SILICON **14**(10), 5133–5143 (2022)
8. Gautam, R., Saxena, M., Gupta, R.S., Gupta, M.: Numerical model of gate-all-around MOSFET with vacuum gate dielectric for biomolecule detection. IEEE Electron Dev. Lett. **33**(12), 1756–1758 (2012). https://doi.org/10.1109/LED.2012.2216247
9. Pratap, Y., Kumar, M., Kabra, S., Haldar, S., Gupta, R.S., Gupta, M.: Analytical modeling of gate-all-around junctionless transistor based biosensors for detection of neutral biomolecule species. J. Comput. Electron. **17**(1), 288–296 (2018)
10. Goel, A., Rewari, S., Verma, S., Gupta, R.S.: Dielectric modulated triple metal gate all around MOSFET (TMGAA)for DNA bio-molecule detection. In: Proceedings of the International Conference on 2018 IEEE Electron Device Kolkata Conference EDKCON 2018, pp. 337–340 (2018)
11. Goel, A., Rewari, S., Verma, S., Deswal, S.S., Gupta, R.S.: Dielectric modulated junctionless biotube FET (DM-JL-BT-FET) bio-sensor. IEEE Sens. J. **21**(15), 16731–16743 (2021)
12. Padmanaban, B., Sathiyamoorthy, S., Ramesh, R.: Modeling and simulation of gate engineered gate all-around MOSFET for bio-molecule detection. Indian J. Sci. Technol. **9**(43), 1–6 (2016)
13. Das, R., Chanda, M., Sarkar, C.K.: Analytical modeling of charge plasma-based optimized nanogap embedded surrounding gate MOSFET for label-free biosensing. IEEE Trans. Electron Devices **65**(12), 5487–5493 (2018)
14. Li, C., Liu, F., Han, R., Zhuang, Y.: A vertically stacked nanosheet gate-all-around FET for biosensing application. IEEE Access **9**, 63602–63610 (2021)
15. Jana, G., Sen, D., Debnath, P., Chanda, M.: Power and delay analysis of dielectric modulated dual cavity Junctionless double gate field effect transistor based label-free biosensor. Comput. Electr. Eng. **99**, 107828 (2022)
16. Rahman, E., Shadman, A., Khosru, Q.D.M.: Effect of biomolecule position and fill in factor on sensitivity of a dielectric modulated double gate junctionless MOSFET biosensor. Sens. Bio-Sensing Res. **13**, 49–54 (2017)
17. Wu, H., Si, M., Dong, L., Gu, J., Zhang, J., Ye, P.D.: Germanium nMOSFETs with recessed channel and S/D: contact, scalability, interface, and drain current exceeding 1 A/mm. IEEE Trans. Electron Dev. **62**(5), 1419–1426 (2015)
18. Brunco, D.P., et al.: Germanium MOSFET devices: advances in materials understanding, process development, and electrical performance. J. Electrochem. Soc. **155**(7), H552 (2008)
19. Saha, R., Hirpara, Y., Hoque, S.: Sensitivity analysis on dielectric modulated ge-source DMDG TFET based label-free biosensor. IEEE Trans. Nanotechnol. **20**, 552–560 (2021)
20. Han, K., Long, S., Deng, Z., Zhang, Y., Li, J.: A novel germanium-around-source gate-all-around tunnelling field-effect transistor for low-power applications. Micromachines **11**(2), 1–11 (2020)

21. Bitnar, B.: Silicon, germanium and silicon/germanium photocells for thermophotovoltaics applications. Semicond. Sci. Technol. **18**(5), 221–227 (2003). https://doi.org/10.1088/0268-1242/18/5/312
22. Gamble, H., et al.: Germanium processing. Eng. Mater. (2011). https://doi.org/10.1007/978-3-642-15868-1_1
23. Nguyen, T.H., Lee, M.S.: A review on germanium resources and its extraction by hydrometallurgical method. Miner. Process. Extr. Metall. Rev. **42**(6), 406–426 (2021)
24. Lee, C.S., Kyu Kim, S., Kim, M.: Ion-sensitive field-effect transistor for biological sensing. Sensors **9**(9), 7111–7131 (2009)
25. Yojo, L.S., Rangel, R.C., Sasaki, K.R.A., Martino, J.A.: Study of BE SOI MOSFET reconfigurable transistor for biosensing application. ECS J. Solid State Sci. Technol. **10**(2), 027004 (2021)
26. Park, J., Hiep, H., Woubit, A., Kim, M.: Applications of field-effect transistor (FET) -type biosensors. Appl. Sci. Converg. Technol. **23**(2), 61–71 (2014)
27. Khakshoor, A., Belhassen, J., Bendayan, M., Karsenty, A.: Doping modulation of self-induced electric field (SIEF) in asymmetric GaAs/GaAlAs/GaAs quantum wells. Results Phys. (2022). https://doi.org/10.1016/j.rinp.2021.105093
28. Kang, S.-M., Yusuf, L.: CMOS Digital Integrated Circuits, 3rd ed. (2003)
29. Hong, S., et al.: Improved CO gas detection of Si MOSFET gas sensor with catalytic Pt decoration and pre-bias effect. Sensors Actuators B Chem. **300**, 127040 (2019)
30. Ayadi, Y., et al.: Novel concept of gas sensitivity characterization of materials suited for implementation in FET-based gas sensors. Nanoscale Res. Lett. (2016). https://doi.org/10.1186/s11671-016-1687-z
31. Gautam, R., Saxena, M., Gupta, R.S., Gupta, M.: Gate-all-around nanowire MOSFET with catalytic metal gate for gas sensing applications. IEEE Trans. Nanotechnol. **12**(6), 939–944 (2013)
32. Pradhan, K.P., Kumar, M.R., Mohapatra, S.K., Sahu, P.K.: Analytical modeling of threshold voltage for cylindrical gate all around (CGAA) MOSFET using center potential. Ain Shams Eng. J. **6**(4), 1171–1177 (2015)
33. ATLAS User's Manual: Device Simulation Software, Santa Clara, CA (2018)
34. Zhang, Y., Han, K., Li, J.: A simulation study of a gate-all-around nanowire transistor with a core-insulator. Micromachines **11**(2), 1–12 (2020)
35. Ganesh, A., Goel, K., Mayall, J.S., Rewari, S.: Subthreshold analytical model of asymmetric gate stack triple metal gate all around MOSFET (AGSTMGAAFET) for improved analog applications. SILICON **14**, 4063–4073 (2021)
36. Ye, S., Yamabe, K., Endoh, T.: Ultimate vertical gate-all-around metal–oxide–semiconductor field-effect transistor and its three-dimensional integrated circuits. Mater. Sci. Semicond. Process. **134**, 106046 (2021)
37. Choi, S.J., Il Moon, D., Kim, S., Duarte, J.P., Choi, Y.K.: Sensitivity of threshold voltage to nanowire width variation in junctionless transistors. IEEE Electron Dev. Lett. **32**(2), 125–127 (2011)
38. Kumar, M., Haldar, S., Gupta, M., Gupta, R.S.: Ambipolarity reduction in DMG asymmetric vacuum dielectric Schottky Barrier GAA MOSFET to improve hot carrier reliability. Superlatt. Microstruct. **111**, 10–22 (2017)
39. Sze, S.M.: VLSI Technology. McGraw Hill education, New York (2017)
40. Goley, P.S., Hudait, M.K.: Germanium based field-effect transistors: Challenges and opportunities. Materials (Basel) **7**(3), 2301–2339 (2014)
41. Parihar, M.S., Kranti, A.: Enhanced sensitivity of double gate junctionless transistor architecture for biosensing applications. Nanotechnology **26**(14), 145201 (2015). https://doi.org/10.1088/0957-4484/26/14/145201
42. Preethi, S., Venkatesh, M., Karthigai Pandian, M., Lakshmi Priya, G.: Analytical modeling and simulation of gate-all-around junctionless mosfet for biosensing applications. Silicon **13**(10), 3755–3764 (2021). https://doi.org/10.1007/s12633-021-01301-2
43. Hafiz, S.A., Ehteshamuddin, I.M., Loan, S.A.: Dielectrically modulated source-engineered charge-plasma-based schottky-FET as a label-free biosensor. IEEE Trans. Electron Devices **66**(4), 1905–1910 (2019)
44. Kumari, M., Singh, N.K., Sahoo, M., Rahaman, H.: Work function optimization for enhancement of sensitivity of dual-material (DM), double-gate (DG), junctionless MOSFET-based biosensor. Appl. Phys. A Mater. Sci. Process. **127**(2), 1–8 (2021)
45. Jang, D.-Y., Kim, Y.-P., Kim, H.-S., KoPark, S.-H., Choi, S.-Y., Choi, Y.-K.: Sublithographic vertical gold nanogap for label-free electrical detection of protein-ligand binding. J. Vac. Sci. Technol. B Microelectron. Nanom. Struct. **25**(2), 443 (2007)
46. Parihar, M.S., Ghosh, D., Armstrong, G.A., Yu, R., Razavi, P., Kranti, A.: Bipolar effects in unipolar junctionless transistors. Appl. Phys. Lett. (2012). https://doi.org/10.1063/1.4748909
47. Kim, S., Ahn, J.H., Park, T.J., Lee, S.Y., Choi, Y.K.: A biomolecular detection method based on charge pumping in a nanogap embedded field-effect-transistor biosensor. Appl. Phys. Lett. **94**(24), 1–4 (2009). https://doi.org/10.1063/1.3148340
48. Busse, S., Scheumann, V., Menges, B., Mittler, S.: Sensitivity studies for specific binding reactions using the biotin/streptavidin system by evanescent optical methods. Biosens. Bioelectron. **17**(8), 704–710 (2002)
49. Lodhi, A., Rajan, C., Kumar, A., Dip, B., Samajdar, P., Soni, D.: Sensitivity and sensing speed analysis of extended nano - cavity and source over electrode in Si / SiGe based TFET biosensor. Appl. Phys. A **126**(11), 1–8 (2020)
50. Kim, C.H., Jung, C., Park, H.G., Choi, Y.K.: Novel dielectric-modulated field-effect transistor for label-free DNA detection. Biochip J. **2**(2), 127–134 (2009)
51. Narang, R., Saxena, M., Gupta, M.: Modeling and simulation investigation of sensitivity of symmetric split gate junctionless fet for biosensing application. IEEE Sens. J. **17**(15), 4853–4861 (2017)
52. Narang, R., Saxena, M., Gupta, M.: Modeling of gate underlap junctionless double gate MOSFET as bio-sensor. Mater. Sci. Semicond. Process. **71**, 240–251 (2017)
53. Jung, H.: Analysis of subthreshold swing in symmetric junctionless double gate MOSFET using high-k gate oxides. Int. J. Electr. Electron. Eng. Telecommun. **8**(6), 334–339 (2019)
54. Wei, J., Lei, J., Tang, X., Li, B., Liu, S., Chen, K.J.: Channel-to-channel coupling in normally-off GaN double-channel MOS-HEMT. IEEE Electron Device Lett. **39**(1), 59–62 (2018). https://doi.org/10.1109/LED.2017.2771354
55. Abarca-Cabrera, L., Fraga-García, P., Berensmeier, S.: Bio-nano interactions: binding proteins, polysaccharides, lipids and nucleic acids onto magnetic nanoparticles. Biomater. Res. **25**(1), 1–18 (2021)
56. Chiu, C.Y., Ruan, L., Huang, Y.: Biomolecular specificity controlled nanomaterial synthesis. Chem. Soc. Rev. **42**(7), 2512–2527 (2013)
57. Jain, S. K., Joshi, A. M.: Dielectric-modulated double gate bilayer electrode organic thin film transistor-based biosensor for label-free detection: simulation study and sensitivity analysis. pp. 1–16 (2022)
58. Chiang, T.-K., Liou, J.: An analytical subthreshold current/swing model for junctionless cylindrical nanowire FETs (JLCNFETs). Facta Univ. - Ser. Electron. Eng. **26**(3), 157–173 (2013). https://doi.org/10.2298/fuee1303157c
59. Sharma, D., Vishvakarma, S.K.: Precise analytical model for short channel Cylindrical Gate (CylG) Gate-All-Around (GAA) MOSFET. Solid State Electron. **86**, 68–74 (2013)

Nature Environment and Pollution Technology
An International Quarterly Scientific Journal
p-ISSN: 0972-6268
(Print copies up to 2016)
e-ISSN: 2395-3454
Vol. 22 No. 1 pp. 159-168 2023
Original Research Paper
https://doi.org/10.46488/NEPT.2023.v22i01.014
Open Access Journal

Optimization of Biodiesel Parameters Using Response Surface Methodology and Production of Biodiesel

Y. K. Singh*
*Department of Biotechnology, Delhi Technological University, New Delhi-110042, India
†Corresponding authors: Y. K. Singh; ykumars@yahoo.com

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 31-05-2022
Revised: 20-07-2022
Accepted: 22-07-2022

Key Words:
Biodiesel
Response surface methodology
CCD
Methanol/oil ratio
KOH catalyst

ABSTRACT

The requirement for a renewable and environmentally gracious alternative resource of energy has grown in recent years as a result of increased knowledge of the negative impacts of petroleum-based fuels on the environment and the regular rise in crude oil prices. Biodiesel has been proven to be the ideal replacement for diesel because of its unique qualities, such as a huge decrease in greenhouse gas emissions, nonparticulate matter pollutants, non-sulfur emissions, less toxicity, and degradability. This article examines the pre-treatment stage, the physiological and chemical features of WCO, transesterification, esterification, and the manufacturing of biofuel from waste-cooked oil using several techniques and catalyst types. The elements that influence the stated process parameters are investigated, with a particular focus on the methanol to oil ratio (molar ratio), time of reaction, the temperature of the reaction, catalyst percentage, and yield of biodiesel. After the production of biodiesel, we can optimize the process parameters, for example, methanol to oil ratio, the temperature of the reaction, duration of reaction, and catalyst percentage, and also optimize the yield of biofuel generation with the CCD design of the Response surface methodology (RSM) algorithm using Design Expert software.

INTRODUCTION

Waste cooking oil refers to the production of oil from different frying activities, such as oil used in restaurants for frying purposes. Two categories of second-hand cooking are formed and used: primary and secondary-hand cooking oils. Primarily used cooking oil prefers to squander oil from clean vegetable oils and is usually generated by restaurants and shops. While second- or secondary-used cooking oil is waste oil derived from first- or primary-used cooking oil, it is typically generated by street vendors. These days, the oil is generally just thrown away, lacking any treatment. Then it will infect the whole environment when we just pay no attention to it. One single alternative to treating this second or secondary-used cooking oil is by conversion into biodiesel. That substitute will not only have environmental advantages but also be economical (Kawentar & Budiman 2013, Uddin et al. 2013). In today's world, power/energy is a crucial dynamic component for socioeconomic advancement. It has an impact on all aspects of human endeavors, for example, crop production, education, and transportation, amongst others. Petro-linked fuels are the most common type of fuel used in the transportation industry in practically all developed countries. Though climate change and rising pumping costs have shifted research focus to sustainable energy resources (Samuel et al. 2013, Phan & Phan 2008). The search for green energy sources is a topical subject that is gaining widespread communal and political attention owing to its abridged greenhouse gas emissions, biodegradability, sustainability, and spirited nature in comparison to fossil fuels and food supplies. Transesterification produces biodiesel from vegetable oil (waste cooking oil). According to the American Society for Testing and Materials (ASTM), biodiesel is distinct as a single alkyl ester of a lengthy chain of fatty acids resulting from sustainable feedstocks. The main disadvantage is the cost, which is significantly greater than that of oil-derivative diesel. The increased price of virgin or fresh oils, which might account for up to 75% of the overall built-up price, has resulted in biodiesel manufacturing prices being around 1.5 times more than petro-diesel. Waste cooking oils are 2 to 3 times less expensive than new virgin oils.

As a result, the total built-up price of biodiesel can be considerably reduced (Samuel et al. 2013). Though there are several successful reports on biodiesel generation from used cooking oil, it is not highly explored owing to

ORCID details of the authors:

Y. K. Singh
https://orcid.org/0000-0002-6225-495X

the difficulty in transesterification as a result of high free fatty acid constituents. In recent work, we report the direct-scale manufacture of biofuel from waste cooking oil with a free fatty acid (FFA) content in the range of 4 to 5%. The generation is achieved in a single stage without any preceding acid treatment. That's why the utilization of used oil for fatty acid methyl ester (FAME) production or formation is highly suggested (Unni et al. 2013).

REACTIONS OF WASTE OIL AND BIODIESEL

Transesterification

As indicated in Fig. 1, the triglyceride constituent of oil combines with the methanol in the presence of sodium hydroxide or another catalyst to produce esters and glycerol. In common, when using vegetable oil and animal fat as an initial material, there are three types of transesterification systems: homogeneous, heterogeneous, and enzymatic, depending on the catalyst used. Because methanol is more efficient, UVO is usually reacted with alcohol. Ethyl alcohol is used for animal fats, but ethyl alcohol and isopropyl alcohol can be used as well. Transesterification is supposed to be influenced by a variety of factors, such as temperature for reaction, pressure, time of reaction, agitation rate, type of alcohol (whether ethanol or methanol is used) and molar ratio, kind and concentrations of catalysts used, and dampness and FFA concentration in the feedstock oil (Sarno & Iuliano 2019, Rizwanul Fattah et al. 2020). The physical and chemical qualities of the feedstock oil determine the best values for these parameters to achieve higher conversion. Today, the majority of biodiesel is made from edible vegetable oils that have been transesterified using a homogenous alkali catalyst. Homogeneous catalysts, which might be liquid or gaseous, are soluble during the process. Acidic and alkaline are the two types of them. For esterification, acidic catalysts such as H_2SO_4 are commonly employed, while transesterification uses alkaline catalysts, for example, NaOH and KOH (Sarno & Iuliano 2019). Homogeneous catalysts have the following advantages: (i) the ability to catalyze reactions at lower reaction temperatures and air pressures; (ii) the ability to achieve a higher level of conversion in a shorter period of time; and (iii) availability and cost. This method produces a high-quality artifact with a quick turnaround time. Only refined vegetable oil with a low level of 0.5 wt. percent or less is permitted. Free fatty acid or an acid value of not greater than 1 mg $KOH.g^{-1}$ can be used effectively with an alkaline homogeneous catalyst. Furthermore, after the reaction is completed, the separation of these catalysts necessitates washing biodiesel through water, which may result in the slaughter of fatty acid alkyl (methyl or ethyl) esters, energy utilization, and the generation of huge amounts of dissipated water. As a catalyst is not easy to recover and catalyst can induce reactor deterioration, this raises the overall cost of biodiesel production. To avoid soap generation (due to alkaline catalyst use) and low product yields, the triglyceride and alcohol (methanol or ethanol) must be anhydrous, and the raw material must have a low free fatty acid (FFA) concentration (Sarno & Iuliano 2019, Rizwanul Fattah et al. 2020).

Esterification

Because FFAs can cause deposits and engine damage, most biodiesel requirements have a maximum FFA level. As illustrated in Fig. 2, esterification can be utilized to switch free fatty acids to biodiesel while also reducing FFAs. Fatty acids interact using alcohol in the absence of a catalyst to form fatty acid alkyl (methyl or ethyl) ester in this reaction (Biodiesel). The goal of the esterification process is to

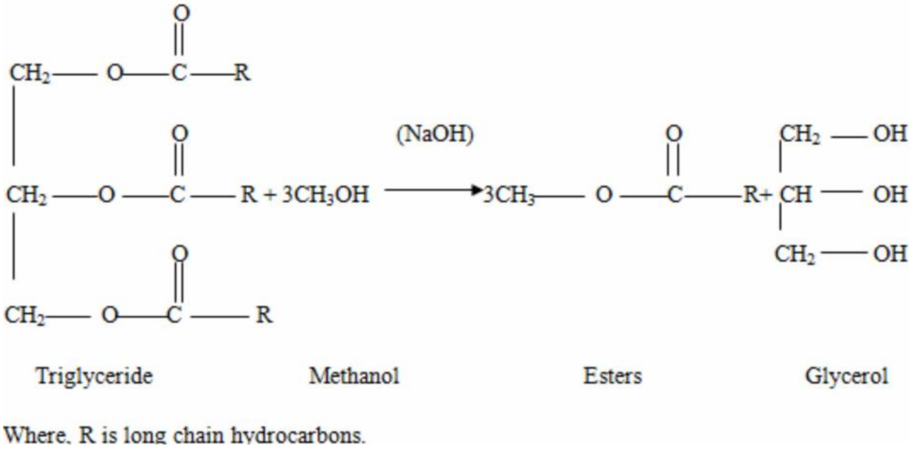


Fig. 1: A schematic illustration of the transesterification reaction (Sarno & Iuliano 2019).

reduce WCO's acidity. As conventional acid catalysts in the esterification process, sulphuric acid (H_2SO_4), hydrochloric acid (HCl), butyl-methyl imidazolium hydrogen sulfate ($BMIMHSO_4$), and sulfonic acid are commonly used (Sarno & Iuliano 2019, Ghiaci et al. 2011). Titration of oil through ethanol and diethyl ether (1:1) alongside potassium hydroxide (KOH) via phenolphthalein as a marker determines the acid values of the oil. The acid value is equal to $56.1*CV.m^{-1}$, where V represents the quantity of KOH (mL), C represents the concentration of potassium hydroxide (KOH) in M, and m represents the heaviness of the oil sample in g. For official techniques, AOCS Cd 3d-63 and ASTMD-664 were followed in this titration. The catalyst is chosen based on acidity. The feedstock can be transesterified without any pretreatment if the FFA content is less than 1%. According to research findings, maximum conversion is achieved at 2% v/v H_2SO_4. Because the reaction is reversible, equilibrium is the greatest stumbling block to its completion. The FFA can be reduced by reducing water by preheating in an oven. The Alcohol to Methanol Ratio, the catalyst and its amount used, and the process temperature are the primary factors determining the esterification reaction (Sarno & Iuliano 2019, Ghiaci et al. 2011).

MATERIALS AND METHODS

If the free fatty acid content in oil exceeds 5% of the feedstock, then a pretreatment process is required before reacting with the alkaline base catalyst (Ribeiro et al. 2011).

Materials

The WCO used in the making of biodiesel was collected from the local street shops and FFA was measured with two different oil samples collected from different shops (0.7% and 0.2%). For example, methanol with 99% purity, potassium hydroxide (KOH) with 90% purity, and for some quality checks for oil and biodiesel, phenol red indicator LR grade, isopropyl alcohol with 99% purity, bromophenol blue, hydrochloric acid 0.01N LR grade for soap content, and 1% phenolphthalein indicator were used for excess catalyst in the process. Alcohol (methanol) is used for the transesterification process, and the KOH base catalyst is used as the base catalyst (Table 1).

Synthesis of Methyl Esters

The synthesis or production of biodiesel initially requires pretreatment if the FFA content is high. First, the oil is heated

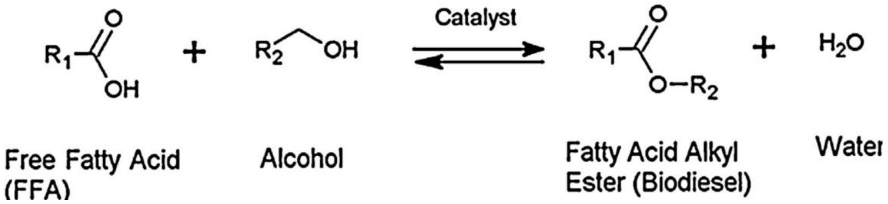


Fig. 2: Schematic illustration of the esterification reaction (Sarno & Iuliano 2019).

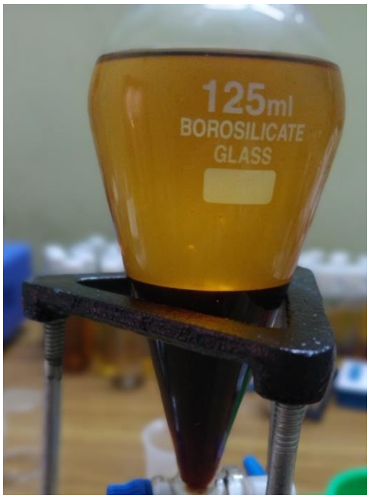


Fig. 3a: Shows two layers of upper layer of biodiesel and the bottom layer of glycerol.

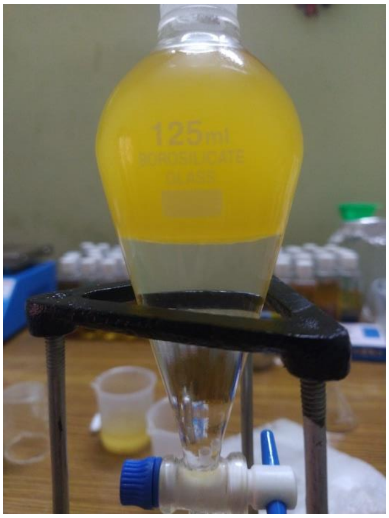


Fig. 3b: Shows biodiesel after washing.

Table 1: Quality analysis of oil and biodiesel.

Quality parameters	Analysis result
Acid value of oil	9mg.KOH^{-1}.g^{-1}
Free fatty acid content in oil	4.5%
Soap content(ppm)	285ppm

to a temperature of 100°C to eliminate any moisture content available in the oil, then the heated oil is cooled down. Again, heat the oil to a different temperature range, from 40 to 75°C, the process temperatures are given in the process Table 4. After heated oil reaches the desired temperature, KOH (normally 0.3 to 1 percent of oil according to FFA content of oil, the catalyst % is taken) with methanol is mixed (ratio of methanol to oil is calculated as per desired data given in Table 4) and added to the process for transesterification reaction with continuous stirring of the process mixture at a desired temperature. The stirring was also continuous for about 45 min to 120 min (all data in Table 4 show the minimum range and maximum range of different parameters). Thereafter, two layers were produced; the upper layer is of biodiesel, and the lower layer is of glycerol, as shown in Fig. 3a and 3b. Then, the mixture was allowed 24 h to properly settle so that all the biodiesel was properly separated from the glycerol. After 24 h, the glycerol was separated from the biodiesel and further processing was done (washing and testing). Washing of biodiesel is done through hot water with 3 to 5 washes with water and then drying of the biodiesel with heating at a temperature of above 100°C for 1 h.

Analysis of Process (Biodiesel)

After the synthesis of biodiesel and before washing, the quality check for biodiesel is done. By using the 3/27 methanol test (Heisner 2020), you can check whether the oil is properly reacted or not. In this test, 3mL of prepared biodiesel was taken and added to 27 mL of methanol, then mixed vigorously in the vial for 5 to 10 seconds. If there is any oil or unreacted oil or fall seen at the bottom of the vial, it means the oil is not properly reacted. If there is no fall seen at the bottom, it means the oil is properly reacted. The 3/27 methanol test was performed both before and after washing the biodiesel (see Fig. 4a and 4b).

Excess Catalyst in Biodiesel

The high level of catalyst content in biodiesel leads to the problem of soap formation and increases the soap ppm level in biodiesel. By eliminating or removing excess catalyst (KOH) in prepared biodiesel, take 100 mL of isopropyl alcohol into a 250 mL beaker and then add about 12 mL of biodiesel. Mix properly. Add 5 drops of 1% phenolpthalein indicator to the beaker. If the solution in the beaker stays clear, it means there is no extra catalyst in the biodiesel. If the solution turns magenta after the addition of the indicator, it means there is some extra catalyst present in the biodiesel. The biodiesel requires some treatment to neutralize it, so take 0.01 N HCL and put the HCL drop-wise in the beaker slowly until the solution color changes from magenta to clear solution. After the excess catalyst removal process, the next step is the soap content test for the biodiesel.

Soap Content Test for Biodiesel

The high level of soap content in biodiesel results in the clogging of filters and engines of automobiles. The soap

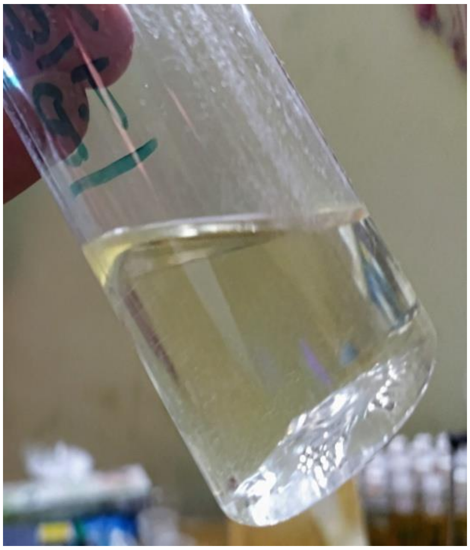


Fig. 4a: Conversion complete (no fall seen).

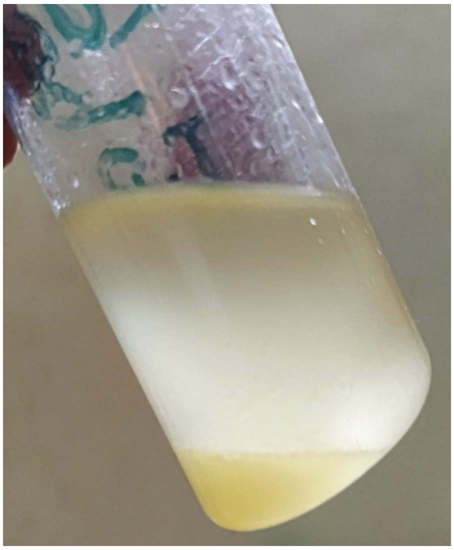


Fig. 4b: Incomplete conversion (fall seen).

Table 2: Analyzing the quality of biodiesel based on the soap content chart.

Soap Content	Fuel Quality
at or below 41 ppm (NaOH) or 66 ppm (KOH)	Within ASTM standards
Above ASTM Standards but Below 200 ppm	Should not pose any threat to a fuel filter or engine
200-300 ppm	maximum soap content which should be allowed in fuel
300-400 ppm	May clog fuel filters, not recommended, wash more
400-500 ppm	High soap content, not recommended, wash more
Above 500 ppm	Can possibly leave ash in your engine and cause long-term damage, not recommended, wash more

content of fuels should be according to the ASTM standard as shown in Table 2. The testing of soap content for biodiesel requires 0.01 N HCL, bromophenol blue, and isopropyl alcohol. Take 100 mL of isopropyl alcohol into a 250 mL beaker, then add about 12 mL of biodiesel into the beaker and mix them. Add 15 to 20 drops of bromophenol blue into the beaker until the solution turns a dark blue color. After that, titrate the solution with 0.01 N HCL. Note that the mL of HCL is required to change the color of the solution from a dark blue color to a yellowish color. Soap content should be checked before and after washing and drying. In the case of the KOH catalyst, the 320 value factor is taken, and in the case of the NaOH catalyst, the 304 value factor is used. The ppm is calculated by multiplying the catalyst factor by the amount of HCL required to get the PPM of the biodiesel sample.

RESULTS AND DISCUSSION

Experimental Design and Parameters Optimization

Box-Behnken design (BBD) and central composite design are the two major experimental designs utilized for response surface optimization (CCD). In this study, we used design expert software to apply the CCD design of the response surface methodology. In the response surface approach, two essential models are typically used, namely the first-degree and second-degree models (Kumar Ghosh & Mittal 2021). When the response can be well explained by a linear function of independent variables, a first-degree model is used. However, when the system has curvature, a second-degree model is used, and a high-degree polynomial is used. In all of these models, there is a correlation between independent variables like time of reaction, temperature, molar ratio, catalyst weight percent, and the resulting variable (yield percent). Table 3 shows the practical amounts and ranges of several independent variables used in the production of biodiesel. In this work, 30 experimental runs were done and consisted of 16 factorial, 8 axial, and 6 center points. The 2nd-degree model is applied in this article, which suggests 30 runs. We already discussed how this system shows curvature.

Experimental design for the production of biodiesel: the coded values of different independent variables are specified in Table 4. The methanol to oil (molar ratio) and catalyst percent are represented by the coded variables x_1 and x_2. The x_3 and x_4 denote the temperature of the reaction and time, respectively (Kumar Ghosh & Mittal 2021).

Quadratic equation Eq. (1) states the performance of the system. For multiple regression data analysis, a statistical program was utilized. Calculating the regression equation and studying the response of 3D surface plots and contour plots provides the optimum value of selected variables.

$$\boldsymbol{Y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{i=1}^{k} \beta_{jj} x_j^2 + \sum\sum_{i<j}^{k} \beta_{ij} x_i x_j + \varepsilon} \ldots(1)$$

Whereas Y denotes the biodiesel yield percentage, and *xi*, and *xj* represent actual independent variables in the appearance of encoding; β_0, β_j, β_{jj}, *and* β_{ij} expressed as intercept,linear,quadratic, and interaction constant coefficients also ε denotes a random error.

Regression Equation for Yield of Biodiesel

The essential parameters that affect the resultant (biodiesel yield) are; the molar ratio (methanol to oil ratio (*x1*)), catalyst percentage (*x2*), the temperature of reaction (*x3*), time of reaction (*x4*) (Kumar Ghosh & Mittal 2021). Experimental

Table 3: Levels of independent variables for the experimental design.

Factor	Name	Units	Minimum	Maximum	Mean
A	Methanol/oil ratio (x_1)	Mol.mol^{-1}	1.0000	13.00	7.00
B	KOH catalyst (x_2)	%	-0.0500	1.75	0.8500
C	Temperature (x_3)	°C	42.50	72.50	57.50
D	Time (x_4)	Min	7.50	157.50	82.50

Table 4: CCD design for biodiesel production.

Runs	Independent variables				Points	Yield
	(x_1)	(x_2)	(x_3)	(x_4)		
1	7	0.85	72.5	82.5	Axial	98.4
2	4	0.4	65	120	Factorial	70.6
3	10	0.4	50	120	Factorial	96.8
4	7	0.85	57.5	82.5	Center	86.8
5	4	0.4	50	45	Factorial	32.6
6	7	0.85	57.5	82.5	Center	86.2
7	7	0.85	57.5	82.5	Center	98.7
8	7	0.85	57.5	82.5	Center	98.7
9	10	1.3	65	120	Factorial	80
10	4	1.3	65	45	Factorial	84.2
11	4	0.4	65	45	Factorial	38.7
12	4	1.3	50	45	Factorial	82.2
13	10	0.4	50	45	Factorial	82
14	4	1.3	65	120	Factorial	92.5
15	7	0.85	57.5	82.5	Center	98.7
16	7	0.85	57.5	157.5	Axial	98.5
17	10	0.4	65	120	Factorial	94.8
18	7	1.75	57.5	82.5	Axial	78
19	1	0.85	57.5	82.5	Axial	38.9
20	7	0.85	57.5	82.5	Center	98.7
21	4	0.4	50	120	Factorial	85.7
22	10	1.3	65	45	Factorial	92.3
23	4	1.3	50	120	Factorial	90.6
24	10	1.3	50	120	Factorial	86.5
25	7	0.05	57.5	82.5	Axial	41.3
26	7	0.85	42.5	82.5	Axial	95.2
27	10	1.3	50	45	Factorial	94.6
28	13	0.85	57.5	82.5	Axial	88.8
29	7	0.85	57.5	7.5	Axial	60.1
30	10	0.4	65	45	Factorial	92.3

runs are carried out to find the coordination between different parameters. The observed verdicts of the whole factorial central CCD were compared to the polynomial Eq. (1) using multiple regression analysis in Table 4. The equation of multiple regression for the yield of biodiesel formation as a function of many variables is shown in Eq. (2).

$$Y = - 1.73211 + 1.47311x_1 + 10.17824x_2 - 0.145439x_3 + 0.107262x_4 - 0.045366x_1^2 - 2.26447x_2^2 + 0.001667x_3^2 - 0.000111x_4^2 - 0.391099x_1x_2 + 0.000304x_1x_3 - 0.003777x_1x_4 - 0.006894x_2x_3 - 0.025322x_2x_4 - 0.000504x_3x_4 \quad \ldots(2)$$

The sign attached to the coefficient predicts the impact of the regression coefficients on the result or response. A negative sign indicates a combative effect, while a positive sign indicates a coadjuvant result..x_1,x_2,x_3,x_4 are four linear factors, and the interaction of x_1x_4 The coadjuvant effect is represented by the remaining quadratic intercepts.x_1^2, x_2^2, x_3^2, x_4^2 and relations of x_1x_2 , x_1x_3 , x_1x_4 , x_2x_3, x_2x_4, x_3x_4 predicts the combative effect. Confirmation of adequacy of the model is determined by the use of analysis of variance (ANOVA) (Kumar Ghosh & Mittal 2021) given in Table 5. Coefficient of determination R^2 is utilized to test whether the model is fit or not, the R^2 is calculated as 0.9363, suggesting that previously model states or explain 93.63% of the response variability, the transesterification experiment factors exhibited a total variation of 93.63(R^2) and adj.R^2

Table 5: ANOVA analysis of variance for a yield of biodiesel.

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	41.41	14	2.96	15.75	< 0.0001	significant
A-Methanol/oil ratio	9.65	1	9.65	51.39	< 0.0001	
B-KOH catalyst	5.94	1	5.94	31.62	< 0.0001	
C-Temperature	0.0011	1	0.0011	0.0060	0.9391	
D-Time	4.85	1	4.85	25.83	0.0001	
AB	4.46	1	4.46	23.75	0.0002	
AC	0.0007	1	0.0007	0.0040	0.9505	
AD	2.89	1	2.89	15.38	0.0014	
BC	0.0087	1	0.0087	0.0461	0.8329	
BD	2.92	1	2.92	15.55	0.0013	
CD	0.3215	1	0.3215	1.71	0.2105	
A^2	4.57	1	4.57	24.34	0.0002	
B^2	5.77	1	5.77	30.71	< 0.0001	
C^2	0.2411	1	0.2411	1.28	0.2750	
D^2	0.6692	1	0.6692	3.56	0.0786	
Residual	2.82	15	0.1878			
Lack of Fit	2.28	10	0.2281	2.12	0.2099	not significant
Pure Error	0.5369	5	0.1074			
Cor Total	44.23	29				

“R^2 = 93.63% and adj.R^2 = 87.68%

of 87.68%. This states to facilitate the model has the best association and makes an accurate prediction. In an analysis of variance, (ANOVA) of Table 5 shows the probability of p-value is not greater than 0.0001 which means the model is significant (Anbessa & Karthikeyan 2019).

Analysis of the Impact of Transesterification Parameters

Graphically, contour and 3D surface plots show the effects of transesterification parameters on the result (biodiesel yield).

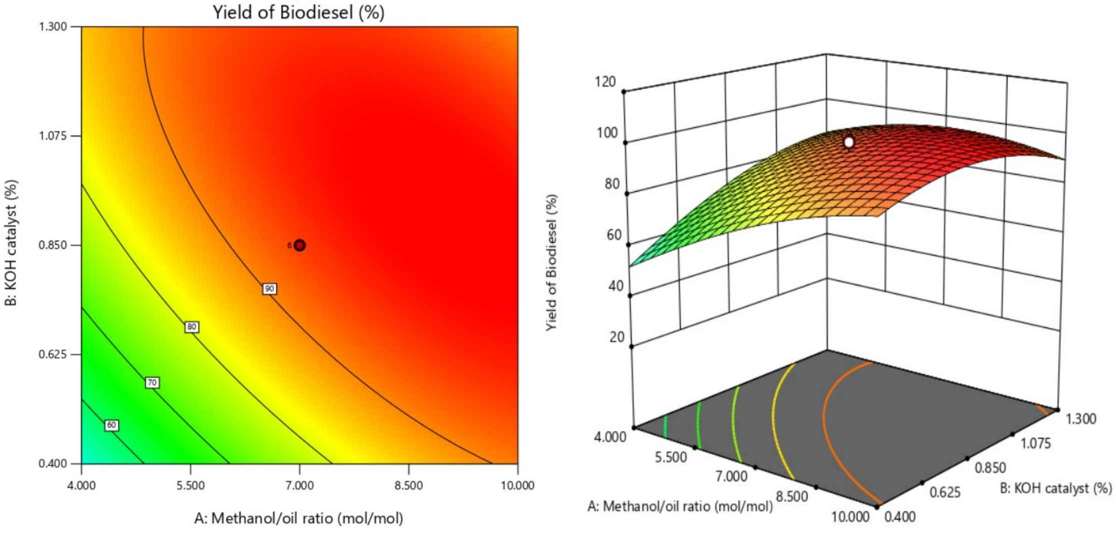


Fig. 5: (a) Represents a Contour plot and (b)shows a 3D surface plot showing the interaction of methanol/oil ratio and catalyst wt%.

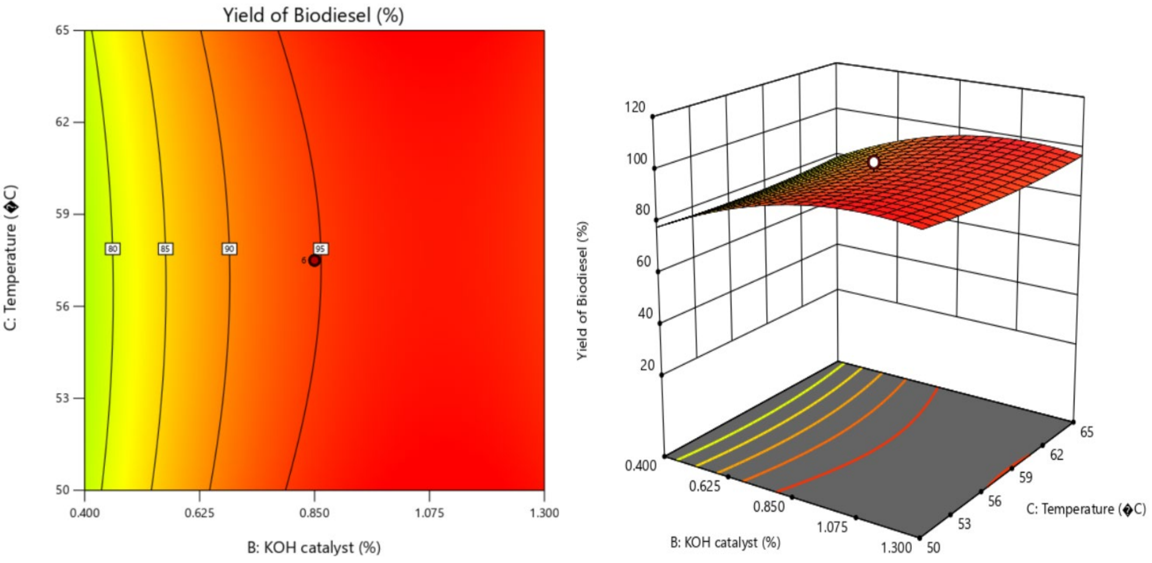


Fig. 6: (a) Represents contour plot (b) 3D surface plot shows the interaction of catalyst % and temperature.

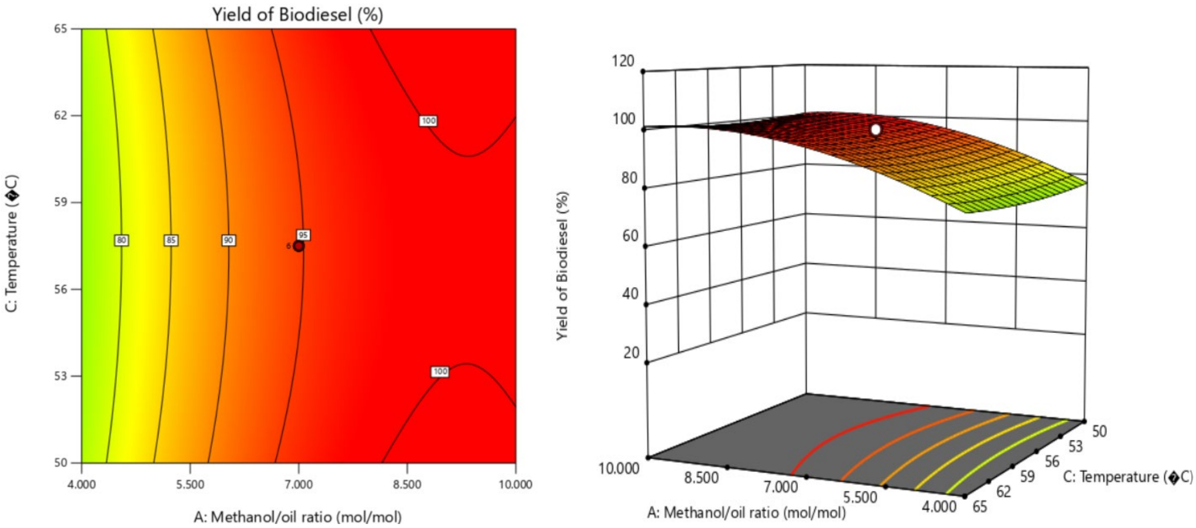


Fig. 7: (a) Represents contour plot (b) Shows 3D surface plot andrelations of methanol/oil ratio and temperature.

Fig. 5(a) depicts the relationship between the methanol/oil ratio and the catalyst percent, as well as the effect on yield. According to Fig. 5(b) of the 3D surface plot, as the molar ratio (methanol/oil ratio) increases, so does the yield of biodiesel, which ranges from 4:1 to 10:1. the optimal methanol/oil ratio is determined by optimization. The best optimum ratio that was achieved is a 10:1 methanol/oil ratio, and this gives a yield of 98.84% for biodiesel. By observing the data, it is found that increasing the methanol/oil ratio with catalyst gives an increment in biodiesel yield due to the higher number of active sites. However, too much catalyst percent results in excess emulsion (Hazmi et al. 2021). maximum yield is obtained at optimized conditions of methanol/oil ratio (10:1) and catalyst 1.3%, which gives 98.84 yields.

Likely, Fig. 6(a) and (b) indicate the effects of interactive factors such as KOH catalyst percentage and temperature of reaction on the resultant response. Fig. 7(a) and (b) show the response of correlated factors to the methanol/oil ratio and temperature of the reaction. A 3D surface plot represents the increase in yield of biodiesel as temperature increments from 50°C to 65°C. This increase in yield is because the speed of transesterification ranges increases as the temperature increments due to the enhancement of a

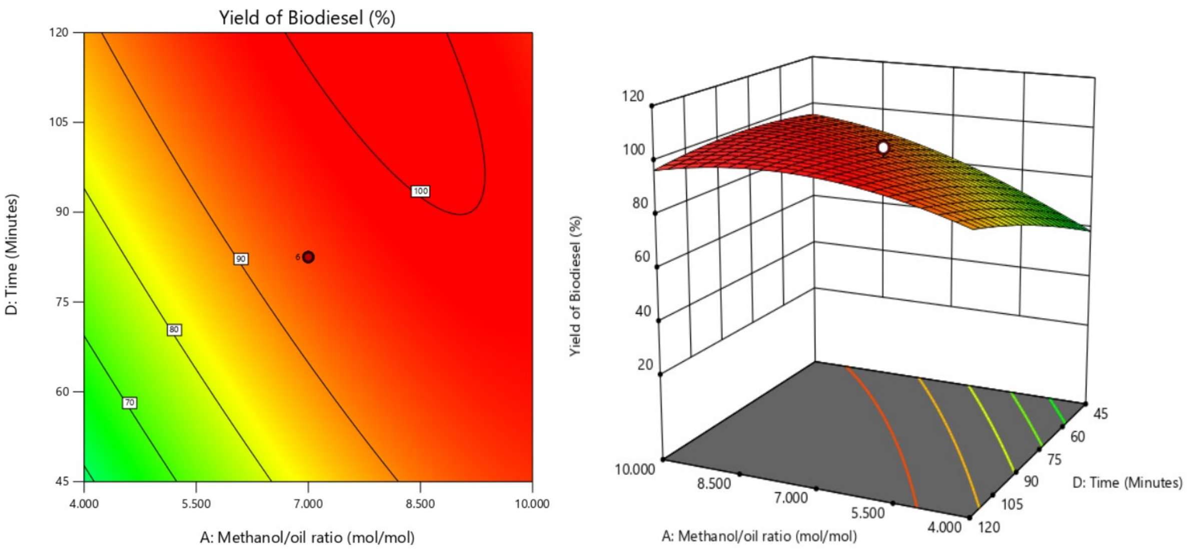


Fig. 8: (a) Represents contour plot (b) Shows 3D surface plot and relations of methanol/oil ratioand reaction time.

Table 6: Optimized result of the process.

Transesterification parameters	Optimum values
Yield of biodiesel	98.84%
Methanol/oil ratio	10:1
Catalyst%	1.3%
temperature	65°C
time	45 min

homogenous mixture (miscibility) when methanol and oil are mixed at high temperatures (Kumar Ghosh & Mittal 2021). The optimum temperature for the best yield is 65°C, which is optimized through RSM optimization with 1.3% catalyst loading for a higher yield.

Fig. 8(a) and (b) represent the interaction of molar ratio (methanol to oil ratio) and reaction time and its effect on the resultant (yield) of biodiesel. Higher ratios of methanol to oil lead to a more rapid conversion of biodiesel. Also, the time of reaction for the process depends on the nature of the catalyst (acid or base catalyst). Typically, the catalyst requires less significant time (1–2 h) for the conversion of biodiesel from oil. As the yield of biodiesel increases with reaction time, excess time of the reaction can lead to deteriorated yield and more glycerol production (Kumar Ghosh & Mittal 2021). After optimization, the optimum reaction time was 45 min for high conversion.

The optimized values are calculated from the regression equations. The different transesterification parameters are summarized. After studying the contour plot and 3D surface plots, we get optimum values for the highest yields of biodiesel production. The maximal yield of biodiesel is calculated to be 98.84% and was predicted using design expert software as the methanol/oil ratio =10:1 catalyst =1.3%, temperature =65°C, and time =45 min. We can conclude from the analysis of all contours and surfaced plots that the maximum yield of biodiesel obtained was 98.84%. The optimized results are given in Table 6.

CONCLUSION

The conversion of biodiesel from triglycerides is based on important parameters and the response surface methodology. The optimized results are obtained by solving the regression equation by using the CCD of the response surface methodology. The response surface methodology is a suitable method to optimize the best or highest level of yield. Thirty experimental runs were carried out for analysis using CCD-based RSM. Studying contours and 3D surface plots were utilized to find optimum results. Whereas we get 98.84% of the yield achieved at methanol/oil ratio (10:1), catalyst percentage (1.3%), temperature (65°C), and time (45 min). This study represents a better yield of biodiesel production and a long-term solution for environmental benefits.

REFERENCES

Anbessa, T.T. and Karthikeyan, S. 2019. Optimization and mathematical modeling of biodiesel production using homogenous catalyst from waste cooking oil. Int. J. Eng. Adv. Technol., 9(1): 1733-1739. https://doi.org/10.35940/ijeat.F9005.109119

Ghiaci, M., Aghabarari, B. and Gil, A. 2011. Production of biodiesel by esterification of natural fatty acids over modified organoclay catalysts. Fuel, 90(11): 3382-3389. https://doi.org/10.1016/j.fuel.2011.04.008

Hazmi, B., Rashid, U., Ibrahim, M.L., Nehdi, I.A., Azam, M. and Al-Resayes, S.I. 2021. Synthesis and characterization of bifunctional magnetic nano-catalyst from rice husk for production of biodiesel.

Environ. Technol. Innov., 21: 101296. https://doi.org/10.1016/j.eti.2020.101296

Heisner, B. 2020. Utilizing the 3 / 27 conversion test to measure the effects of temperature on the base-catalyzed transesterification of waste vegetable oils into fatty acid methyl esters. J. Autom. Technol., 15: 3-14.

Kawentar, W.A. and Budiman, A. 2013. Synthesis of biodiesel from second-used cooking oil. Energy Procedi., 32: 190-199. https://doi.org/10.1016/j.egypro.2013.05.025

Kumar Ghosh, U. and Mittal, V. 2021. Application of response surface methodology for optimization of biodiesel production from microalgae through nano catalytic transesteriication process. Fuel Process. Technol., 92(3): 407-413. https://doi.org/10.21203/rs.3.rs-771200/v1

Phan, A.N. and Phan, T.M. 2008. Biodiesel production from waste cooking oils. Fuel, 87(17-18): 3490-3496. https://doi.org/10.1016/j.fuel.2008.07.008

Ribeiro, A., Castro, F. and Carvalho, J. 2011. Influence of free fatty acid content in biodiesel production on non-edible oils. International Conference Waste Sol. Treat. Oppor., 12: 141.

Rizwanul Fattah, I.M., Ong, H.C., Mahlia, T.M.I., Mofijur, M., Silitonga, A.S., Ashrafur Rahman, S.M. and Ahmad, A. 2020. State of the art of catalysts for biodiesel production. Front. Energy Res., 8: 1-17. https://doi.org/10.3389/fenrg.2020.00101

Samuel, O.D., Waheed, M.A., Bolaji, B.O. and Dario, O.U. 2013. Production of biodiesel from Nigerian restaurant waste cooking oil using blender. Int. J. Renew. Energy Res., 3(4): 976-979. https://doi.org/10.20508/ijrer.35021

Sarno, M. and Iuliano, M. 2019. Biodiesel production from waste cooking oil. Green Process. Synth., 8(1): 828-836. https://doi.org/10.1515/gps-2019-0053

Uddin, M.R., Ferdous, K., Uddin, M.R., Khan, M. and Islam, M.A. 2013. Synthesis of biodiesel from waste cooking oil. Chem. Eng. Sci., 1(2): 22-26. https://doi.org/10.12691/ces-1-2-2

Unni, K.S., Yaakob, Z., Pudukudy, M., Mohammed, M. and Narayanan, B.N. 2013. Single step production of biodiesel from used cooking oil. Proceedings of 2013 International Renewable and Sustainable Energy Conference, IRSEC 2013, 461-464. https://doi.org/10.1109/IRSEC.2013.6529712

Wireless Networks
https://doi.org/10.1007/s11276-023-03292-x

ORIGINAL PAPER

Optimization of energy and delay on interval data based graph model of wireless sensor networks

Radhika Kavra[1] · Anjana Gupta[1] · Sangita Kansal[1]

Accepted: 17 February 2023

Abstract

Emerging applications of wireless sensor networks (WSNs) in various domain of real-life require establishment of such routing topology for wireless sensors which can balance energy consumption with delay minimization. Despite a plethora research on energy and delay optimization in this field, work on developing delay minimum and energy efficient connected routing topology on heterogenous bi-directional WSNs where each node is assigned with a power interval, remain untouched. Considering this problem, we have introduced 'Interval data based graph model (IDGM)' and 'Sorted-interval data based graph model (S-IDGM)' of WSNs to explicitly deal with nodes' power interval and proposed 'Energy and Delay Optimization (EDO)' algorithm to optimize S-IDGM such that the maximum topology delay, total topology delay and maximum node's power interval become minimum in polynomial time complexity. A new function is formulated to estimate topology delay based on link distance and link interference after showing dependency analysis between link distance and link interference on large number of WSNs towards achieving optimal solutions. Extensive simulation work, graphical and statistical *t*-test analysis have been carried out to show the performance of EDO algorithm in minimizing topology delay and nodes' power consumption, better than the existing algorithms from similar grounds. *t*-test analysis shows that the proposed EDO algorithm achieves optimal energy saving of nodes at 5% level of significance along with optimal minimization of max and total topology delay at 2% level of significance on S-IDGMs.

Keywords Wireless sensor networks (WSNs) · Algorithm · Energy-Delay Tradeoff · Minimum spanning tree (MST) · Path cover

1 Introduction

Wireless sensor network (WSN) is the most evolutionary technology now-a-days because of the facilitation of low-cost multi-functional wireless sensors that are small in size and communicate untethered using radio signals, contributing to large number of applications in various fields involving tasks like environmental monitoring, health and well-ness monitoring, power monitoring, inventory and location monitoring, military surveillance, tracking objects, animals, humans and vehicles etc [3, 27, 44]. Each sensor has capability of sensing, processing and transmitting radio signals by organizing a multi-hop routing topology among themselves [4, 32, 38]. Since sensors are battery powered and once their energy gets drained, they will no longer contribute to any of the task in the network. Depletion of their power is largely depending on transmission than sensing and processing [30]. Therefore in order to conserve nodes power, authors in history develop and engage many strategies like minimization of transmission energy by following shortest path, transmission power adaptation, data gathering or clustering, duty-cycle concept, construction of low interference topology and etc [9, 11, 12, 21, 28, 41].

1.1 Energy-delay tradeoff

It has been observed that routing topologies which aim to minimize nodes' power consumption or focus on saving nodes' power adversely affect delivery of data packets [10, 20]. This means that route with minimum energy

✉ Anjana Gupta
anjanagupta@dce.ac.in

[1] Department of Applied Mathematics, Delhi Technological University, Bawana Road, Rohini, Delhi, India

Published online: 16 March 2023

consumption causes longer delay in transmitting data packets between the nodes. This is because nodes use smaller transmission range to communicate in order to save their energy and prefer to have hop-by-hop transmission [25, 29]. Delay increases with number of intermediate hops existing between source to destination nodes because data is sensed, processed and transmitted as many time as number of intermediate hops and this contributes processing, propagation, transmission and queuing delay [24, 31]. Therefore delay minimization costs maximization of energy consumption since it requires longer transmission range of every source node to directly transmit data packets to its destination node. Work alone to minimize end-to-end routing delay shortens the network lifetime and work alone to minimize network energy consumption promotes routing delay. This issue in WSNs is called Energy-Delay Tradeoff [1, 5, 8, 22, 42] and has been a prime concern from past many years to resolve. Many authors work towards this Energy-Delay Tradeoff by formulating weighted function of some energy consumption causing factors as well as delay causing factors and use value of their formulated function to achieve desired routing goals. Some authors attempt to minimize delay factor more if the goal is to meet packet delivery deadline and some prioritize low transmission cost path in order to save nodes energy with some delay constraint [5, 13, 16, 36, 43, 45].

1.2 Research gaps

Existing approaches on Energy-Delay Tradeoff result delay constraint energy aware path between source to destination node based on hop counts by ignoring rest of the topology. Also work on minimization of nodes power consumption with delay optimization at topology level is so rare. Existing attempts do not visualise nodes power consumption explicitly while minimizing delay [15, 43] whether nodes are having variable transmission range or not. Some authors estimate queuing, transmission and propagation delay [18, 36] with the ignorance of transmission failure or retransmission and collision of data packets due to interference factor in large and dense networks. This retransmission and packets collision contribute to more delay in data delivery along with nodes' power consumption [8, 17, 31, 34]. Selvi et al. in [36] estimate end-to-end delay including propagation delay, processing delay, transmission delay and queuing delay, from source S to destination D using Eq. (1).

$$Delay_{end-to-end}(S,D) = \sum_{i,j}\left(\left(\frac{1}{\alpha-\beta}\right)+\frac{1}{\phi}+\frac{d(i,j)}{\mu}\right) \tag{1}$$

where α, β are constant, ϕ is bandwidth link, μ is the propagation speed and $d(i,j)$ is the distance between intermediate nodes, called hop length or link distance. Here authors didn't consider retransmission or interference factor in estimating end-to-end delay.

Moreover solution to Energy-Delay Tradeoff on bi-directional connected heterogenous WSNs where nodes are allowed to vary their transmission range and are assigned with appropriate power interval is still unexplored on topology level.

1.3 Motivation

These existing gaps motivate us to bring out a methodology to obtain delay minimum energy efficient connected routing topology considering interference factor which simultaneously work to minimize nodes' power consumption and topology delay on bidirectional connected heterogenous WSNs where each sensor node has been allowed to vary its transmission range within a power interval. To estimate delay, hop counts won't be considered since this work is dealt with bi-directional connected topology and each node has to present to maintain the topology connectivity therefore hop length (or link distance) between every communicating node is the first factor to estimate delay between every communicating sensor nodes as Eq. (1) implies that $Delay_{end-to-end}(S,D) \propto \sum_{i,j}(d(i,j))$ where ϕ and μ are assumed to be same for each sensor nodes in the network.

As delivery time between any two nodes increases with distance but opting a shortest distance doesn't imply fast data delivery between any two nodes until the path carries low interference in a dense network. In view of this, the new link-level interference model, introduced by Sun et al. in [39] has been engaged to mitigate interference in the network which in turn reduces topology delay as well as nodes' power consumption, is the new metric and second factor to estimate topology delay in the proposed work.

Briefing the points of motivation and improvements:

(1) Proposed methodology attempts to minimize overall topology delay with power consumption while the existing work focuses to minimize end-to-end delay between sender and receiver nodes only.

(2) Proposed work shows minimization of nodes power consumption explicitly in the form of power interval while no delay minimization work from literature shows reduction of nodes power consumption explicitly.

(3) Our work on developing energy and delay minimum topology on those WSNs whose nodes are assigned with power interval, considers interference factor

along with distance factor to estimate topology delay by engaging the link interference model introduced in [39] while most of the existing work ignores this retransmission or delay causing factor.

1.4 Contribution

The manifold contributions of the proposed work are as follows:

(1) Introduction of 'Interval data based graph model (IDGM)' and 'Sorted-interval data based graph model (S-IDGM)' for type of bi-directional connected heterogenous WSNs where each node is weighted with a power interval so that nodes can vary their transmission range in the power interval according to communication requirement.

(2) A new weighted 'Delay estimation (DE) function' based on distance and interference factors has been formulated after showing dependency analysis between link distance (hop length) and link interference towards optimality.

(3) A polynomial time 'Energy and Delay Optimization (EDO) algorithm' has been proposed to obtain an energy and delay optimized connected routing topology on S-IDGMs using DE function. EDO algorithm works under an optimal delay constraint in an energy efficient way and so it minimizes node power consumption and topology delay simultaneously.

(4) Energy efficiency of EDO algorithm is shown in term of percentage of energy saving by reducing the upper bound of maximally assigned power intervals along with minimization of topology delay.

(5) Extensive simulation work, graphical and statistical *t*-test analysis have been carried out to show the performance and comparison of EDO algorithm with delay minimum spanning tree, algorithms described in [23] and [19] on S-IDGMs in term of maximum and total topology delay, maximum power consumption interval and percentage of energy saving.

1.5 Organization of paper

The rest of the paper is organized in the following section: Sect. 2 surveys the related work, Sect. 3 introduces important terminologies, details of opted and proposed models. Section 4 presents the proposed problem with linear programming, pseudo-code and explanation of proposed algorithm. Performance evaluation and comparison are shown in Sect. 5. Section 6 concludes the paper.

2 Related work

There has been lot of research work carried out during past many years to optimize power consumption and routing delay as well as to have an optimal solution to the Energy-Delay Tradeoff in WSNs. Authors worked with different strategies and introduced many methodologies to either optimize transmission energy or routing delay or both with the involvement of different estimation metrics. Here, we have mentioned some of them targeting the same goal with different approaches. Some may involve different models of WSNs.

Authors in [40] proposed an approach to minimize energy consumption where packets need to be delivered within a deadline. The approach involves optimal scheduling for packet transmission by varying power level of nodes and transmission time so that the packets must be delivered within a given time frame. In this, delay causing factors while routing has not been considered explicitly. In [2], an energy-aware multi-hop centralized routing scheme can be seen where static as well as mobile gateway nodes have been engaged to handle packet delivery tasks to minimize energy consumption and end-to-end transmission delay. Authors used weighted fair queuing (WFQ) packet scheduling methodology to acknowledge traffic at each intermediate node and a node with least load had been considered for data forwarding from source to destination node. However, WFQ mechanism might be complicated and costly for resource constraint sensor nodes. Authors in [30] come-up with an algorithm to find lowest energy consuming path following a delay bound. In this, path delay has been estimated undertaking medium access layer contention. Authors in [15] introduced energy efficient routing scheme with delay bound using Bellman-Ford algorithm, path length and hop count concept. Approaches involved in this paper limits the number of iterations and hop counts while determining the delay constraint routing path. A real-time delay attack detection and isolation scheme has been proposed in [33] where methodology is based on machine learning mechanism along with a route-handoff mechanism. A concentric circular bands (CCBs) based data forwarding scheme from source to sink node optimizing transmission energy or routing delay or both has been proposed in [5]. In this, authors considered dense and uniformly distributed homogenous sensor network to estimate energy consumption and delay. An energy optimizing technique has been proposed in [14] where data collection in large and dense WSNs are accomplished using projection-based compressive data gathering which aids to conserve nodes' energy and extend network lifetime. Authors in [37] proposed an algorithm for detecting cluster head based on trust value among neighbor nodes based on the Received Signal Strength Index (RSSI) to reduce the network energy drainage and enhance the network lifetime. Authors in [8] introduced delay concerned routing protocol by believing that delay can be minimized by reducing intermediate hop counts. Using realistic unreliable link models with hop counts, hop length and transmission power factors, authors in [45] obtained lower bound of energy-delay tradeoff in Rayleigh fast fading and Rayleigh block fading channels. Authors in [16] used multiple mobile base stations to minimize latency and engaged minimum Wiener index spanning tree (MWST) as a routing topology. In this, number of hop counts had been considered to estimate latency. Although adopting multiple mobile base stations make the network costly. Further, a 'delay-constrained energy multi-hop (DCEM)' algorithm, proposed in [18], provides a solution of TED (Tradeoff between Energy and Delay) optimization problem in real-time cluster-based routing by adjusting the distance and residual energy factor. Authors used queuing theory to find queuing delay to estimate end-to-end routing delay. However, authors ignored the unreliability factor in queuing delay estimation, also it is defined for homogenous WSN only. Authors in [17] estimated delay considering link reliability as well as unreliability factors by involving access delay and collision delay. Collision delay had been taken equivalent to the number of interfering nodes at a time of transmission lying inside the transmission disc of other nodes. Working of the algorithm is very local. Authors in [43] dealt with the Energy-Delay Tradeoff considering unreliable links in both hop-by-hop routing and end-to-end routing of data packets. With the minimization of delay based on transmission and queuing delay, authors provided an optimal power assignment to each links involved in transmission. Optimal power assignment to the nodes in the topology had been left in this work. Authors in [10] proposed MDET algorithm to construct minimum delay and energy-efficient tree for data flooding in low-duty-cycle WSNs. This algorithm finds shortest path using expected transmission count and improves the energy efficiency in a local manner. Authors involved sleep-wakeup delay of duty-cycle with packet transmission delay of each hop under unreliability situation to estimate end-to-end delay. Minimization of delay and energy consumption by engaging optimal duty-cycle can be also be seen in [42]. Authors believed to take minimum intermediate hops in dense WSNs undergoing duty-cycle. Authors in [36] proposed DCEMRA algorithm to minimize energy consumption in cluster-based routing without increasing end-to-end delay. In this, nodes' remaining energy was computed using distance and data delivery time. DFS algorithm was used to have low cost multiple paths and one of the best path was chosen based on beacon messages. Authors formulated a function to estimate end-end delay comprises of transmission delay, propagation delay, queuing delay and processing delay which majorly depends on distance between the intermediate nodes. However, the algorithm works on homogenous WSNs.

Major research in this domain has been done to optimize end-to-end routing delay between sender to receiver node and transmission energy cost of data packets travelling the same route. No work has explicitly shown reduction of nodes' power consumption with delay minimization, neither at transmission level nor at topology construction level.

Different from previously related work, we first attempt to obtain energy and delay optimized connected topology for a heterogenous bidirectional WSN where power assignment of each node has been explicitly considered throughout the methodology. In this paper, we work with those WSNs whose nodes have been assigned with power intervals so that each node can adjust its transmission range according to the communication requirement. Furthermore, two factors that are distance between the nodes [36] and link interference model, introduced in [39], have been considered to estimate topology delay to obtain delay minimized and energy efficient reliable routing topology on an interval data based graph model of WSNs.

3 Proposed work

Our work comprises of previously defined definitions, formulas, models, newly proposed models, new functions and algorithms. Moving ahead step-by-step with proper description of every important terminology, models, functions and algorithms.

3.1 Preliminaries

Here we introduce some important terminologies, concepts and abbreviations which have been a part of the proposed work.

(1) *Graphical layout of WSNs* A bidirectional WSN can be represented as a connected undirected graph, denoted as $G(V, E)$ where V denotes the set of all sensor nodes and E denotes the set of all transmission links.

(2) *Transmission cost TC* TC is the cost of energy required by a node u to transmit data packets to another node v and is equivalent to $c_1 + c_2 d^\alpha$, where $d = ||uv||$, α $(2 \leq \alpha \leq 4)$ is the path loss gradient depending on transmission environment, c_1 and c_2 are some constant depending on electronic characteristics of wireless devices [26]. It has been denoted as $TC(u, v)$ when the transmission

is from u to v and $TC(v, u)$ when the transmission is from v to u. In bi-directional WSNs, $TC(u, v) = TC(v, u) = max\{TC(u, v),\ TC(v, u)\}$, can be taken symmetrically [28] and written as $TC(uv)$, uv depicts the bi-directional link. Panda and Shetty in [28] compute transmission cost of a bi-directional link uv as $TC(uv) = ||uv||^2$, an approximate estimation.

(3) *Nodes power assignment* In WSNs, a power function $P : V \rightarrow R^+$ has been defined to assign sufficiently high transmit power to each sensor node u so that it can directly communicate its farthest node v i.e. $P(u) = max_{uv \in E}\{TC(uv)\}$ [35].

(4) *Power interval PI* When the power assignment to each node has been done in the form of interval so that nodes can vary their transmission range from minimum to maximum power level depending upon the transmission need. It is denoted as $PI(u) = [P_{min}(u), P_{max}(u)]$ for any arbitrary node u. $P_{min}(u) = min_{uv \in E}\{TC(uv)\}$ and $P_{max}(u) = max_{uv \in E}\{TC(uv)\}$ [19]. Let's see an illustration of PI assignment on a network, redrawn from [28], using the theory introduced in [19]. The weight on each edge of Fig. 1 is the square of the Euclidean distance between end nodes i.e. $||ab|^2$ for two adjacent nodes a and b, taken as a transmission cost of bi-directional link ab [28]. Now consider node 'a' of Fig. 1, mentioning all the links incident on 'a' with Euclidean distance i.e. $||ab||^2 = 252, ||ac||^2 = 14, ||ad||^2 = 589, ||ae||^2 = 34$. Minimum and maximum weight among all the links incident on 'a' is 14 and 589 respectively. Therefore $PI(a) = [14, 589]$ is the range of power assigned to node a i.e. $P_a \leftarrow PI(a)$ and likewise $PI(b) = [75, 252], PI(c) = [12, 428], PI(d) = [75, 589]$ and $PI(e) = [12, 452]$ are computed power intervals corresponding to nodes b, c, d, and e respectively.

Hence
P_a = [14,589],
P_b = [75,252],
P_c = [12,428],
P_d = [75,589]
and $P_e = [12, 452]$ are assigned power intervals in Fig. 1. Moreover, interval graph theory results a power interval to each link which is the intersection of power intervals of its end nodes as for example available power interval on a bi-directional link ab is $PI(a) \cap PI(b)$ and denoted by PI_{ab}. Table 1 shows the available power interval on each link of the WSN given in Fig. 1 [19].

(5) *Maximally assigned power node* A node having power interval with maximum upper bound i.e. *max PI*, is called maximally assigned power node. A maximally assigned power node retains a communication link with maximum transmission cost as in Fig. 1 nodes a and d retain a communication link of max transmission cost 589 and so their power intervals have maximum upper bound. Therefore a and d are called maximally assigned power nodes.

(6) *Minimally assigned power node* A node having power interval with minimum upper bound i.e. *min PI*, is called minimally assigned power node as in Fig. 1 node b is minimally assigned power node.

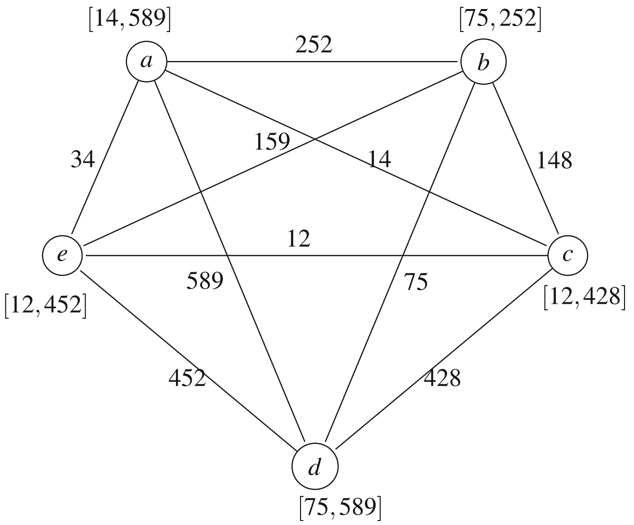


Fig. 1 Weighted complete graph with power intervals

3.2 Interval data based graph model (IDGM) and Sorted-interval data based graph model (S-IDGM) of WSNs

In this section, we introduce IDGM and S-IDGM of a bidirectional WSN where each node is carrying a weight in interval form. The S-IDGM has been used in our work to compare the nodes' power intervals by just seeing their labels which consequently reduces the algorithmic complexity. Not every WSN can form an interval graph layout even having interval weight on each node [7]. Therefore the type of WSNs whose underlying graph models show the interval weight on each node, do not require to satisfy interval graph properties, are called 'interval data based

Table 1 Available power interval on each link of Fig. 1

$PI_{bc} \in [75, 252]$	$PI_{ac} \in [14, 428]$
$PI_{be} \in [75, 252]$	$PI_{dc} \in [75, 428]$
$PI_{bd} \in [75, 252]$	$PI_{ae} \in [14, 452]$
$PI_{ba} \in [75, 252]$	$PI_{de} \in [75, 452]$
$PI_{ce} \in [12, 428]$	$PI_{ad} \in [75, 589]$

graph model (IDGM)' of WSNs. Figure 2 is one of an illustration of IDGM.

Every interval graph is an IDGM but converse doesn't hold always because in an interval graph, two nods are adjacent if and only if their power intervals intersect [6]. In Fig. 2, one can see power intervals of non-adjacent nodes are also intersecting hence this can't be an interval graph so it is called an IDGM of WSN.

Sorted-interval data based graph model of WSN (S-IDGM) Sorting power intervals of an IDGM in an increasing order of their upper bounds results a S-IDGM. Figure 4 shows S-IDGM of Fig. 2.

An intersection model, described in [6] has been used for sorting the power intervals because of its low complexity. Even though the intersection model, given in [6] has been described for interval graphs, it can work for IDGM also but in one way only, not viceversa i.e. one can have an intersection model of an IDGM for a sorting purpose but can't form an IDGM from an intersection model if the IDGM is not an interval graph. Figure 3 shows the intersection model of Fig. 2.

Sorting and relabelling the nodes as $u_1, u_2, ..., u_n$ following increasing order of power intervals is the very first requirement of our methodology. A node with higher subscript is the node with higher power interval. Figure 4 which is a S-IDGM of Fig. 2, visualising that power intervals are increasing with nodes' subscript and viceversa. Although there are two cases, which have been observed during sorting; first is two or more nodes can have same upper bound of power consumption as seen in Fig. 2, therefore labelling order depends on the lower bound of their power intervals i.e. node with lesser lower bound of power consumption has been labelled with lesser subscript and the node with higher power consuming lower bound has been labelled with higher subscript, as seen in Fig. 4. Secondly, if two or more nodes have same power intervals as shown in Fig. 5a, labelling of the same power nodes can take any consecutive order of the running subscript since all the resultant S-IDGM come out to be isomorphic, see Fig. 5b and c representing the isomorphic models of Fig. 5a. The second requirement of our methodology is to rename minimally assigned power node with leftmost vertex (*LMV*) and maximally assigned power node with rightmost vertex (*RMV*) of the obtained S-IDGM. Therefore it can be said that a vertex whose subscript is the minimum number is called leftmost vertex (*LMV*) and a vertex whose subscript is the maximum number among all vertices' subscript of a S-IDGM is called rightmost vertex (*RMV*). Hence *LMV* is the minimally assigned power node, having *min PI* and *RMV* is the maximally assigned power node, having *max PI* [19]. Let $G(V, E)$ be a S-IDGM, then

$$LMV(G(V,E)) = \{v_j | j = min_{v_i \in V}(i)\}$$
$$RMV(G(V,E)) = \{v_k | k = max_{v_i \in V}(i)\}$$

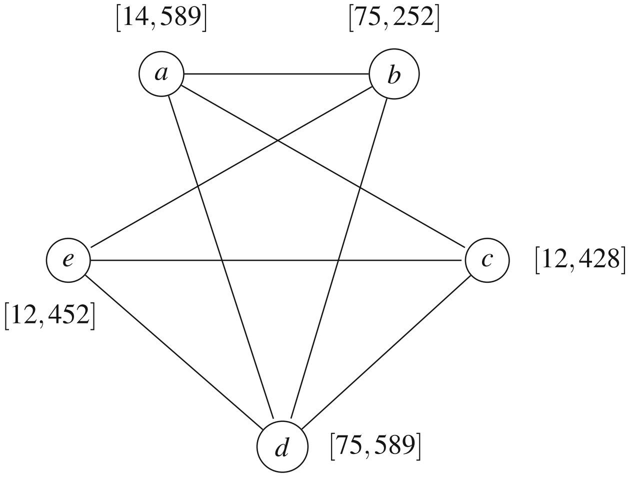


Fig. 2 Interval data based graph model (IDGM) of bidirectional WSN

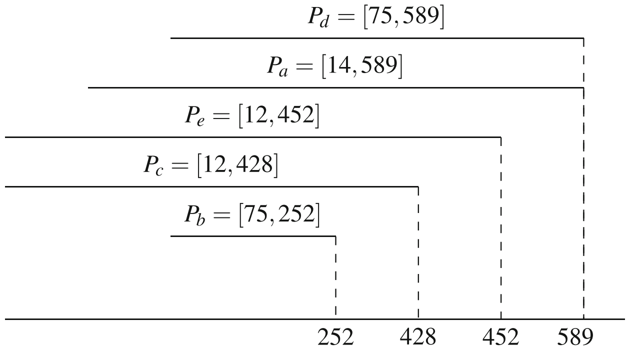


Fig. 3 Intersection model of IDGM, given in Fig. 2

3.3 Link-level interference model

Now introducing Link-level interference model which has been proposed by Sun et al. in 2015 [39] to compute link interference in bidirectional WSNs. For a given graph

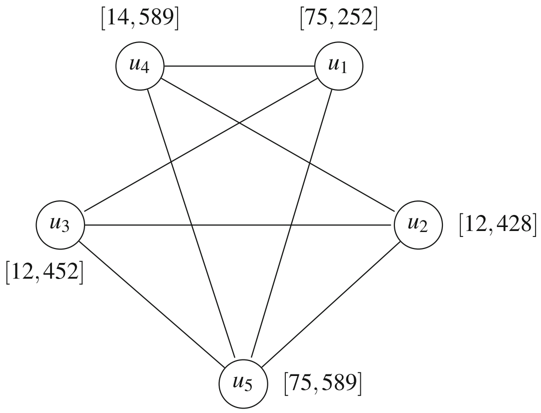


Fig. 4 S-IDGM of IDGM, given in Fig. 2

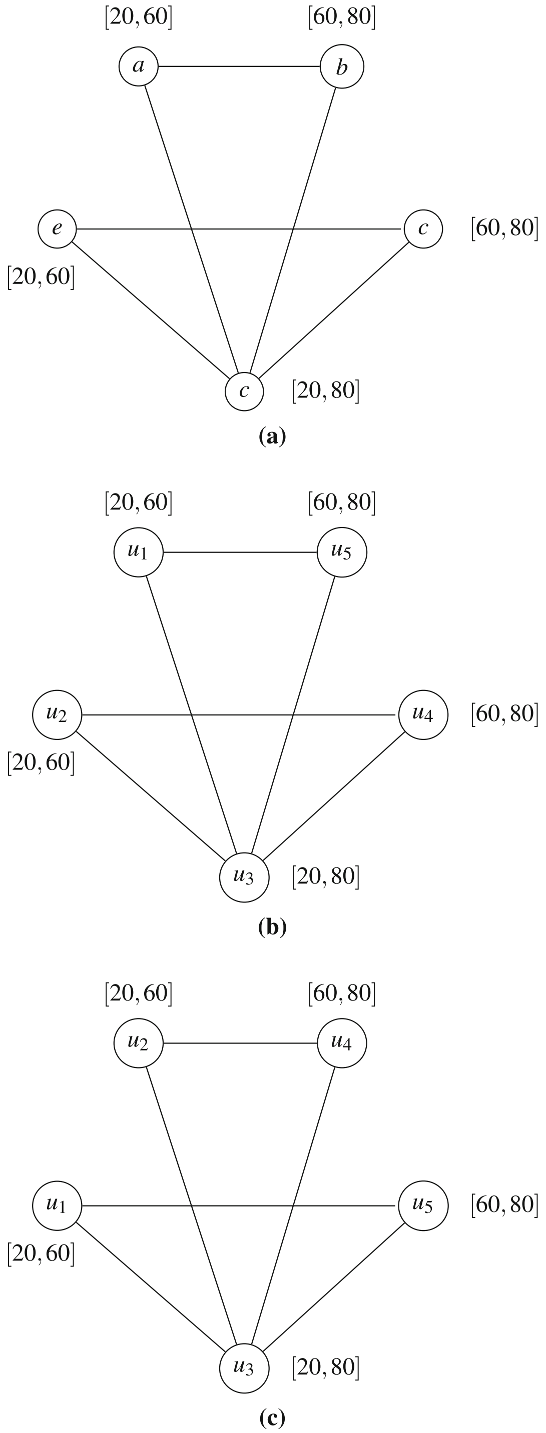


Fig. 5 **a** An IDGM where nodes are having same power intervals **b** First S-IDGM of Fig. 5(a) **c** Second S-IDGM of Fig. 5(a)

$G(V, E)$, interference on any bidirectional link uv is denoted by I_{uv} and is equal to the number of distinct interfering links if any of their end node resides inside the disc of node u or v, is called Link-level interference model and is given by Eq. (2).

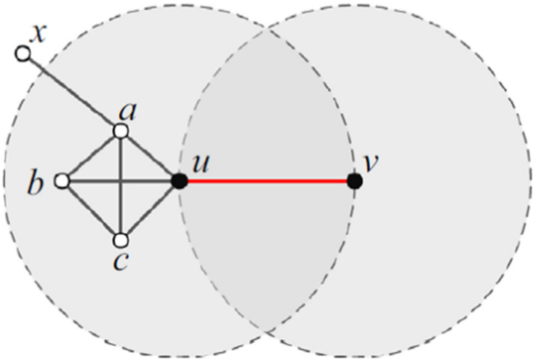


Fig. 6 Disc topology of link uv

$$I_{uv} = |\{xy \mid xy \in E(G),\ x \text{ or } y \text{ or both } \in \text{Disc}\ (u, d(u,v)) \text{ or Disc}\ (v, d(v,u))\}| \tag{2}$$

where xy is another bidirectional link, $d(u, v)$ is an Euclidean distance between u and v, Disc $(u, d(u, v))$ is the transmission disc whose centre is u and radius is $d(u, v)$. Similarly Disc $(v, d(v, u))$ is the transmission disc whose centre is v and radius is $d(v, u)$, where $d(v, u) = d(u, v)$ [41].

An illustration of computing link-level interference model using Eq. (2) are shown on network, given in Fig. 6, taken from [39]. The link-level interference of link uv using Eq. (2) come out to be 7. This link-level interference model outperforms the link interference model described in [28] which counts the number of nodes existing inside the transmission disc of u or v. According to the model described in [28], the link interference of link uv, given in Fig. 6, come out 3 based on nodes count. Hence link-level interference model, proposed by Sun et al. [39] provides more significant results.

3.4 Dependency analysis of link distance and link-level interference model towards optimality

It has been claimed in many research papers that link distance and its interference are so much depending on each other such as more the distance between the nodes more will be the interference between them [17, 26, 41]. Sun et al. in his paper [39] says that link interference increases with link distance and shows the effect of link distance on link-level interference model with positive correlation.

This motivates us to analyze the effect on interference when the approach is related to obtain distance minimum spanning tree as well as the effect on distance when the approach is related to obtain interference minimum spanning tree. For this, first we put the weight on each link is its

Table 2 Total link distance and total link interference calculation on 20 distinct networks

Sr. no	WSNs		Euclidean distance-based MST		Link-level interference model-based MST	
	Nodes	Edges	Total distance	Total interference	Total distance	Total interference
1	9	27	4.044	38	4.154	34
2	11	30	3.54	44	3.63	43
3	13	38	3.68	60	3.88	55
4	15	49	3.73	97	4.17	88
5	18	100	5.394	96	5.65	87
6	20	120	3.68	124	3.84	117
7	23	129	4.152	144	4.36	137
8	25	141	4.106	180	4.425	172
9	28	202	6.177	165	6.459	161
10	30	268	5.731	174	5.754	173
11	32	291	5.29	240	5.485	218
12	33	265	4.91	245	5.17	243
13	35	307	5.507	254	5.753	239
14	37	304	5.284	310	5.468	283
15	39	433	7.88	251	8.25	230
16	41	461	6.49	248	6.65	239
17	43	436	6.034	353	6.28	300
18	45	479	7.53	341	7.75	334
19	48	608	7.27	363	7.56	326
20	50	608	6.92	394	7.35	377

distance, compute the MST using Prim's algorithm and observe the reduction in total distance as well as total interference. Then, we put weight on each link is its interference using Eq. (2), compute the MST using Prim's algorithm and observe the reduction in total interference as well as total distance. This process has been done on 20 distinct networks and resultant data values have been tabulated in Table 2.

Some concluding points have been made after observing the tabulated data

- Resultant data in Table 2 reveals that there is always a difference between value of total interference obtained in interference-based MST and distance-based MST i.e. optimal interference reduction is achieved in interference-based MST only, not in distance-based MST, same things happen with link distance. This implies that one cannot optimally reduce both the distance and interference factor by considering the single factor; interference is optimally reduced in interference-based MST and distance is optimally reduced in distance-based MST. Therefore to minimize both the factors simultaneously, a weighted function of link distance and link interference needs to formulate. (Discussion of weighted function is carried forward to next section.)
- Link-level interference is more depending on network density than link distance. More dense the network is, more will be data collision and so more retransmission.
- In dense network, minimization of interference factor when the hop length are not much longer, reduces total distance to some optimal extend as well but this doesn't happen in sparse network. This implies that in dense network, interference factor has more dominance over distance factor.

3.5 Delay estimation (DE) function

After analysing the results obtained in previous section, a conclusion is made to construct a new function which optimally balances the minimization of link interference with hop length (link distance) when any algorithm is applied after it. Since both the hop length and link-level interference are used metrics to estimate delay between every communicating nodes, therefore this new function is known as 'Delay estimation (DE) function' and is given by Eq. (3).

This DE function assigns weight on each edge which is a combination of link-level interference and link distance, using Eq. (3). Since $\alpha = 1$ of DE function results interference-based MST providing optimal total interference value and $\alpha = 0$ results distance-based MST providing optimal total link distance value. Exclusion of both these alpha values from Eq. (3), the DE function assigns weight to each link with the combination of link distance and link interference which results a weighted MST whose total interference and total hop length values lie close to their respective optimal values (better than when $\alpha = 0$ and $\alpha = 1$).

Moreover, for maintaining the generality, $\alpha = 0.5$ is taken to assign the weight on each link in the network. Therefore Eq. (3) reduces to Eq. (4).

$$DE(uv) = \alpha * I_{uv} + (1 - \alpha) * d(u, v) \tag{3}$$

where $\alpha \in (0, 1)$ and uv is any bidirectional link.

$$DE(uv) = I_{uv} + d(u, v) \tag{4}$$

Role of DE function

- DE function assigns weight to each edge using Eq. (4) which is nothing but the required delay estimation of each link in the network.
- DE function is used to obtain optimal delay constraint Δ for the complete execution of proposed algorithm.
- DE function estimates total and maximum topology delay (*TD*) in the resultant topology using Eq. (4).

4 Problem definition and proposed algorithm

To find an energy and delay optimized connected routing topology on S-IDGM $G(V, E)$ of WSN, following is the formulated linear programming problem (LPP) demonstrating the required optimization problem.

$$min\ max_{uv \in E} DE(uv) = min\ max_{uv \in E}\{I_{uv} + d(u, v)\} \tag{5}$$

$$min\ max_{u \in V} PI(u) = min\ max_{u \in V}[P_{min}(u), P_{max}(u)] \tag{6}$$

subject to

$$\sum_{uv \in E} uv = |V| - 1 \tag{7}$$

$$DE(uv) \leq \Delta \tag{8}$$

$$P_{min}(u) \leq PowerConsumption(u) \leq P_{max}(u) \tag{9}$$

$$max\{P_{min}(u), P_{min}(v)\} \leq TC(uv) \leq min\{P_{max}(u), P_{max}(v)\} \tag{10}$$

$$u, v > 0\ \forall\ u, v \in V \tag{11}$$

where $uv's$ are bidirectional transmission links, Δ is an optimal delay constraint i.e. an optimal max delay bound obtained using delay minimum spanning tree by the proposed algorithm during its execution. Equations (6) and (7) depict the desired optimization goals that are minimization of maximum topology delay and minimization of maximum node's power interval respectively. Equations (8)–(11) depict the required constraints attached with the given optimization problem. Equation (8) gives a bound on number of links i.e. total number of links in the resultant topology should be equal to one less than the number of nodes, Eq. (9) directs to keep only those link whose delay estimation (DE) must be less than or equal to the obtained delay constraint, Eq. (10) allows each node to spend its energy within its assigned power interval, Eq. () allows variation of transmission cost of each link within the given bounds (which is actually an available power interval on each link) as transmission cost is directly depending on the power used by a link during transmission and Eq. (11) requires each node to present in the resultant topology.

Energy and Delay Optimization (EDO) algorithm

To the defined optimization problem, a polynomial time algorithm has been proposed, named as 'Energy and Delay Optimization (EDO)' algorithm. The proposed EDO algorithm undergoes three main phases. The very first phase (step 1 and 2) provides an optimal delay constraint i.e. Δ based on DE function and DE MST, second phase (from step 3 to 17) computes the delay constraint energy efficient path cover on S-IDGM of WSN and third phase (step 18 onwards) constructs energy and delay optimized connected routing topology of S-IDGM if the obtained path cover contains more than one path component. Methodology of the proposed EDO algorithm is the motivation of the algorithms proposed in [19] and [23].

Algorithm 1 Energy and Delay Optimization (EDO) algorithm

Input : A S-IDGM $G(V,E)$ where $|V| = n$ and $|E| = m$ of a bidirectional WSN.

Output : A delay-constraint and energy-efficient connected routing topology which results Eq. (5) and (6) of the given LPP following all the attached constraints.

```
1:  Compute $DE(uv)$ of each link $uv$ of $G(V,E)$ using Eq. (4).
2:  Obtain MST of the delay weighted graph $G(V,E)$ using Prim's algorithm and optimal delay constraint $\Delta = max_{uv \in MST}\{DE(uv)\}$.
3:  Define path cover, $PC = \phi$ initially
4:  Let $UncoverV = \{v_1, v_2, ..., v_n\}$
5:  Start $num = 1$ and $Cnum = v_n$        ▷ ($num$ shows the number of path components; $Cnum$ shows the running path component)
6:  $UncoverV = UncoverV / v_n$
7:  while $UncoverV \neq \phi$ do
8:      Compute $S = \{v_k \mid v_k \in UncoverV,\ DE(v_k v_n) \leq \Delta$ and $k > n\}$ and $T = \{v_k \mid v_k \in UncoverV,\ DE(v_k v_n) \leq \Delta$ and $k < n\}$
9:      if $S$ and $T$ of $v_n$ is empty then
10:         $[v = RMV(UncoverV); num = num + 1; Cnum = v]$
11:     if $S \neq \phi$ then
12:         $[v = RMV(S); Cnum = Cnum \cup v]$        ▷ ($\cup$ shows the joining of vertices)
13:     else
14:         $[v = LMV(T); Cnum = Cnum \cup v]$
15:     $v_n = v$
16:     $uncoverV = uncoverV / \{v_n\}$
17: return $PC = \{C_1, C_2, ..., Cnum\}$
18: if $|PC| = 1$ then
19:     $C_1$ is the required solution
20: else
21:     Choose a component $C_i$ from $PC$ and $PC \leftarrow PC/\{C_i\}$
22: while $PC \neq \phi$ do
23:     Pick another component $C_j \in PC$ such that $C_j$ and $C_i$ are intersecting
24:     Compute $nghC_j = \{v_i \mid v_i \in C_i$, providing the direct connectivity to $C_j$ under $\Delta\}$ and $nghC_i = \{v_j \mid v_j \in C_j$, providing the direct connectivity to $C_i$ under $\Delta\}$
25:     if $LMV(nghC_i) < LMV(nghC_j)$ then
26:         $C_i = C_i \cup C_j$ will be the new path with the joining of $LMV(nghC_i)$ to the leftmost adjacent vertex from $nghC_j$.
27:     else
28:         $C_i = C_i \cup C_j$ will be the new path with the joining of $LMV(nghC_j)$ to the leftmost adjacent vertex from $nghC_i$
29:     $PC = PC/\{C_j\}$
30: $C_i$ is the required delay constraint energy efficient connected routing topology.
31: For every link $uv \in C_i$, update $DE(uv)$ using Eq. (4) and compute $min\ max_{uv \in C_i} DE(uv)$, $min\ \sum_{uv \in C_i} DE(uv)$.
32: For each node $v \in C_i$, update $PI(v) = [min\{PI_{vu, \forall vu \in C_i}\}, max\{PI_{vu, \forall vu \in C_i}\}]$ and compute $min\ max_{v \in C_i} PI(v)$.
```

*Working illustration with step-by-step explaination of EDO algorithm*consider a S-IDGM of an arbitrary WSN given in Fig. 7 where each node has been weighted with a power interval which results to power interval on each edge as shown in Fig. 7, $PI(v_1) = [50, 105]$ and $PI(v_2) = [78, 150]$, then the edge between v_1 and v_2 has $PI_{v_1 v_2} = PI(v_1) \cap PI(v_2) = [78, 105]$. Likewise each edge has been resulted with a power interval due to its end nodes and so these power intervals are called available power intervals. Weight other than the available power interval on each edge in Fig. 7 is a delay estimation.

Step 1 Computes delay estimation on each edge of graph, given in Fig. 7, using Eq. (4) and so $DE(v_1 v_2) = 2.617$, $DE(v_1 v_3) = 8.923$, $DE(v_2 v_3) = 8.923$, $DE(v_3 v_4) = 7.76$, $DE(v_3 v_5) = 7.76$, $DE(v_3 v_6) = 8.99$, $DE(v_4 v_5) => 5.78$, $DE(v_4 v_6) = 5.52$.

Step 2 DE-MST is obtained using Prim's algorithm of Fig. 7 with edges $\{v_1 v_2, v_2 v_3, v_3 v_4, v_4 v_5, v_4 v_6\}$ and results $\Delta = max\{6.803, 8.697, 7.584, 7.717, 7.876\} = 8.697$.

Step 3 Path cover is defined as $PC = \phi$ initially depicts that no vertex is covered by any path component of PC.

Step 4 *UncoverV* is the set of unvisited vertices which are not covered by PC. So in this illustration $UncoverV = \{v_1, v_2, v_3, v_4, v_5, v_6\}$ since no vertex is covered by PC.

Step 5 $num = 1$ shows that PC starts to obtain its first path component $C1$ as PC is like $\{C_1, C_2, ..., Cnum\}$. C_1 visits maximally assigned vertex first i.e. $RMV(UncoverV)$, therefore $C_1 = v_6$.

Step 6 *UncoverV* updates to $UncoverV = \{v_1, v_2, v_3, v_4, v_5\}$ as v_6 is covered by C_1.

Step 7 Initialization of while loop brings execution of steps from 8 to 16 until *UncoverV* becomes empty.

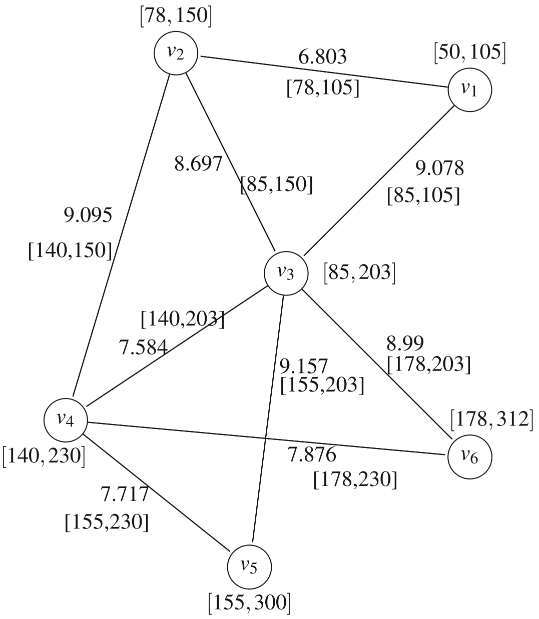


Fig. 7 A S-IDGM of an arbitrary WSN where weights on each edge is the estimated delay of each link and available power interval on each link

Step 8 S and T has been computed for each node present in the original graph. For $v_n = v_6$, $S(v_6) = \phi$ since S comprises only those unvisited neighbour nodes of v_6 whose subscript is greater than 6 with link delay ≤ 8.697. Unlikely $T(v_6) = \{v_4\}$ comprises of those unvisited neighbour nodes of v_6 whose subscript is less than 6 with link delay ≤ 8.923.

Step 9 Checks the emptiness of S and T.

Step 10 If incase both S and T are empty then C_1 gets terminated with the initialization of C_2 which again starts with *RMV* of *UncoverV* and jumps to step 15.

Step 11 Checks the non-emptiness of set S if the condition of step 9 is not satisfied. If S comes out non-empty then

Step 12 C_1 extends itself by adjoining *RMV* from S.

Step 13 Checking of T's emptiness if the condition of step 11 is not approved like in the given illustration where S is empty and T is non-empty then

Step 14 C_1 extends itself by adjoining *LMV* from T so C_1 becomes $C_1 = v_6v_4$.

Step 15 Now v_4 becomes v_n i.e $v_n = v_4$.

Step 16 Therefore $UncoverV = \{v_1, v_2, v_3, v_5\}$. (Since $UncoverV \neq \phi$ execution of while loop causes computation of S and T for v_4, results $S = \{v_5\}$ and $T = \{v_3\}$ and $C_1 = v_6v_4v_5$. Now $v_n = v_5$, $UncoverV = \{v_1, v_2, v_3\}$, $S(v_5) = \phi$ and $T(v_5) = \phi$ therefore step 9 and 10 terminate C_1 with the initiation of $C_2 = RMV(UncoverV)$ i.e $C_2 = v_3$. Now $v_n = v_3$, $UncoverV = \{v_1, v_2\}$, $S(v_3) = \phi$ and $T(v_3) = \{v_2\}$ so C_2 extends to v_3v_2 and likewise $v_n = v_2$ results $C_2 = \{v_3v_2v_1\}$. Finally *UncoverV* becomes empty and while loop gets terminated.)

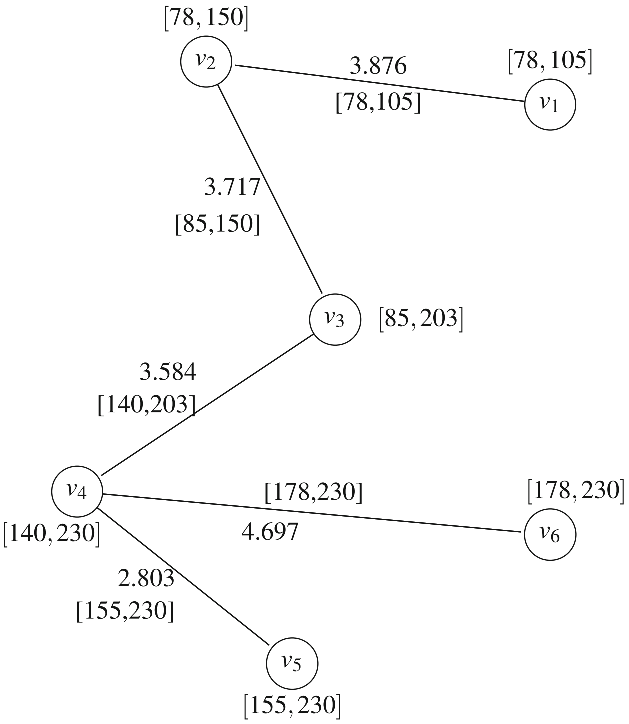


Fig. 8 Connected topology by EDO algorithm of S-IDGM of Fig. 7

Step 17 Return $PC = \{C_1, C_2\}$ i.e. $PC = \{v_6v_4v_5, v_3v_2v_1\}$.

Step 18 Checks the cardinality of *PC*, if $|PC| = 1$ then

Step 19 C_1 is the required solution.

Step 20 In the given illustration we follow the else part as $|PC| = 2$.

Step 21 Let $C_i = v_6v_4v_5$ and update $PC = \{v_3v_2v_1\}$.

Step 22 Initialization of while loop since $PC \neq \phi$.

Step 23 Let $C_j = v_3v_2v_1$, also C_i and C_j are intersecting in the original graph, given in Fig. 7.

Step 24 $nghC_i$ and $nghC_j$ shows the the direct neighbourhood of path component C_i and C_j respectively under the delay constraint. Therefore $nghC_j = \{v_4\}$ and $nghC_i = \{v_3\}$ for the given illustration.

Step 25 $LMV(nghC_i)$ is v_3, $LMV(nghC_j)$ is v_4 and $v_3 < v_4$, this satisfies the condition that $LMV(nghC_i) < LMV(nghC_j)$.

Step 26 Therefore C_i is joined to C_j by opting the connection from v_3 to v_4, see in Fig. 8.

Step 27 If incase condition of step 25 doesn't satisfy then

Step 28 C_i is joined to C_j by making connection of $LMV(nghC_j)$ to its leftmost adjacent vertex from $nghC_i$.

Step 29 *PC* becomes empty for the given illustration, therefore end of while loop.

Step 30 Resultant C_i, given in Fig. 8 is the optimized connected routing topology of Fig. 7. *(Fig. 8 shows the updated weights on nodes as well as on edges.)

Step 31 Computation of link delay of the optimized topology using Eq. (4). Delay weights are updated in Fig. 8 which results $min\ max_{uv \in C_i} DE(uv) = 4.697$ and $min\ \sum_{uv \in C_i} DE(uv) = 18.677$.

Step 32 For each node of the optimized topology, let's find the updated power assignment to each node; $P(v_1) = [78, 105]$,

$P(v_2) = [min\{78, 85\}, max\{105, 150\}] = [78, 150]$,

$P(v_3) = [min\{85, 140\}, max\{150, 203\}] = [85, 203]$,

$$P(v_4) = [min\{140, 155, 178\}, max\{203, 230\}] = [140, 230],$$

$P(v_5) = [155, 230]$ and $P(v_6) = [178, 230]$. Therefore $min\ max\ PI = [140, 230]$ at node v_4.

Remarks

(1) Max *TD* in the original graph, given in Fig. 7 is 9.157 and max node *PI* is [178, 312]. EDO algorithm reduces max *TD* from 9.157 to 4.697 and max node *PI* from [178, 312] to [140, 230] (refer Fig. 8). Also total *TD* in the original graph, given in Fig. 7 is 74.997 reduces to 18.677 in the optimized topology of Fig. 8.

(2) Minimization of max node *PI* leads to the minimization of max power consumption as 320 is the maximum power consumption of atleast one link in the original graph and hence it reduces to 230 in the optimized topology (refer Figs. 7 and 8). This leads to network energy saving. Therefore % of energy saving by EDO algorithm when max node *PI* reduces from [178, 312] to [140, 230] is obtained by doing $\frac{(312-230)}{312} \times 100 = 26\%$.

Theorem 1 *Δ is an optimal delay constraint obtained by EDO algorithm.*

Proof Δ is the maximum edge weight on the obtained MST where the edge weight is the delay estimation of each link (computed using Eq. (4)). Assume that Δ is not an optimal constraint this means that there exists another spanning tree say *Tr* whose maximum edge weight must be less than Δ. But this is not possible as *Tr* cannot be another MST giving different solution of the same graph. Hence Δ is an optimal delay constraint. □

Table 3 Data obtained on maximum topology delay and total topology delay after simulation

Number of nodes	Number of edges	Maximum topology delay				Total topology delay			
		DE MST	EDO algorithm	Algorithm in [19]	Algorithm in [23]	DE MST	EDO algorithm	Algorithm in [19]	Algorithm in [23]
6	9	4.923	4.923	4.997	4.923	15.345	15.606	18.081	15.606
9	27	6.931	6.931	7.970	7.89	38.154	42.898	56.615	48.6
11	30	7.445	8.445	9.722	9.577	45.595	52.083	81.405	55.646
13	38	9.566	11.566	11.793	11.614	58.875	87.303	95.438	92.79
15	49	11.336	11.512	13.6	13.6	95.002	116.324	145.06	136.54
18	100	11.7	11.7	16.923	16.915	91.626	126.336	253.425	176.82
20	120	17.479	18.479	18.602	18.586	121.793	246.42	294.52	214.1
23	129	17.557	17.557	21.59	20.653	142.275	265.92	340.22	228.676
25	141	17.614	17.614	23.424	19.542	178.243	281.27	378.288	303.522
28	202	14.3	16.468	27.046	27	170.4	302.278	503.12	419.67
30	268	11.373	20.44	28.84	28.614	177.744	232.017	600.5	456.34
32	291	13.282	23.35	30.856	29.714	226.422	305.87	761.024	674.07
35	307	17.685	18.31	33.712	33.67	244.72	364.86	740	634.65
37	304	26.786	28.786	35.667	33.582	288.433	672.15	856.106	698.47
39	433	12.363	23.853	38.1	38	237.133	406.834	1041.672	668.4
41	461	10.305	16.367	40	38.654	246.535	355.252	1065.673	774.827
43	436	20.513	20.513	41.819	37.58	306.282	461.506	1192.872	796.134
45	479	14.3	20.368	42.758	42.854	340.684	412.144	1279.416	943.376
48	608	15.178	22.446	46.847	45.817	333.56	666.5	1483	1100
50	608	16.27	18.22	49	48.757	384.235	486	1679.355	1324.142

Theorem 2 *EDO algorithm results energy and delay optimized connected routing topology which minimizes maximum delay as well as node's power interval simultaneously on S-IDGM of a WSN.*

Proof We proceed the proof by considering two possible cases. Case (i) Maximum power consumption and maximum delay exist on distinct links: Let v_i be any node having *max PI* on the obtained topology T by EDO algorithm. This implies that there exists atleast one link of max power consumption incident on v_i, let l be that link. Let $l' \in T$ be another link, distinct from l, having maximum delay. Let us assume that T is not an energy and delay optimal topology obtained by EDO algorithm and T' is an optimal one where the cardinality of path cover obtained for T' must not be greater than the cardinality of path cover obtained for T's construction. Hence links l and l' which are having max power consumption and max delay respectively must not belong to T' (otherwise T would have been optimal). This means that a connected routing topology can be constructed using links whose power consumption and delay estimation are less than l and l' respectively. In that case EDO algorithm would have terminated before processing l and l' and this leads to contradiction. Hence T is the optimal topology.

Case (ii) Maximum power consumption and maximum delay lie on a same link:

Let l be the link of max power consumption and max delay on the topology T obtained by EDO algorithm, incident on node v_i having *max PI*. The proof carries forward by assuming the contradicting situation in the same way as we did in case(i) which finally results that T is the energy and delay optimal topology, obtained by EDO algorithm. □

Theorem 3 *Time complexity of EDO algorithm is $\mathcal{O}(n^2)$ where n is number of nodes.*

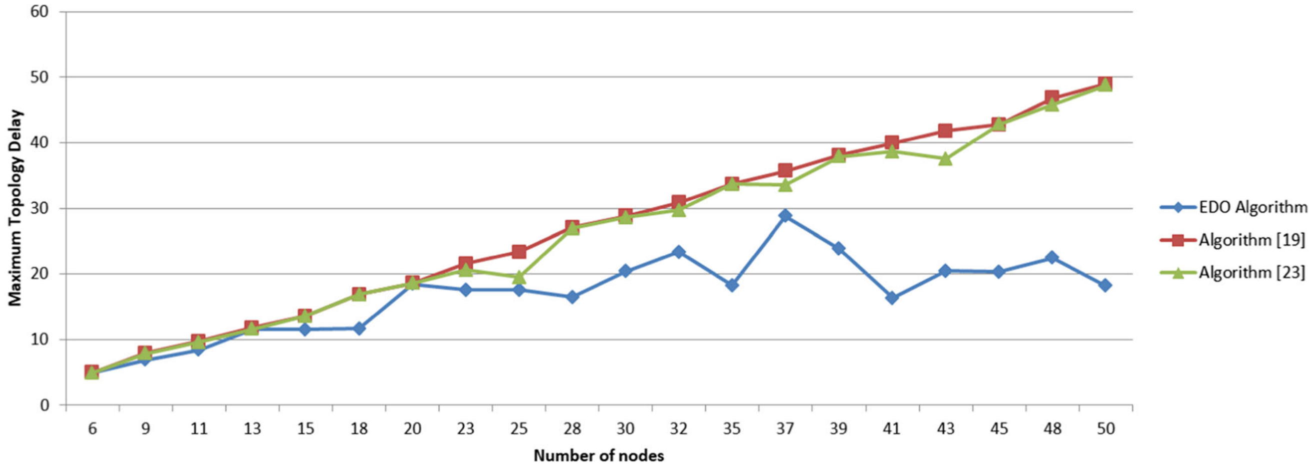


Fig. 9 Max Topology Delay with network density

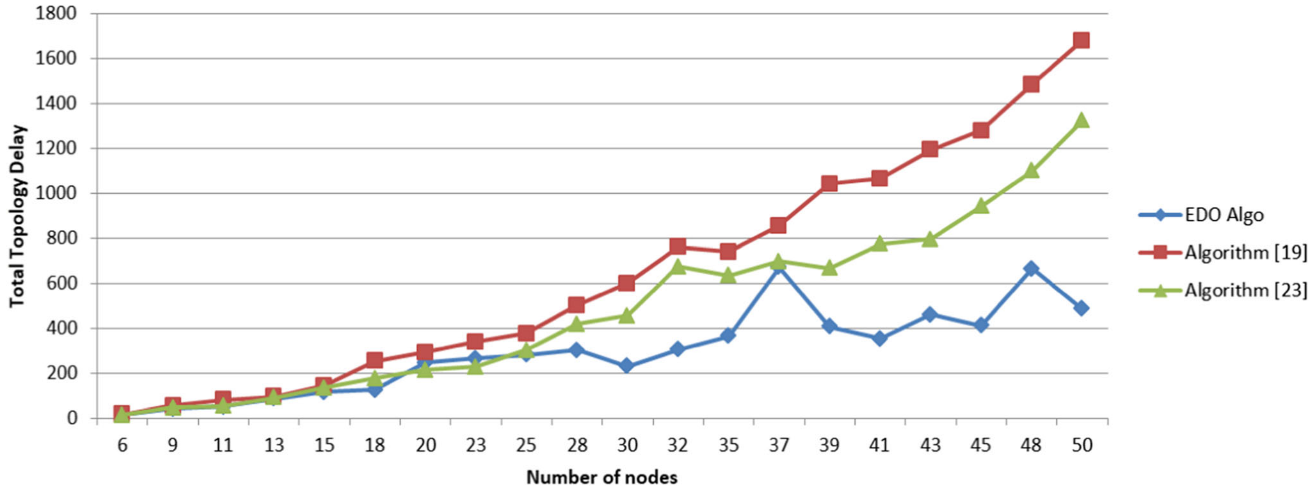


Fig. 10 Total Topology Delay with network density

Proof Step 1 takes $\mathcal{O}(n^2)$ to compute weighted graph. Step 2 also takes $\mathcal{O}(n^2)$ to obtain MST using Prim's algorithm. Step 4 takes $\mathcal{O}(n)$ and steps 3,5,6 are done in constant time. Step 7 initiates the first while loop and executes the steps from 8 to 16 atmost $\mathcal{O}(n)$ time. Step 8 of finding S and T takes $\mathcal{O}($ degree $(v_k))$ and maximum complexity from steps 9 to 14 can be $\mathcal{O}(n)$ since finding *LMV* and *RMV* of any component can take $\mathcal{O}(n)$. Steps 15 and 16 take constant time and here's the end of first while loop. Now step 18 is done in $\mathcal{O}(n)$ time and steps 19 to 21 take constant time. Step 22 initiates the second while loop and executes the steps from 23 to 29 atmost $\mathcal{O}(n)$. Steps 23 and 24 can take $\mathcal{O}(n^2)$ to find intersection and adjacency of two distinct components. Step 25 takes $\mathcal{O}(|V(C_i)|)+\mathcal{O}(|V(C_j)|)$ where $|V(C_i)|$ and $|V(C_j)|$ depict the number of nodes in components C_i and C_j respectively. Step 26 takes $\mathcal{O}(|V(C_j)|)$ and step 28 takes $\mathcal{O}(|V(C_i)|)$. Final steps 31 and 32 can be done in $\mathcal{O}(n^2)$. Therefore the maximum time complexity of EDO algorithm is $\mathcal{O}(n^2)$. □

5 Performance evaluation

Let us now evaluate the performance of EDO algorithm in term of topology delay minimization and energy efficiency by computing max node's power interval and percentage of energy saving and show the comparison with predefined algorithms, given in [23] and [19] on interval data based graph models of WSNs. Algorithm, given in [23], constructs a spanning tree of minimum number of pendant vertices by joining disjoint path component of path cover, obtained by Arikati and Rangan in [6] on an interval graph. The path cover obtained in [6] is called optimal in term of minimum cardinality (minimum number of path components) of the obtained path cover. Algorithm, given in [19] provides energy optimal path cover and energy efficient connected routing topology which minimizes max node's power interval and max link's transmission cost of worst case scenario. This routing topology [19] is more energy efficient than the spanning tree obtained in [23] on interval graphs.

Table 4 Data obtained on maximum node's power interval and % of energy saving after simulation

Number of nodes	Number of edges	Max node power interval				Percentage of energy saving			
		DE MST	EDO algorithm	Algorithm in [19]	Algorithm in [23]	DE MST	EDO algorithm	Algorithm in [19]	Algorithm in [23]
6	9	[155,312]	[155,230]	[155,230]	[161,312]	0	26	26	0
9	27	[16,133]	[42,116]	[16,116]	[22,116]	4	16	16	16
11	30	[81,169]	[103,169]	[55,131]	[103,193]	20.3	20.3	38	9
13	38	[123,176]	[72,162]	[72,162]	[121,236]	31	36	36	7
15	49	[68,227]	[9,189]	[9,246]	[93,277]	23	35.5	16	5.5
18	100	[87,226]	[87,207]	[47,190]	[84, 263]	21	28	34	8
20	120	[197,369]	[10,366]	[10,366]	[319,394]	8	8.5	8.5	2
23	129	[86,371]	[99,349]	[99,349]	[272,349]	7	12.5	12.5	12.5
25	141	[226,478]	[273,457]	[273,478]	[226,481]	1	5	1	0
28	202	[165,408]	[165,375]	[24,404]	[165,446]	9.5	16.9	10.4	1
30	268	[332,492]	[12,477]	[15,485]	[93,560]	18	20.5	19.2	6.7
32	291	[308,569]	[350,543]	[41,534]	[308,581]	6.9	11.13	12.6	5
35	307	[516,663]	[479,663]	[455,663]	[516,679]	5	5	5	2.6
37	304	[44,646]	[328,612]	[544,619]	[359,694]	9.3	14	13	2.5
39	433	[486,725]	[412,671]	[412,671]	[156,771]	6.2	13.2	13.2	0.3
41	461	[287,746]	[282,696]	[287,696]	[270,806]	8.13	14.3	14.3	0.71
43	436	[546,825]	[548,807]	[396,803]	[548,825]	4	6.1	6.5	4
45	479	[205,875]	[205,875]	[351,823]	[367,894]	2.2	2.2	8	0.1
48	608	[641,927]	[16,803]	[568,919]	[641,940]	3	16	4	1.7
50	608	[513,942]	[443,883]	[163,870]	[604,959]	5.5	11.4	12.7	4

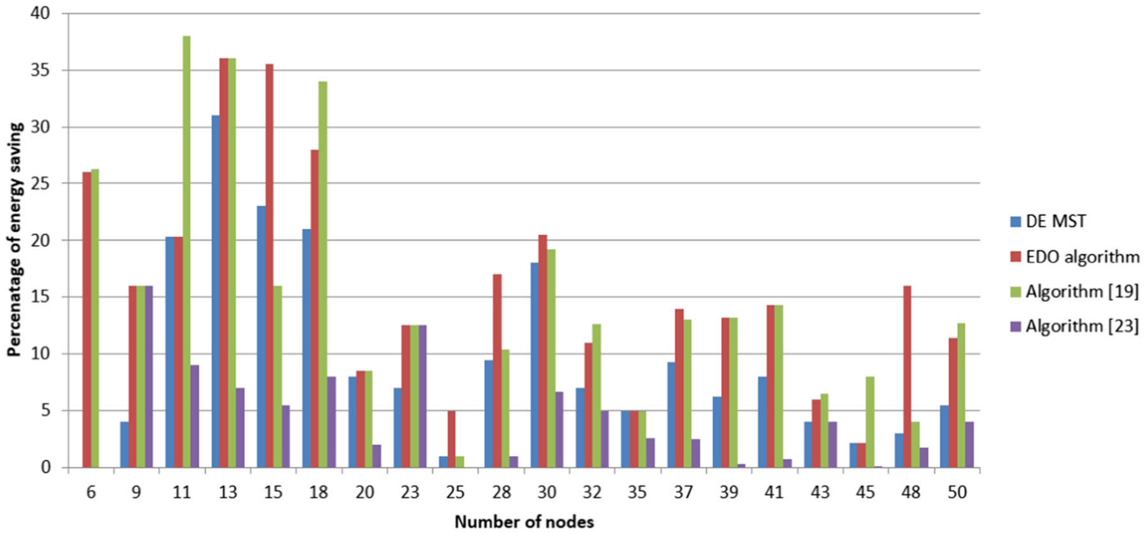


Fig. 11 Percentage of energy saving

Here we will show that EDO algorithm outperforms both the algorithms on S-IDGMs of WSNs in simultaneous minimization of topology delay and nodes' power consumption interval using simulation results, graphical plotting and statistical *t*-test tool. Using *t*-test, we show that topology formed by EDO algorithm has no significant difference in optimizing power consumption to save network energy, comparing with the energy efficient topology obtained by algorithm [19].

In addition, comparison of EDO algorithm with delay minimum spanning tree where the weight on each edge is computed using Eq. (4) and MST is obtained using Prim's algorithm, for conveniency we call it DE MST, is also shown. No doubt, this DE MST is the most optimal in minimizing max topology delay and total topology delay but through our evaluation we show that DE MST is not as much energy efficient as the topology formed by EDO algorithm.

5.1 Simulation environment

We run the simulation on large number of WSNs whose underlying graphs are interval graphs in Spyder (Python 3.7) to show the required comparison. Input of arbitrary interval graphs are randomly generated in term of number of nodes, number of edges, list of edges and list of all intervals in increasing order of their upper bounds. The data of each interval graph has been given to the python code of EDO algorithm to evaluate a connected routing topology and compared it with the topology formed by using algorithms given in [19, 23] and DE MST.

Simulation parameter $\alpha$of EDO algorithm The variable α when approaching to 1 gives more weightage to link level interference and α approaching to 0 gives more weightage to link distance by Eq. (3). After observing the relationship between link distance and link interference on distinct WSNs towards optimality in Sect. 3.4, α is fixed to a constant 0.5 in simulation work of EDO algorithm in order to consider both the metrics of delay estimation at equal weightage.

5.2 Metric 1: Topology delay minimization

The obtained data after simulation on maximum topology delay and total topology delay of the routing topology by each algorithm has been noted down in Table 3. It has been seen from the data given in Table 3 and their corresponding plotting in Fig. 9 that EDO algorithm performs better in minimizing maximum topology delay than Algorithm [19] and Algorithm [23]. Moreover, data plotting in Fig. 10 depicts that EDO algorithm performs better in minimizing total topology delay than Algorithm [19] in each case. Comparing with Algorithm [23], EDO algorithm minimizes more total topology delay in dense networks. Further, we use two-tailed *t*-testing to show that EDO algorithm optimizes topology delay at 2% level of significance, comparing with DE MST which has been optimal in minimizing topology delay.

*t -testing on the data given in*Table 3: *t*-test compares mean of data obtained by two different methodologies where results which are obtained on samples, are believed to exist on large number of networks (or we call it population of S-IDGMs).

To examine the performance of EDO algorithm with respect to DE MST through 'two tailed *t*-test', two types of hypothesis have to make. The first hypothesis is called null hypothesis comprises of a statement which we want to claim/proof and the other one is called alternate hypothesis whose statement is against the claim. Presently our claim for a null hypothesis is that the topology delay minimization by EDO algorithm is as optimal as DE MST. For this to test, null hypothesis becomes 'mean of data by EDO algorithm is equal to the mean of data by DE MST' and alternate hypothesis contradicts the null hypothesis and becomes 'mean of data by EDO algorithm is not equal to the mean of data by DE MST'.

Mean of max *TD* by DE MST is 13.844 and mean of total *TD* by DE MST is 187.175 on the networks, given in Table 3. Let μ_0 be the mean of max TD and μ_1 be the mean of total TD by EDO algorithm on population of S-IDGMs.

Null hypothesis H_0: $\mu_0 = 13.844$ and $\mu_1 = 187.175$.

Alternate hypothesis H_a: $\mu_0 \neq 13.844$ and $\mu_1 \neq 187.175$.

Now computing *t*-values for both the max *TD* and total *TD* using formula $t_0 = \frac{\overline{x_0}-\mu_0}{SD_0/\sqrt{n}}$ and $t_1 = \frac{\overline{x_1}-\mu_1}{SD_1/\sqrt{n}}$ respectively where n is number of sample networks, $\overline{x_0}$ is mean of max *TD*, SD_0 is standard deviation of max *TD* by EDO algorithm on the networks, given in Table 3. Likewise $\overline{x_1}$ is mean of total *TD* and SD_1 is standard deviation of total *TD* by EDO algorithm on the networks, given in Table 3.

Therefore $n = 20$, $\overline{x_0} = 16.89$, $\overline{x_1} = 294.978$, $SD_0 = 6.978$ and $SD_1 = 190.688$. Substituting these values in the formulas (given above) result $t_0 = 2.23395$ and $t_1 = 2.52827$. Hence the null hypothesis is accepted for both μ_0 and μ_1 with 2% level of significance. It can be said that topology delay minimization by EDO algorithm is as optimal as DE MST at 2% level of significance.

Remark *t*-values of Algorithm [19] and Algorithm [23] for max *TD* come-out to be 4.36 and 4.20 respectively, are significantly different from optimal value at 2% level of significance. Also *t*-values of Algorithm [19] and Algorithm [23] for total *TD* come out to be 4.0 and 3.536, are again significantly different from optimal value at 2% level of significance. Therefore overall performance of EDO algorithm in optimizing *TD* is much better than Algorithm [19] and Algorithm [23] on interval graph models of WSNs.

5.3 Metric 2: Energy efficient topology

The obtained data after simulation on max node's power interval and percentage of energy saving on the topology by each algorithm has been noted down in Table 4. It has been remarked already that percentage of energy saving is directly depending on max node's power interval; more is the upper bound of *PI* lesser is the energy saving. Figure 11 which shows the percentage of energy saving by each respective algorithm, also depicts that more energy saving by any algorithm implies lesser maximum power consumption and this can be seen in Table 4 as well.

Data in Table 4 and plotting in Fig. 11 are explicitly showing that EDO algorithm performs better in minimizing max node's power consumption than DE MST and algorithm, given in [23] in each case and so EDO algorithm conserves more energy than DE MST and algorithm, given in [23] on interval data based graph models of WSNs.

However, it has been observed that in some networks EDO algorithm conserves more energy than algorithm, given in [19] while in other networks EDO algorithm conserves equal and lesser energy than Algorithm [19]. This is because algorithm, proposed in [19] has been an energy optimal in providing connected routing topology on interval graphs without considering delay factors and EDO algorithm which works to obtain an energy and delay minimized connected routing topology, considers DE function depending on link interference and distance factor which helps in minimizing topology delay and power consumption simultaneously. Since tradeoff between energy consumption and delay minimization brings fluctuation in energy saving while minimizing topology delay. Therefore to evaluate the performance of EDO algorithm in compare to the Algorithm, given in [19], two-tailed $t - test$ has been used on the obtained data, given in Table 4 to show that topologies formed by EDO algorithm and algorithm, given in [19] have no significant difference in saving nodes' energy.

*t-testing on the data given in*Table 4: Considering the topology formed by Algorithm [19] is energy optimal and claiming that topology formed by EDO algorithm is also energy optimal while optimizing topology delay. So null hypothesis becomes 'Mean of percentage of energy saving by EDO algorithm is equal to the mean of percentage of energy saving by Algorithm [19]' and alternate hypothesis contradicts the null hypothesis and becomes 'Mean of percentage of energy saving by EDO algorithm is not equal to the mean of percentage of energy saving by Algorithm [19]'.

Mean of percentage of energy saving by Algorithm [19] is 15.36 on the networks, given in Table 4 and let μ_0 is the mean of percentage of energy saving by EDO algorithm on population of S-IDGM.

Let the null hypothesis H_0: $\mu_0 = 15.36$ and alternate hypothesis H_a: $\mu_0 \neq 15.36$.

Mean of percentage of energy saving by EDO algorithm on the networks, given in Table 4 is 15.92 i.e. $\bar{x} = 15.92$ and standard deviation is $SD = 9.53026$. Computing *t*-value using formula $t = \frac{\bar{x}-\mu_0}{SD/\sqrt{n}}$ where n is the number of sample networks, in this case $n = 20$. Substituting the values in the given formula results $t = 0.2628$. Hence the null hypothesis is accepted with 5% level of significance. This shows that there is no significant difference in minimizing maximum nodes' power interval and percentage of energy saving between EDO algorithm and Algorithm [19]. So it can be said that EDO algorithm results energy optimal topology while minimizing topology delay on S-IDGMs of WSNs.

Remark *t*-values of DE MST and Algorithm [23] for average percentage of energy saving come out to be -−3.0381 and -−11.12 respectively, are significantly different from optimal value at 5% level of significance. Therefore overall performance of EDO algorithm in optimizing max power consumption is much better than DE MST and Algorithm [23] on interval graph models of WSNs.

6 Conclusion

In this work, we have proposed an energy and delay optimizing polynomial time algorithm, named as EDO algorithm, to construct an energy and delay optimized connected routing topology on heterogenous bi-directional connected WSNs where nodes are allowed to vary their transmission range within their assigned power interval. A sorted interval data based graph model (S-IDGM) is introduced to show nodes power interval explicitly and proposed EDO algorithm runs on S-IDGM to result an optimized topology which minimizes max topology delay, total topology delay and max nodes' power interval. Minimization of max nodes' power interval consequently results efficient energy saving on S-IDGMs. Additionally, dependency analysis between link distance and link interference has been shown which helps to formulate a new weighted 'Delay estimation (DE)' function towards the optimality of EDO algorithm. Extensive simulation work, graphical and *t*-test analysis have been carried out to compare the performance of EDO algorithm with delay minimum spanning tree (DE MST) and algorithms [19, 23] on S-IDGMs of WSNs. The proposed EDO algorithm outperforms algorithm [19] and algorithm [23] in minimizing max and total topology delay while conserving more energy than algorithm [23] and DE MST. The *t*-test analysis has shown that EDO algorithm results an energy efficient topology on S-IDGM at 5% level of significance, comparing with algorithm [19] and delay minimization at 2% level of significance, comparing with DE MST.

In future, work on energy and delay optimization can be extended to dynamic WSNs where topologies are required to be more adaptable and stable with respect to node's movement. Moreover we look forward to develop new interference and delay models under different possible factors like network density, speed of node's movement and data collision in this direction.

Data Availability Authors can confirm that all relevant data are included in the article.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Ahmed, A. A. (2013). An enhanced real-time routing protocol with load distribution for mobile wireless sensor networks. *Computer Networks, 57*(6), 1459–1473.
2. Akkaya, K., & Younis, M. (2004). Energy-aware delay constrained routing in wireless sensor networks. *International Journal of Communication Systems, 17*(6), 663–687.
3. Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: A survey. *Computer Networks, 38*(4), 393–422.
4. Al Aghbari, Z., Khedr, A. M., Osamy, W., Arif, I., & Agrawal, D. P. (2020). Routing in wireless sensor networks using optimization techniques: A survey. *Wireless Personal Communications, 111*(4), 2407–2434.
5. Ammari, H. M., & Das, S. K. (2008). A trade-off between energy and delay in data dissemination for wireless sensor networks using transmission range slicing. *Computer Communications, 31*(9), 1687–1704.
6. Arikati, S. R., & Rangan, C. P. (1990). Linear algorithm for optimal path cover problem on interval graphs. *Information Processing Letters, 35*(3), 149–153.
7. Booth, K. S., & Lueker, G. S. (1976). Testing for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms. *Journal of Computer and System Sciences, 13*(3), 335–379.
8. Boughanmi, N., & Song, Y. (2008). A new routing metric for satisfying both energy and delay constraints in wireless sensor networks. *Journal of Signal Processing Systems, 51*(2), 137–143.
9. Bukhsh, M., Abdullah, S., Rahman, A., Asghar, M. N., Arshad, H., & Alabdulatif, A. (2021). An energy-aware, highly available, and fault-tolerant method for reliable IoT systems. *IEEE Access, 9*, 145363–145381.
10. Cheng, L., Niu, J., Luo, C., Shu, L., Kong, L., Zhao, Z., & Gu, Y. (2018). Towards minimum-delay and energy-efficient flooding in low-duty-cycle wireless sensor networks. *Computer Networks, 134*, 66–77.
11. Chincoli, M., Syed, A. A., Exarchakos, G., & Liotta, A. (2016). Power control in wireless sensor networks with variable interference. *Mobile Information Systems*. https://doi.org/10.1155/2016/3592581
12. Daoud, W. B., Mchergui, A., Moulahi, T., & Alabdulatif, A. (2022). Cloud-IoT resource management based on artificial intelligence for energy reduction. *Wireless Communications and Mobile Computing*. https://doi.org/10.1155/2022/2248962
13. Dutt, S., Agrawal, S., & Vig, R. (2021). Delay-sensitive, reliable, energy-efficient, adaptive and mobility-aware (dream) routing protocol for WSNS. *Wireless Personal Communications, 120*, 1675–1703.
14. Ebrahimi, D., Sharafeddine, S., Ho, P. H., & Assi, C. (2018). UAV-aided projection-based compressive data gathering in wireless sensor networks. *IEEE Internet of Things Journal, 6*(2), 1893–1905.

15. Ergen, S. C., & Varaiya, P. (2007). Energy efficient routing with delay guarantee for sensor networks. *Wireless Networks, 13*(5), 679–690.
16. Han, S.-W., Jeong, I.-S., & Kang, S.-H. (2013). Low latency and energy efficient routing tree for wireless sensor networks with multiple mobile sinks. *Journal of Network and Computer Applications, 36*(1), 156–166.
17. Hu, Y., Liu, D., Wu, Y. (2016). A new distributed topology control algorithm based on optimization of delay in ad hoc networks. In *2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI)*, pp. 148-152.
18. Huynh, T.-T., Dinh-Duc, A.-V., & Tran, C.-H. (2016). Delay-constrained energy-efficient cluster-based multi-hop routing in wireless sensor networks. *Journal of Communications and Networks, 18*(4), 580–588.
19. Kavra, R., Gupta, A., & Kansal, S. (2021). Interval graph based energy efficient routing scheme for a connected topology in wireless sensor networks. *Wireless Networks, 27*(8), 5085–5104.
20. Ketshabetswe, L. K., Zungeru, A. M., Mangwala, M., Chuma, J. M., & Sigweni, B. (2019). Communication protocols for wireless sensor networks: A survey and comparison. *Heliyon, 5*(5), e01591. https://doi.org/10.1016/j.heliyon.2019.e01591
21. Khalily-Dermany, M. (2021). Transmission power assignment in network-coding-based-multicast-wireless-sensor networks. *Computer Networks, 196*, 108203. https://doi.org/10.1016/j.comnet.2021.108203
22. Kim, B.-S., Park, H., Kim, K. H., Godfrey, D., & Kim, K.-I. (2017). A survey on real-time communications in wireless sensor networks. *Wireless Communications and Mobile Computing*. https://doi.org/10.1155/2017/1864847
23. Li, X., Feng, H., Jiang, H., Zhu, B. (2016). A polynomial time algorithm for finding a spanning tree with maximum number of internal vertices on interval graphs. *International Workshop on Frontiers in Algorithmics*, 92-101.
24. Li, Y., Chen, C. S., Song, Y.-Q., Wang, Z., & Sun, Y. (2009). Enhancing real-time delivery in wireless sensor networks with two-hop information. *IEEE Transactions on Industrial Informatics, 5*(2), 113–122.
25. Majid, M., Habib, S., Javed, A. R., et al. (2022). Applications of wireless sensor networks and internet of things frameworks in the industry revolution 4.0: A systematic literature review. *Sensors, 22*(6), 2087.
26. Moaveninejad, K., & Li, X.-Y. (2005). Low interference topology control for wireless ad hoc networks. *Ad Hoc & Sensor Wireless Networks, 1*(1–2), 41–64.
27. Noueihed, H., Harb, H., & Tekli, J. (2022). Knowledge-based virtual outdoor weather event simulator using unity 3D. *The Journal of Supercomputing, 78*(8), 10620–10655.
28. Panda, B., & Shetty, D. P. (2013). Minimum interference strong bidirectional topology for wireless sensor networks. *International Journal of Ad Hoc and Ubiquitous Computing, 13*(3–4), 243–253.
29. Pantazis, N. A., Nikolidakis, S. A., & Vergados, D. D. (2013). Energy-efficient routing protocols in wireless sensor networks: A survey. *IEEE Communications Surveys and Tutorials, 15*(2), 551–591.
30. Pothuri, P.K., Sarangan, V., Thomas, J.P. (2006). Delay-constrained, energy-efficient routing in wireless sensor networks through topology control. In *2006 IEEE International Conference on Networking, Sensing and Control*, pp. 35-41.
31. Rachamalla, S., & Kancherla, A. S. (2016). A two hop based adaptive routing protocol for real-time wireless sensor networks. *SpringerPlus, 5*(1), 1–12.
32. Ramani, S. V., & Jhaveri, R. H. (2022). SDN framework for mitigating time-based delay attack. *Journal of Circuits, Systems and Computers, 31*(15), 2250264. https://doi.org/10.1142/S0218126622502644
33. Ramani, S., & Jhaveri, R. H. (2022). ML-Based delay attack detection and isolation for fault-tolerant software-defined industrial networks. *Sensors, 22*, 6958. https://doi.org/10.3390/s22186958
34. Roy, A., Pachuau, J. L., & Saha, A. K. (2021). An overview of queuing delay and various delay based algorithms in networks. *Computing, 103*, 2361–2399.
35. Santi, P. (2005). Topology control in wireless ad hoc and sensor networks. *ACM Computing Surveys (CSUR), 37*(2), 164–194.
36. Selvi, M., Velvizhy, P., Ganapathy, S., Nehemiah, H. K., & Kannan, A. (2019). A rule based delay constrained energy efficient routing technique for wireless sensor networks. *Cluster Computing, 22*(5), 10839–10848.
37. Shahid, J., Muhammad, Z., Iqbal, Z., Almadhor, A. S., & Javed, A. R. (2022). Cellular automata trust-based energy drainage attack detection and prevention in wireless sensor networks. *Computer Communications, 191*, 360–367. https://doi.org/10.1016/j.comcom.2022.05.011
38. Singla, P., & Munjal, A. (2020). Topology control algorithms for wireless sensor networks: A review. *Wireless Personal Communications, 113*, 2363–2385.
39. Sun, G., Zhao, L., Chen, Z., & Qiao, G. (2015). Effective link interference model in topology control of wireless ad hoc and sensor networks. *Journal of Network and Computer Applications, 52*, 69–78.
40. Uysal-Biyikoglu, E., Prabhakar, B., & El Gamal, A. (2002). Energy-efficient packet transmission over a wireless link. *IEEE/ACM Transactions on Networking, 10*(4), 487–499.
41. Von Rickenbach, P., Wattenhofer, R., & Zollinger, A. (2009). Algorithmic models of interference in wireless ad hoc and sensor networks. *IEEE/ACM Transactions on Networking, 17*(1), 172–185.
42. Wang, F., Liu, W., Wang, T., Zhao, M., Xie, M., Song, H., Li, X., & Liu, A. (2019). To reduce delay, energy consumption and collision through optimization duty-cycle and size of forwarding node set in wsns. *IEEE Access, 7*, 55983–56015.
43. Xu, M., Yang, Q., & Shen, Z. (2017). Joint design of routing and power control over unreliable links in multi-hop wireless networks with energy-delay tradeoff. *IEEE Sensors Journal, 17*(23), 8008–8020.
44. Yick, J., Mukherjee, B., & Ghosal, D. (2008). Wireless sensor network survey. *Computer Networks, 52*(12), 2292–2330.
45. Zhang, R., Berder, O., Gorce, J.-M., & Sentieys, O. (2012). Energy-delay tradeoff in wireless multihop networks with unreliable links. *Ad Hoc Networks, 10*(7), 1306–1321.

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

Radhika Kavra is currentlypursuing Ph.D. from Departmentof Applied Mathematics,Delhi Technological University,New Delhi, India. She has doneM.Sc. Mathematics from IITRoorkee and B.Sc. (Hons)Mathematics from DelhiUniversity. Her research area isGraph and Optimization.

Anjana Gupta is a Professor inthe Department of Applied Mathematics, Delhi TechnologicalUniversity, New Delhi,India. She received her Ph.D.degree from MathematicsDepartment, Delhi University.Her field of specializationincludes Optimization Techniques,Fuzzy Logics, MulticriteriaGroup Decision MakingProblems. She has more than 24years of teaching experienceand more than 50 researchpublications in well-reputedInternational and National Journals and many in Internationalconferences..

Dr. Sangita Kansal is a Professor inthe Department of AppliedMathematics, Delhi TechnologicalUniversity, New Delhi,India. She received her Ph.D.degree from IIT Delhi. Her fieldof specialization includes Graphtheory and Petri nets. She hasmore than 28 years of teachingexperience and more than 32years of research experience.She has research publications inwell-reputed International andNational Journals.

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-14816-0

Optimized ensemble-classification for prediction of soil liquefaction with improved features

Nerusupalli Dinesh Kumar Reddy[1] · Ashok Kumar Gupta[1] · Anil Kumar Sahu[1]

Received: 4 January 2022 / Revised: 22 June 2022 / Accepted: 6 February 2023

Abstract
The occurrence of soil liquefaction is an interesting and complicated field in the geotechnical earthquake, which has attained the consideration of a lot of analysts in current years. Liquefaction is a process, where the stiffness and strength of soil are minimized by sudden cyclic loading or earthquakes. Liquefaction and associated phenomenon were accountable for the massive quantity of damages during earlier earthquakes around the globe. Here, pre-processing is done with data normalization. Subsequently, the features including "statistical and raw features, higher-order statistical features, and improved entropy and Mutual Information (MI) features" are derived. Further, ensemble classifiers like "Deep Belief Network (DBN), Long Short Term Memory (LSTM), and Recurrent Neural Network (RNN)" are deployed during prediction. Here, the outputs obtained from DBN and LSTM are fused and then given to optimized RNN, which provides the final predicted output. Particularly, the weights of RNN are fine-tuned by Opposition based Self Adaptive SSO (OSA-SSO) model. Eventually, the advantage of the adopted model is proven on diverse metrics. The accuracy of the developed approach was 9.09%, 8.08%, and 10.1% higher than the values obtained for traditional schemes such as EC + SSO, EC + SSA, EC + PRO, and EC + BOA at the 90th LP, respectively.

Keywords Soil liquefaction · Statistical features · Ensemble classifiers · Optimized RNN · OSA-SSO algorithm

1 Introduction

Currently, the forecast of soil liquefaction is proven to be a most complicated and motivating issue in geotechnical earthquakes owing to the complexity and uncertainties of numerous factors [1, 21, 45]. The most important cause of damage throughout an earthquake is land crack. It might occur owing to the existence of cracks, anomalous settlement, gaps, and

✉ Nerusupalli Dinesh Kumar Reddy
nerusupallidineshkumarreddy@gmail.com

[1] Department of Civil engineering, Delhi Technological University, Delhi, India

Published online: 07 March 2023

inequality or loss of stiffness and strength in reaction to an unexpected transformation in the stress situations like an earthquake, leading the soil to perform like a fluid. [17, 52]. At this state, movable sand has a propensity to compact and have a propensity to expand. While soil is liquefied, constructing a structure on soil turns out to be unsteady. Even though numerous studies were performed for evaluating the potential of liquefaction, the complication of this normal occurrence points out the requirement for widespread research to measure the potential of its happening [34].

As the soil liquefaction is said to be the most important reason for natural disasters and engineering failure, numerous researchers were attempting to forecast and assess the liquefaction potential of a certain situation earlier, as a result, that the pertinent loss could be alleviated [9, 12]. The basic process of evaluating the Cyclic Shear Ratio (CSR) with Cyclic Resistance Ratio (CRR) is a landmark in introducing the forecast model of soil liquefaction [49, 51, 25]. This process is constantly depending upon the in situ analysis that mostly contains the SPT and CPT [49, 44, 30]. Owing to several benefits, various analysts were developing and proposing the experimental prediction approach depending on CPT data.

In recent times, the exploitation of ML schemes such as Artificial Neural Network (ANN) [18, 13, 14], Decision Tree (DT) and Bayesian Belief Network (BBN), etc. for predicting the liquefaction of soil is proved to be more rapid than conventional techniques and appropriate for accurately predicting nonlinear issues. The development of intelligent video surveillance had a significant impact on multimedia surveillance systems throughout the last decade [22]. Videos are audiovisual material that has been captured or recorded in chronological order to reveal the live states of events or activities [24, 23]. Delta, Event bagging: A unique event summarizing approach in multiview surveillance footage, Deep event learning boost-up approach [6]. Processing numerous pictures to build a 3D volume object takes a long time [27]. Recommender Systems are programs that make recommendations to users based on their preferences [5]. The derivation and definition of requirements focus on gathering requirements from various stakeholders. In addition, Synthetic Neural Networks (SNN) or ANNs or Neural Network (NN), are novel computational methods and systems for ML, knowledge expression, and eventually deployed knowledge acquirement to increase the outputs of intricate systems [2, 26, 8, 46, 38, 49]. Soil liquefaction is a major difficulty in geotechnical engineering. Soil liquefaction prediction is still a work in progress, and increasingly researches are going on creating ways to anticipate liquefaction soil. But this research work overcomes certain drawbacks such as the conditional probability being improved. In addition, it includes large datasets and large data samples. The computational cost is less and the liquefaction risks are reduced. The simplicity of the approach and the greatest prediction accuracy are the driving forces behind this research.

Contributions:

- Introduces a novel soil liquefaction prediction model, where improved entropy features along with other features are derived.
- For precise prediction outputs, ensemble classifiers with optimized RNN are used.
- Here, OSA-SSO model is introduced to choose optimal RNN weights.

Here, Section 2 and 3 reviews and interprets the proposed soil liquefaction prediction approach. Section 4 portrays the extraction of proposed features. Section 5 depicts ensemble classifiers with OSA-SSO optimization for prediction. Sections 6 and 7 depict resultants and conclusions.

2 Literature review

2.1 Related works

In 2021, Hu et al. [15] developed 2 novel BN approaches for forecasting gravelly soil liquefaction using a novel hybrid model that combined the maximum domain knowledge and information coefficient depending upon the shear wave velocity test and dynamic penetration test datasets. The proposed hybridised model's performance was validated by comparing it to other existing schemes, and two unique BN techniques outperformed other schemes when compared to traditional methods or models for predicting liquefaction of gravelly soil.

In 2018, Jilei et al. [29] investigated 31 aspirant intensity metrics by the analysis of sufficiency, correlation, proficiency, and efficiency depending on a larger dataset of past soil motion records. Here, 2 novel BN approaches were introduced using the recognized intensity metrics by merging the calculated liquefaction-associated information and the previous information of soil liquefaction depending upon a larger database of SPT analysis. The resultants revealed that the RMSA was finer for evaluating the potential of liquefaction.

In 2021, Zhang et al. [47] has deployed Standard Penetration Test (SPT) data and GWO model for improving the accuracy of the SVM-based prediction scheme. The optimal value of SVM constraints was first computed and chosen using iteration of Grey Wolf Optimization (GWO); then, the chosen constraints were given to SVM to train the prediction technique with the training data. Finally, by studying the test set and computing the performances of the trained scheme until the accuracy goal was attained, the major constraint of GWO was evaluated and revised.

In 2021, Zhang et al. [48] developed an Multi-Layer Fully Connected Network (ML-FCN) for optimizing the Deep Neural Network (DNN) and the prediction scheme was trained depending on Vs and SPT dataset. "The record database was separated into a testing set, a validation set, and a training set by a ratio of 2:2:6 for improved assessment. The SPT database was derived for training a related DNN prediction scheme". The ML-FCN-trained method predicted the possibility for liquefaction with better accuracy based on evaluation results.

In 2019, Rahbarzare et al. [34] used hybrid Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) with an Fuzzy Support Vector Machine (FSVM) for predicting soil liquefaction crisis. Fuzzy logic was deployed for decreasing the outlier sensitivity by computing the significance of every sample for increasing the capability of the classifier's simplification. By means of the suitable amalgamation of optimization schemes, the finest constraints were discovered throughout the training stage by the customer owing to the higher accurateness of the classifier.

In 2021, Alizadeh et al. [3] proposed a new prediction model depending upon ANN for predicting the liquefaction potential sufficiently in a specified range of soil nature. In addition, a whole set of 100 soil data was acquired in Iran for the purpose of determining the liquefaction potential using experiential methodologies. Following that, the experimental schemes' outputs were used to train data in an artificial neural network (ANN), which was then tested as an alternative to forecasting liquefaction in Iran. The ANN formation that was achieved was used to forecast liquefaction for assessment.

In 2021, Ghani et al. [11] evaluated the impact of flexibility on liquefaction activities of fine-grained soil in areas of Bihar by developing a formula depending upon Multi-Linear Regression (MLR) examination for forecasting security in opposition to liquefaction. The analysis' conclusions were aided by reliability analysis. The suggested scheme's dependability has been demonstrated by substantiating the results using actual liquefaction data obtained from non-liquefied and liquefied quake locations.

In 2021, Zhang et al. [49] considered the advantages of CPT over SPT and the appropriateness for handling the nonlinear issues of ELM. Here, ELM was trained for prediction. Initially, 7 prediction constraints were determined and analyzed; subsequently, 226 CPT samples were split into testing and training sets; then the constraint of the ELM scheme was guaranteed by evaluating the training speed and accuracy of the scheme while setting the count of the neuron.

2.2 Review

Table 1 reviews existing systems. Initially, BN was deployed in [15] which offers enhanced accuracy and it also provides a minimal error. BN scheme was exploited in [29] that offered higher accuracy with high recall; however, it requires representation of liquefaction potential. SVM was introduced in [47], which offers superior accuracy and it speeds up the operating rate. But larger datasets were not concerned. Likewise, the DNN model was exploited in [48], which minimizes loss and it offers high accuracy. Moreover, FSVM was exploited in [34] to offer high sensitivity and better classification accuracy, however; feature selection options are not exploited. ANN model was deployed in [3], which offers maximal reliability and it reduces error values; however, cost factors are not analyzed. MLR was suggested in [11] that is highly accurate and it also presents fewer errors. Finally, ELM was used in [49] that offered improved accurateness and it minimizes operating time, however, it needs analysis on non-liquefied cases.

3 A stepwise interpretation of the proposed liquefaction of soil approach

The implemented approach encompasses below stages.

- Initially, data normalization is deployed at pre-processing.
- Then, the features including "statistical features (Fe_{ST}), higher-order statistical features (Fe_{HO}), raw features (Fe_{RF}), improved entropy (Fe_{IEn}), and MI-based features" are derived.

Table 1 Review on soil liquefaction prediction schemes

Author	Adopted Method	Features	Challenges
Hu et al. [15]	BN	• High accuracy • Less error	• Conditional possibility is not concerned
Jilei et al. [29]	BN model	• Superior accuracy • High recall	• Need representation on liquefaction potential
Zhang et al. [47]	SVM+GWO	• High accuracy • Speediness in operating time	• Larger datasets is not concerned
Zhang et al. [48]	DNN	• Less loss High accurateness	• Enlargement of data scale is not concerned
Rahbarzare et al. [34]	FSVM	• Maximal classification accuracy • High sensitivity	• Feature selection options are not exploited
Alizadeh et al. [3]	ANN	• Reduced error values • Better reliability	• Needs analysis of cost factors
Ghani et al. [11]	MLR	• Fewer error • High accuracy	• Liquefaction risks needs focus
Zhang et al. [49]	ELM	• Better accuracy • Minimal operating time	• Need analysis on non-liquefied cases

- The derived features are supplied to LSTM and DBN frameworks for classification. The resulting classified output is further classified via optimized RNN (weight optimization via OSA-SSO) that provides the final classified output.

Figure 1 demonstrates OSA-SSO-based model.

3.1 Pre-processing via data normalization

Data normalization is used to pre-process input data (***In***). This technique generally protected data and database was made reliable by inconsistent dependency and get rid of duplication.

The pre-processed data ($\boldsymbol{In}_{DN}$) is provided for feature extraction.

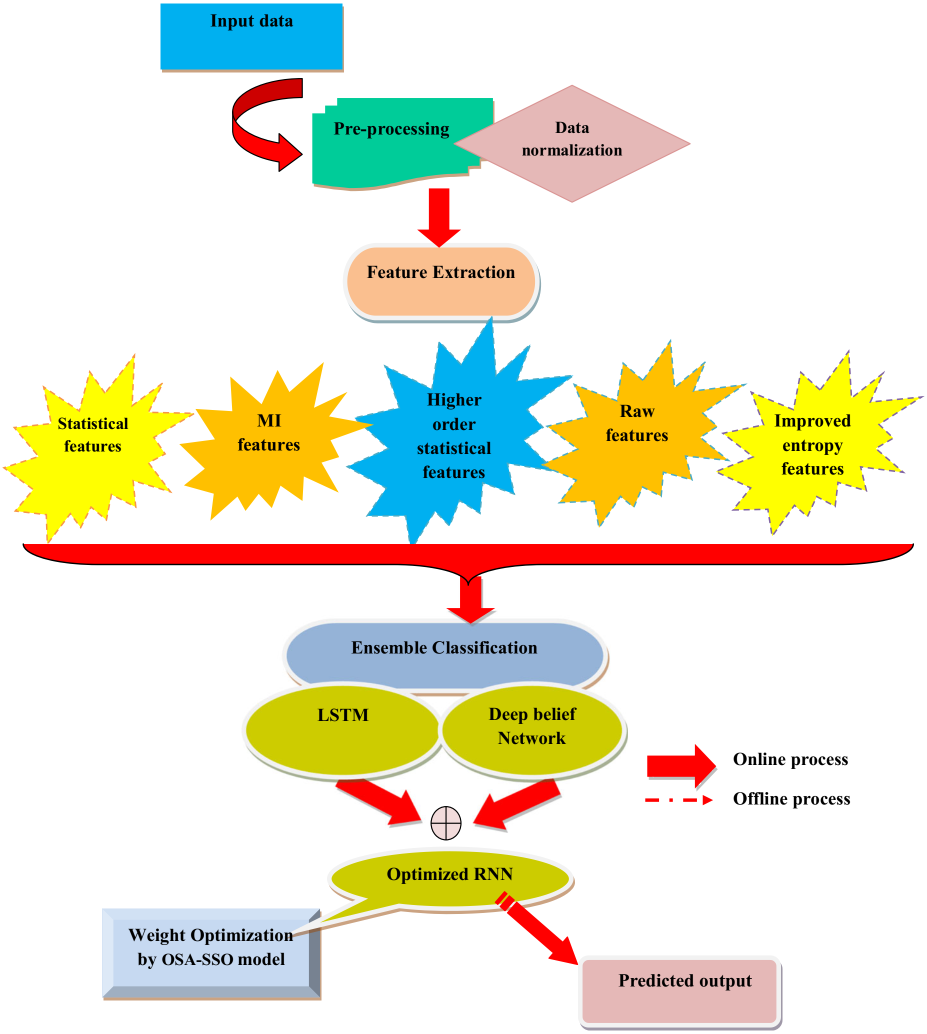


Fig. 1 Illustrative demonstration for developed model

4 Extraction of proposed features

4.1 Extracting proposed features

From input data, the “statistical features indicated by Fe_{ST} (mean, median, min, max, standard deviation), higher-order features (variance, kurtosis, skewness), improved EMA, MI, and improved correlation features” are extracted.

a) Skewness (https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm#:~:text=Skewness%20is%20a%20measure%20of,relative%20to%20a%20normal%20distribution): “It is a symmetry measure or the lack of symmetry exactly. A dataset or distribution is symmetric only if it is similar to the left and right of the center point”. The arithmetic term of skewness $SF_1$is specified in Eq. (1).

$$SF_1 = \frac{\sum_{i=1}^{k}(Y_i - \mu)^3/k}{T^3} \tag{1}$$

In Eq. (1), “$Y_i = Y_1, Y_2, ..., Y_k$, μ implies mean value, T implies standard deviation and k implies a number of data points. Furthermore, T is computed with k in the denominator rather than $k - 1$ whilst computing the skewness”.

b) Kurtosis (https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm#:~:text=Skewness%20is%20a%20measure%20of,relative%20to%20a%20normal%20distribution): “It is a measure that identifies whether the data are light-tailed or heavy-tailed and related to the normal distribution”. Datasets with smaller kurtosis [38] offer minor outliers or tails. The arithmetic method of kurtosis SF_2 is articulated in Eq. (2).

$$SF_2 = \frac{\sum_{i=1}^{k}\left(Y_i - \overline{Y}\right)^4/k}{T^4} \tag{2}$$

c) Variance (http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0001-37652013000301063): It is defined as the mean squared disparity among every data point and the middle of distribution computed by mean. The variance features are signified as SF_3.

The derived higher-order features are entirely indicated by $SF_1 + SF_2 + SF_3 = Fe_{HO}$.

d) MI Features: It is defined as the calculation of exchanged information among two ensembles of random variables N and Z [40]. It is exposed in Eq. (3), ρ ➔ probability.

$$MI = \sum \rho(N, Z) \log_2 \frac{\rho(N, Z)}{\rho(Z).\rho(N)} \tag{3}$$

The extracted MI features are represented as Fe_{MI}.

e) Improved Entropy Features: It is a statistical measurement of arbitrariness, which assists in characterizing the texture (http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0001-37652013000301063). Entropy features can be successfully used to evaluate the structural properties. Thus indicating that entropy may be a good predictor of structural deterioration. Traditionally, entropy (En) is assessed as exposed in Eq. (4), however, as per the developed concept; the proposed entropy Fe_{IEn} is computed based on weight as shown in Eq. (5), where, $log(L^\wedge m)$ refers to the maximal value of Fe_{IEn}, L refers to discretization level and ***we*** refers to weight factor produced using the chaotic map between [0, 1].

$$Fe_{En} = \left[-\sum_{i=1}^{n} P_i log(P_i)\right] \tag{4}$$

$$Fe_{IEn} = \frac{\left[-\sum_{i=1}^{n} P_i \log(P_i)\right]}{\log(L^\wedge m)} {*we} \tag{5}$$

The derived improved entropy features are indicated by Fe_{IEn}.

Accordingly, the derived features are summed up andFe, i.e. $Fe = Fe_{ST} + Fe_{HO} + Fe_{RF} + Fe_{MI} + Fe_{IEn}$ are then provided to ensemble classifiers for prediction.

5 Ensemble classifiers with OSA-SSO optimization for prediction

5.1 Ensemble classifiers

a) LSTM classifier: It [50] includes a sequence of recurring LSTM cells. Each cell of LSTM encompassed 3 units, such as “forget gate, the input gate, and the output gate”. Assume M and D as hidden and cell states. (M_t, D_t) and (X_t, D_{t-1}, M_{t-1})➔ output and input layer.

At the “time t, the output, input and forget gate implies O_t, I_t, F_t”. LSTM is primarily employs F_t for sorting the data to ignore. F_t is formulated as specified in Eq. (6).

$$F_t = \sigma(J_{\boldsymbol{IF}} X_t + B_{\boldsymbol{IF}} + J_{MF} M_{t\text{-}1} + B_{MF}) \tag{6}$$

In Eq. (6),” (J_{MF}, B_{MF}) and $(J_{\boldsymbol{IF}}, B_{\boldsymbol{IF}})$ imply weight and bias constraint to map hidden and input layers to forget gate and activation function is implied byσ.

Input gate is exploited by LSTM as shown in Eq. (7) - Eq. (9), where, (J_{MG}, B_{MG}) and (J_{IG}, B_{IG}) imply weight and bias constraint to map hidden and input layers to cell gate correspondingly. (J_{MI}, B_{MI}) and (J_{II}, B_{II}) imply weight and bias constraint to map hidden and input layers to I_t”.

$$G_t = tanh(J_{IG} X_t + B_{IG} + J_{MG} M_{t\text{-}1} + B_{MG}) \tag{7}$$

$$I_t = \sigma(J_{II} X_t + B_{II} + J_{MI} M_{t\text{-}1} + B_{MI}) \tag{8}$$

$$D_t = F_t D_{t-1} + I_t G_t \tag{9}$$

$$O_t = \sigma(J_{IO} X_t + B_{IO} + J_{MO} M_{t-1} + B_{MO}) \tag{10}$$

$$M_t = O_t tanh(D_t) \tag{11}$$

Further, the LSTM cell obtains the output hidden layer from the output gate as revealed in Eq. (10) and Eq. (11), wherein, (J_{MO}, B_{MO}) & (J_{IO}, B_{IO}) implies weight and bias to map the hidden and input layer to O_t.

b) DBN classifier: DBN [43] contain varied layers involving visible as well as hidden neurons and is specified in Eq. (14). The output denoted by $\overline{PO}$ is modeled in Eq. (13) and possibility function $\overline{P}_q(\zeta)$ is modeled in Eq. (12), wherein, pseudo-temperature is symbolized by t^P.

$$\overline{P}_q(\zeta) = \frac{1}{1 + e^{\frac{-\zeta}{t^P}}} \tag{12}$$

$$\overline{PO} = \left\{ 1 \quad \boldsymbol{with}\ 1 - \overline{P}_q(\zeta) 0 \boldsymbol{with} \overline{P}_q(\zeta) \right\} \tag{13}$$

$$\lim_{t^P \to 0^+} \overline{P}_q(\zeta) = \lim_{t^P \to 0^+} \frac{1}{1 + e^{\frac{-\zeta}{t^P}}} = \begin{cases} 0 & for \quad \zeta < 0 \\ \frac{1}{2} & for \quad \zeta = 0 \\ 1 & for \quad \zeta > 0 \end{cases} \tag{14}$$

The binary state bi is revealed in Eq. (15) and (16), wherein "θ_a implies biases and $L_{a,\,l}$ implies weights amid neurons".

$$EN(bi) = -\sum_a bi_a \theta_a - \sum_{a<l} bi_a L_{a,l} bi_l \tag{15}$$

$$\Delta EN(bi_a) = \sum_l \theta_a + L_{a,l} bi_l \tag{16}$$

The energy for visible and hidden neuron (x, y) is in Eq. (17)–(19), "wherein y_l and x_a implies a binary state of the hidden unit l and visible unit a, k_a and C_l implies biases".

$$EN(x,y) = -\sum_{(a,l)} L_{a,l} x_a y_l - \sum_a k_a x_a - \sum_l C_l y_l \tag{17}$$

$$\Delta EN\left(x_a, \overline{y}\right) = \sum_l L_{al} y_l + k_a \tag{18}$$

$$\Delta EN\left(\vec{x}, y_l\right) = \sum_a L_{al} x_a + C_l \tag{19}$$

The allotment of weight is shown in Eq. (20).

$$\mathrm{M}_{\left(\widehat{m}\right)} = \max_{\mathrm{M}} \prod_{\vec{x}\in N} c\left(\vec{x}\right) \tag{20}$$

The RBM energy function is in Eq. (21), PR^F implies partition terms as in Eq. (22).

$$c\left(\vec{x}, \vec{y}\right) = \frac{1}{PR^F} e^{-EN\left(\vec{x}, \vec{y}\right)} \tag{21}$$

$$PR^F = \sum_{\vec{x}, \vec{y}} e^{-EN\left(\vec{x}, \vec{y}\right)} \tag{22}$$

Let training pattern be $\left(K^{\widehat{H}}, U^{\widehat{H}}\right)$, wherein $K^{\widehat{H}}$ and $U^{\widehat{H}}$ indicate input and output vector, and $1 \le \widehat{H} \le V$, V point out training pattern count. All neuron errors at the output are delineated by Eq. (27). Subsequently, the square error of $\widehat{H}$ pattern is specified in Eq. (28).

$$e_l^{\widehat{H}} = K^{\widehat{H}} - U^{\widehat{H}} \tag{27}$$

$$SE_{\widehat{H}} = \frac{1}{\widetilde{o}_y} \sum_{l=1}^{\widetilde{o}_y} \left(e_l^{\widehat{H}}\right)^2 = \frac{1}{\widetilde{o}_y} \sum_{l=1}^{\widetilde{o}_y} \left(K^{\widehat{H}} - U^{\widehat{H}}\right)^2 \tag{28}$$

$$SE_{avg} = \frac{1}{V} \sum_{\widehat{H}=1}^{V} SE_{\widehat{H}} \tag{29}$$

The DBN and LSTM outputs are then classified via optimized RNN for absolute classification.

c) Optimized RNN classifier: This classifier [20] facilitates the affiliation amid neurons to form a phase; accordingly, the data at a time t is taken whilst the input data is communicated from time t to $t + 1$.

RNN enclose "output layer, a hidden layer, and an input layer with varied neurons. The input layer encompasses N input units (vector sequence) from time$t$, such as $\{\ldots., q_{t-1}, q_t, q_{t+1}, \ldots\}$, in which $q_t = (q_1, q_2, \ldots . q_N)$. Every input unit is correlated with each hidden unit in the hidden layer, whose links are described by weight matrix W_{Ih}. The hidden layers contain M hidden units $\mathrm{A}_t = (\mathrm{A}_1, \mathrm{A}_2, \ldots . \mathrm{A}_M)$ that are connected to one another by means of recurrent links offered by matrix W_{hh}". The hidden layer is shown by Eq. (30), wherein $\theta_h(.)$ and a_{A} ➔ activation function and hidden bias vector.

$$\mathrm{A}_t = \theta_h(W_{Ih}q_t + W_{hh}\mathrm{A}_{t\text{-}1} + a_\mathrm{A}) \tag{30}$$

The output layers involve$q$units that is exposed in Eq. (31), “wherein $\theta_o(.)$ and a_o signifies the activation function and bias vector of units at the output layer, W_{ho} signifies weight matrix”.

$$x_t = \theta_o(W_{ho}\mathrm{A}_t + a_o) \tag{31}$$

d) Objective: The objective *Obj* is to diminish the error *Err* as revealed in Eq. (32).

$$Obj = min(Err) \tag{32}$$

e) RNN Hyperparameters

Some of the hyperparameters of RNN are provided in Table 2. The number of layers, the activation functions are provided in the Table.

Solution encoding As said above, the RNN weights signified by (W) are optimally chosen via the OSA-SSO. Figure 2 shows the solutions, wherein, *Nl* ➔ entire RNN weight count.

5.2 OSA-SSO algorithm

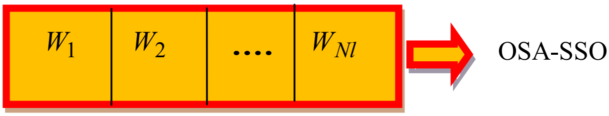
The shark’s sense of smell can be used as a guide. This method aids the shark in locating the source of the smell. The migration of the shark toward the source of the smell. Concentration is crucial in guiding the shark to its prey during this action. To put it another way, a larger concentration causes the shark to move. This property serves as the foundation for the creation of an optimization algorithm for finding the best solution to a problem.

Though the extant SSO [32] model includes many benefits; it endures varied limits like premature convergence etc. Hence, precise modifications were essential and OSA-SSO is developed. Generally, self-enhancement is established to be capable in conventional optimization schemes [35, 36, 41, 10, 37, 42, 39, 16]. The steps followed in the proposed OSA-SSO are as follows.

Table 2 RNN Hyperparameters

S.No	Parameters
1.	Number of Layers = 54
2.	Activation Function = ReLu
3.	Activation Function = Sigmoid
4.	Loss = Sparse categorical cross-entropy
5.	Optimizer = rmsprop
6.	Reshape (1,64)
7.	Epochs = epoch
8.	Batch Size = 10

Fig. 2 Solution encoding



OSA-SSO includes, "initialization, forward movement, rotational movement, and position update".

Initialization The preliminary solution is depicted in Eq. (33) and (34), in which, $Z_i^1 = i^{th}$ initial populace vector position and *np* implies populace size. In addition, OBL-based solutions are created, that is, opposite solutions are generated in SSO.

$$Z^1 = \left[Z_1^1, Z_2^1, \ldots . Z_{np}^1\right] \tag{33}$$

The related issue of optimization is shown in Eq. (34), in which, "$Z_{i,j}^1 = j^{th}$ dimension of i^{th} the position of the shark and *nd* implies decision variable count".

$$Z_i^1 = \left[Z_{i,1}^1, Z_{i,2}^1, \ldots . Z_{i,nd}^1\right] \tag{34}$$

Forward movement Here, "velocity *V*" is computed as exposed in Eq. (35) and (36).

$$V_i^1 = \left[V_1^1, V_2^1, \ldots . V_{np}^1\right] \tag{35}$$

$$V_i^1 = \left[V_{i,1}^1, V_{i,2}^1, \ldots . V_{i,nd}^1\right] \tag{36}$$

Hence, the velocity in all dimensions is evaluated as in Eq. (39), "wherein, $k = 1, 2, \ldots . k_{\max}$ $, \frac{\partial(obj)}{\partial \chi j}\Big|_{\chi_{i,j}^k}$ specify derivative *obj*at position $\chi_{i,j}^k$,$k_{\max}$ specify stage count for shark's forwarding movement, k specify stage count". Conservatively, $\Re 1$, $\Re 2$, $\Re 3$, α_k and β_k are randomly chosen, nonetheless as per OSA-SSO, $\Re 1$, $\Re 2$, $\Re 3$ are calculated as per tent map as in Eq. (37) and α_k and β_k are calculated as per logistic map as revealed in Eq. (38).

$$\Re_{k+1} = \begin{cases} \Re_k / 0.4, 0 < \Re_k \leq 0.4 \\ (1 - \Re_k)/0.6 \;\; 0.4 < \Re_k \leq 1 \end{cases} \tag{37}$$

$$\Re_{k+1} = 4\Re_k(1 - \Re_k) \tag{38}$$

$$V_{i,j}^k = \eta_k . \Re 1 . \frac{\partial(obj)}{\partial \chi j}\Big|_{\chi_{i,j}^k} \tag{39}$$

In all phases of $V_{i,j}^{k}$, the velocity limiter is deployed as exposed in Eq. (40), where α_k denotes coefficient of inertia.

$$V_{i,j}^{k} = \alpha_k.\Re 2.V_{i,j}^{k-1} + \eta_k.\Re 1.\frac{\partial(obj)}{\partial \chi j}\bigg|_{\chi_{i,j}^{k}} \tag{40}$$

The velocity limiter deployed for every phase of the OSA-SSO model is shown in Eq. (41), wherein, β_k implies a velocity limiter ratio for phase k.

$$|V_{i,j}^{k}| = \left[\left|\eta_k.\Re 1.\frac{\partial(obj)}{\partial \chi j}\right|_{\chi_{i,j}^{k}} + \alpha_k.\Re 2.V_{i,j}^{k-1}|, |\beta_k.V_{i,j}^{k-1}|\right] \tag{41}$$

The novel position of shark is in Eq. (42), Δt_k➔ time period.

$$G_i^{k+1} = Z_i^k + V_i^k.\Delta t_k \tag{42}$$

Rotational movement This phase is revealed in (43), wherein, $m = 1, 2\ldots M$.

$$Y_i^{k+1,m} = G_i^{k+1} + \Re 3.G_i^{k+1} \tag{43}$$

Particle position update This phase is revealed in Eq. (44), here; Y_i^{k+1} signify following shark position with high Obj.

$$Z_i^{k+1} = argmax\left\{obj\left(G_i^{k+1}\right), Obj\left(Y_i^{k+1,i}\right), \ldots, Obj\left(Y_i^{k+1,M}\right)\right\} \tag{44}$$

6 Results and discussion

6.1 Simulation set up

The proposed EC + OSA-SSO-based soil liquefaction scheme was executed in Python. Here, the analysis was done using a dataset that was downloaded from (http://cecas.clemson.edu/chichi/TW-LIQ/In-situ-Test.htm). Accordingly, the performance of adopted approach was measured over extant models such as Ensemble classifier with Shark Smell Optimization (EC + SSO) [32], Ensemble classifier with Salp Swarm Optimization (EC + SSA) [31], Ensemble classifier with Poor Rich Optimization (EC + PRO [33], Ensemble classifier with Butterfly Optimization Algorithm (EC + BOA) [7], Ensemble classifier with SVM + GWO [47], Ensemble classifier with Fuzzy Support Vector Machine (FSVM) [34], RF [28], CNN [19], LSTM [50], Naive Bayes (NB) [4] and DBN [43] regarding varied metrics. Statistical and convergence analysis were done to portray the effectiveness of EC + OSA-SSO.

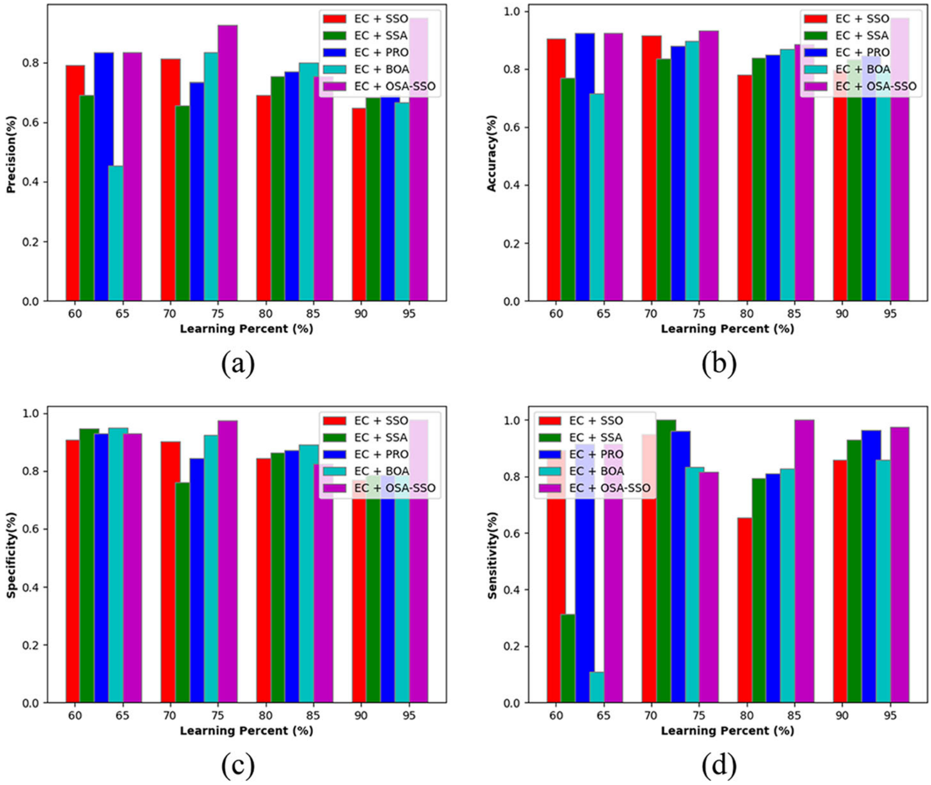
6.2 Performance analysis

The performances of developed EC + OSA-SSO are evaluated over extant optimization models. The analysis of the suggested EC + OSA-SSO model is computed over EC +

BOA, EC + SSA, EC + PRO and EC + SSO, models for varied LP that ranges from 60, 70, 80, and 90. Consequently, analysis was held using the dataset in (http://cecas.clemson.edu/chichi/TW-LIQ/In-situ-Test.htm) and the respective resultants are plotted from Figs. 3, 4 and 5. The analysis of EC + OSA-SSO over existing classifier models is shown in Table 3. On examining Fig. 3d, the sensitivity of the developed approach at 80th LP is 0.99, whereas, at 90th LP, the EC + OSA-SSO has obtained a superior sensitivity of 0.97. In the same way, for all the positive measures, the outputs for EC + OSA-SSO increase, and the outputs for negative measures decrease. In particular, tremendous outputs are attained at 90th LP for EC + OSA-SSO as well as existing methods. Mainly from Fig. 3b; high accuracy values (0.99) are gained by EC + OSA-SSO at 90th LP than at other LPs. At 90th LP, the accuracy of EC + OSA-SSO is 9.09% , 8.08%, and 10.1% better than the values obtained for conventional schemes like SSO, SSA, PRO, and BOA respectively. As a result, the investigation established the enhanced efficacy of ensemble classification with the incorporation of optimization theory.

6.3 Feature analysis

Table 4 depicts the feature analysis of deployed EC + OSA-SSO scheme over EC + OSA-SSO with existing entropy features and EC with no optimization. Here, analysis is done for



(a)

(b)

(c)

(d)

Fig. 3 Analysis using EC + OSA-SSO over others concerning "(**a**) Precision (**b**) Accuracy (**c**) Specificity and (**d**) Sensitivity"

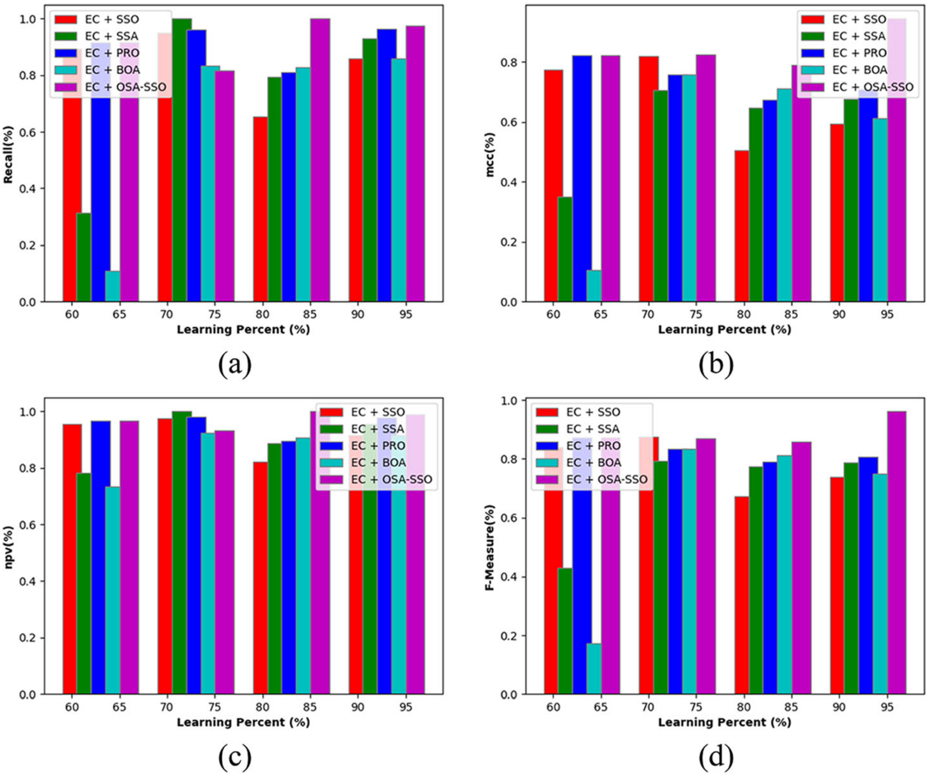


Fig. 4 Analysis using EC + OSA-SSO over others concerning "(**a**) Recall (**b**) MCC and (**c**) NPV (**d**) F-measure"

varied metrics. While observing the results, the suggested EC + OSA-SSO have attained maximal values than EC + OSA-SSO with existing entropy features and EC with no optimization. This ensures the enhancement of the developed model due to the integration of the OSA-SSO theory.

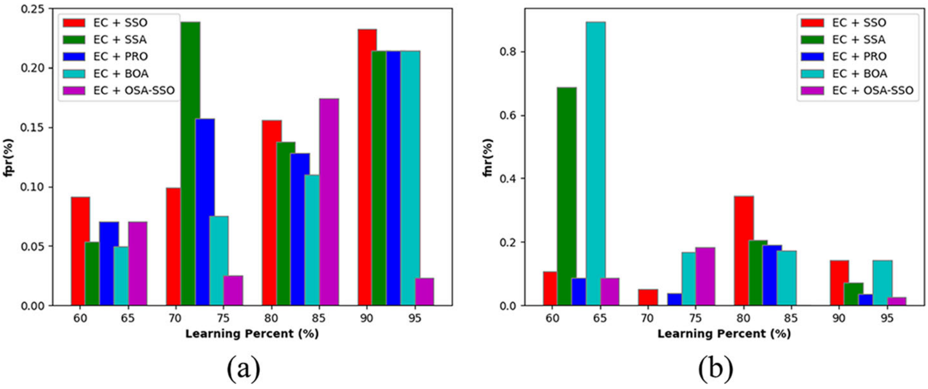


Fig. 5 Analysis using EC + OSA-SSO over others concerning "(**a**) FPR and (**b**) FNR"

Table 3 Analysis on EC + OSA-SSO over extant classifiers

Metrics	SVM+GWO [47]	FSVM [34]	RF	CNN	LSTM	NB	DBN	EC+OSA-SSO
F-Measure	0.58	0.66	0.87	0.41	0.12	0.44	0.09	0.86
Accuracy	0.81	0.84	0.92	0.75	0.70	0.72	0.70	0.93
Sensitivity	0.42	0.51	0.79	0.28	0.06	0.35	0.05	0.81
Precision	0.94	0.95	0.96	0.81	0.72	0.59	0.69	0.92
Recall	0.42	0.51	0.79	0.28	0.06	0.35	0.05	0.81
Specificity	0.98	0.98	0.98	0.97	0.75	0.88	0.78	0.97
MCC	0.54	0.62	0.83	0.37	0.21	0.29	0.18	0.82
NPV	0.79	0.81	0.91	0.74	0.70	0.75	0.69	0.93
FPR	0.01	0.01	0.01	0.02	0	0.11	0	0.02
FNR	0.57	0.48	0.20	0.71	0.93	0.64	0.94	0.18

6.4 Convergence analysis

Figure 6 depicts the cost study of EC + OSA-SSO scheme over conventional schemes namely, Ensemble classifier with Shark Smell Optimization (EC + SSO), EC + SSA, EC + PRO, and Ensemble Classifier with Butterfly Optimization AlgorithmEC + BOA. On noticing the outcomes, the suggested EC + OSA-SSO have obtained minimum values for all iterations. Initially, from 0 to 40th iteration, the cost values are somewhat higher for proposed as well as compared schemes. Here, PRO has revealed the worst performances at the initial iterations (0–20). Primarily, the offered approach has gained slighter cost (almost 1.015) with the amalgamation of the introduced optimization concept. Thus, the assessment confirmed the augmentation of the offered model. This establishes how the hybrid model converges to the specified fitness.

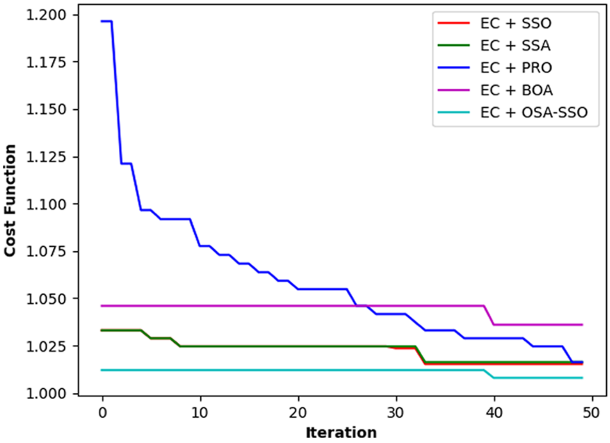
6.5 Statistical analysis

Table 5 describes the statistical analysis in terms of error for the adopted EC + OSA-SSO scheme over conventional schemes. On scrutinizing the investigational results, the adopted EC + OSA-SSO scheme has obtained negligible values. Subsequent to the developed model, EC + OSA-SSO approach has attained reduced cost values than EC + SSA, EC + PRO, EC + SSO, and EC + BOA models. Principally, minimal best-case scenario values are achieved by the developed model since it is enhanced via

Table 4 Analysis of existing and proposed features

Metrics	EC+OSA-SSO existing entropy features	EC without optimization	EC+OSA-SSO
F-Measure	0.82	0.79	0.86
Specificity	0.99	0.84	0.97
Accuracy	0.90	0.79	0.93
Precision	0.98	0.82	0.92
Recall	0.70	0.74	0.81
Sensitivity	0.70	0.74	0.81
MCC	0.78	0.58	0.82
NPV	0.88	0.76	0.93
FPR	0.05	0.15	0.02
FNR	0.29	0.25	0.18

Fig. 6 Convergence Analysis of OSA-SSO over others



optimization theory. Likewise, the EC + OSA-SSO model has acquired minimal outcomes for every scenario, hence showing betterment of adopted concepts.

6.6 Validation results

Table 6 depicts the validation results of EC + OSA-SSO versus existing method. The EC + OSA-SSO is evaluated with existing method such as EC + AC-SSO, EC + SSA, EC + CSO, and EC + GWO. The accuracy of EC + OSA-SSO is 1.15%, 2.08%, 4.81%, and 8.75% superior to existing method.

7 Conclusion

This work presented a novel soil liquefaction prediction scheme, where, pre-processing was done with data normalization. Subsequently, the features including "statistical and raw features, higher-order statistical features, and improved entropy and MI features" were derived. Further, DBN, LSTM, and RNN were deployed during prediction. Here, the outputs obtained from DBN and LSTM were fused and then given to optimal RNN, which provided the final output. In particular, tremendous outputs are attained at 90th LP for EC + OSA-SSO as well as existing methods. Mainly from Fig. 3b; high accuracy values (0.99) are gained by EC + OSA-SSO at 90th LP than at other LPs. At 90th LP, the accuracy of EC + OSA-SSO is 9.09%, 8.08%, and 10.1% better than the values obtained for conventional schemes like SSO,

Table 5 Statistical Analysis

Methods	Minimum	Best	Maximum	Worst	Std dev
EC+SSO	0.150937	0.084	0.149095	0.221557	0.06155
EC+SSA	0.180721	0.161677	0.165333	0.230539	0.028817
EC+PRO	0.124828	0.07485	0.13485	0.154762	0.031764
EC+BOA	0.177661	0.104	0.161106	0.284431	0.0691
EC+OSA-SSO	0.071013	0.024	0.073139	0.113772	0.031834

Table 6 Validation Results

Measures	EC+OSA-SSO	EC+SSA	EC+CSO	EC+GWO	EC+AC-SSO
Accuracy	0.78	0.72	0.83	0.80	**0.87**
Recall	0.83	0.77	0.85	0.82	**0.85**
Specificity	0.70	0.64	0.78	0.75	**0.83**
F-Measure	0.83	0.77	0.83	0.84	**0.85**
MCC	0.54	0.41	0.63	0.57	**0.67**
Precision	0.82	0.78	0.87	0.86	**0.90**
NPV	0.71	0.62	0.74	0.70	**0.75**
Sensitivity	0.77	0.77	0.71	0.74	**0.79**
FPR	0.29	0.35	0.21	0.24	**0.16**
FNR	0.22	0.22	0.28	0.25	**0.20**

SSA, PRO, and BOA respectively.The major advantages are the prediction accuracy is improved and the liquefaction risks are removed. If an earthquake strikes suddenly and unexpectedly, it is critical to estimate seismic damage potentials over the target area quickly to limit damage and give an effective means of emergency control. In the future, non-liquefied cases will be analyzed, and the location will be determined before the occurrence of an earthquake. Before including liquefaction records in the dataset, a magnitude of liquefaction index can be generated using the information on the modifications required.

Data availability statement No new data were generated or analyzed in support of this research.

Author contribution All authors have made substantial contributions to conception and design, revising the manuscript, and the final approval of the version to be published. Also, all authors agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding This research did not receive any specific funding.

Declaration

Conflict of interest The authors declare no conflict of interest.

Acknowledgements I would like to express my very great appreciation to the co-authors of this manuscript for their valuable and constructive suggestions during the planning and development of this research work.

Informed consent Not Applicable.

Ethical approval Not Applicable.

References

1. Ahmad M, Tang XW, Qiu JN, Ahmad F, Gu WJ (2020) A step forward towards a comprehensive framework for assessing liquefaction land damage vulnerability: exploration from historical data. Front Struct Civ Eng 14:1476–1491. https://doi.org/10.1007/s11709-020-0670-z
2. Ahmad M, Tang XW, Qiu JN, Ahmad F, Gu WJ (2021) Application of machine learning algorithms for the evaluation of seismic soil liquefaction potential. Front Struct Civ Eng 15:490–505. https://doi.org/10.1007/s11709-020-0669-5

3. Alizadeh Mansouri M, Dabiri R (2021) Predicting the liquefaction potential of soil layers in Tabriz city via artificial neural network analysis. SN Appl. Sci. 3:719. https://doi.org/10.1007/s42452-021-04704-3
4. Alizadeh SH, Hediehloo A, Harzevili NS (2020) Multi independent latent component extension of naive Bayes classifier. Knowledge-based systems, volume 213 (cover date: 15 February 2021)article 106646, 24
5. AlKhatib AA, Sawalha T, AlZu'bi S (2020) Load balancing techniques in software-defined cloud computing: an overview. 2020 Seventh International Conference on Software Defined Systems (SDS). IEEE
6. AlZu'bi S, Shehab M, al-Ayyoub M, Jararweh Y, Gupta B (2020) Parallel implementation for 3d medical volume fuzzy segmentation. Pattern Recogn Lett 130:312–318
7. Arora S, Singh S (2019) Butterfly optimization algorithm: a novel approach for global optimization. Soft Comput 23:715–734. https://doi.org/10.1007/s00500-018-3102-4
8. Das SK, Mohanty R, Mohanty M et al (2020) Multi-objective feature selection (MOFS) algorithms for prediction of liquefaction susceptibility of soil based on in situ test methods. Nat Hazards 103:2371–2393. https://doi.org/10.1007/s11069-020-04089-3
9. Ferreira C, da Viana Fonseca A, Ramos C et al (2020) Comparative analysis of liquefaction susceptibility assessment methods based on the investigation on a pilot site in the greater Lisbon area. Bull Earthq Eng 18: 109–138. https://doi.org/10.1007/s10518-019-00721-1
10. George A, Rajakumar BR (2013) APOGA: An Adaptive Population Pool Size based Genetic Algorithm. AASRI Procedia - 2013 AASRI Conf Intel Syst Contr (ISC 2013), Vol. 4. pages 288–296 https://doi.org/10.1016/j.aasri.2013.10.043
11. Ghani S, Kumari S (2021) Probabilistic study of liquefaction response of fine-grained soil using multi-linear regression model. J Inst Eng India Ser A 102:783–803. https://doi.org/10.1007/s40030-021-00555-8
12. Haeri H, Sarfarazi V, Shemirani AB, Gohar HP, Nejati HR (2017) Field evaluation of soil liquefaction and its confrontation in fine-grained Sandy soils (case study: south of Hormozgan Province). J Min Sci 53:457–468. https://doi.org/10.1134/S1062739117032356
13. Hoang ND, Bui DT (2018) Predicting earthquake-induced soil liquefaction based on a hybridization of kernel fisher discriminant analysis and a least squares support vector machine: a multi-dataset study. Bull Eng Geol Environ 77:191–204. https://doi.org/10.1007/s10064-016-0924-0
14. Hsein Juang C, Shen M, Wang C, Chen Q (2018) Random field-based regional liquefaction hazard mapping — data inference and model verification using a synthetic digital soil field. Bull Eng Geol Environ 77:1273–1286. https://doi.org/10.1007/s10064-017-1071-y
15. Hu J (2021) A new approach for constructing two Bayesian network models for predicting the liquefaction of gravelly soil. Comput Geotech, volume 137 (cover date: September 2021) article 104304, 18
16. Jadhav AN, Gomathi N (2019) DIGWO: hybridization of dragonfly algorithm with improved Grey wolf optimization algorithm for data clustering. Multimed Res 2(3):1–11
17. Javdanian H (2019) Evaluation of soil liquefaction potential using energy approach: experimental and statistical investigation. Bull Eng Geol Environ 78:1697–1708. https://doi.org/10.1007/s10064-017-1201-6
18. Javdanian H, Heidari A, Kamgar R (2017) Energy-based estimation of soil liquefaction potential using GMDH algorithm. Iran J Sci Technol Trans Civ Eng 41:283–295. https://doi.org/10.1007/s40996-017-0061-4
19. Jiuxiang G, Wang Z, Kuen J, Lianyang MA, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J, Chen T, Recent advances in convolutional neural networks, Pattern Recogn, vol. 77, pp354–377, 2018.
20. Kao L-J, Chiu CC (2020) Application of integrated recurrent neural network with multivariate adaptive regression splines on SPC-EPC process. J Manuf Syst 57:109–118
21. Karafagka S, Fotopoulou S, Pitilakis D (2021) Fragility curves of non-ductile RC frame buildings on saturated soils including liquefaction effects and soil–structure interaction. Bull Earthq Eng 19:6443–6468. https://doi.org/10.1007/s10518-021-01081-5
22. Kumar K, Shrimankar DD (2017) F-DES: fast and deep event summarization. IEEE Transac Multimed 20(2):323–334
23. Kumar K, Shrimankar DD, Singh N (2016) Equal partition based clustering approach for event summarization in videos. 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). IEEE
24. Kumar K, Shrimankar DD, Singh N (2018) Eratosthenes sieve based key-frame extraction technique for event summarization in videos. Multimed Tools Appl 77(6):7383–7404
25. Kumar D, Samui P, Kim D, Singh A (2021) A novel methodology to classify soil liquefaction using deep learning. Geotech Geol Eng 39:1049–1058. https://doi.org/10.1007/s10706-020-01544-7
26. Kurnaz TF, Kaya Y (2019) A novel ensemble model based on GMDH-type neural network for the prediction of CPT-based soil liquefaction. Environ Earth Sci 78:339. https://doi.org/10.1007/s12665-019-8344-7
27. Lafi M, Hawashin B, AlZu'bi S (2021) Eliciting requirements from Stakeholders' responses using natural language processing. Comput Model Engin Sci 127(1):99–116

28. Li Q, Chen L, Li X, Xia S, Kang Y (2020) An improved random forests approach for interactive lobar segmentation on emphysema detection. Granul Comput 5:503–512. https://doi.org/10.1007/s41066-019-00171-9
29. Liu JH (2018) Identification of ground motion intensity measure and its application for predicting soil liquefaction potential based on the Bayesian network method. Eng Geolog, Volume 248 (Cover date: 8 January 2019). Pages 34-49
30. Mahmood A, Tang XW, Qiu JN et al (2020) A hybrid approach for evaluating CPT-based seismic soil liquefaction potential using Bayesian belief networks. J. Cent. South Univ 27:500–516. https://doi.org/10.1007/s11771-020-4312-3
31. Mirjalili S, Gandomi AH, Mirjalili SM (2017) Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems. Advances in Engineering Software 114:163–191
32. Mohammad-Azari S, Bozorg-Haddad O, Chu X (2018) Shark smell optimization (SSO) algorithm. In: Bozorg-Haddad O (ed) Advanced optimization by nature-inspired algorithms. Studies in computational intelligence, vol 720. Springer, Singapore. https://doi.org/10.1007/978-981-10-5221-7_10
33. Moosavi S, Bardsiri V (2019) Poor and rich optimization algorithm: A new human-based and multi populations algorithm. Engineering Applications of Artificial Intelligence 86. 165–181. https://doi.org/10.1016/j.engappai.2019.08.025
34. Rahbarzare A, Azadi M (2019) Improving prediction of soil liquefaction using hybrid optimization algorithms and a fuzzy support vector machine. Bull Eng Geol Environ 78:4977–4987. https://doi.org/10.1007/s10064-018-01445-3
35. Rajakumar BR (2013) Impact of Static and Adaptive Mutation Techniques on Genetic Algorithm. Int J Hybrid Intel Syst 10(1):11–22. https://doi.org/10.3233/HIS-120161
36. Rajakumar BR (2013) Static and Adaptive Mutation Techniques for Genetic algorithm: A Systematic Comparative Analysis. Int J Comput Sci Eng 8(2):180–193. https://doi.org/10.1504/IJCSE.2013.053087
37. Rajakumar BR, George A (2012) A New Adaptive Mutation Technique for Genetic Algorithm. In: Proceedings of IEEE International Conference on Computational Intelligence and Computing Research (ICCIC). pages: 1–7, December 18–20, Coimbatore, India. https://doi.org/10.1109/ICCIC.2012.6510293
38. Sabbar AS, Chegenizadeh A, Nikraz H (2019) Prediction of liquefaction susceptibility of clean Sandy soils using artificial intelligence techniques. Indian Geotech J 49:58–69. https://doi.org/10.1007/s40098-017-0288-9
39. Sadashiv Halbhavi B, Kodad SF, Ambekar SK, Manjunath D (2019) Enhanced invasive weed optimization algorithm with Chaos theory for weightage based combined economic emission dispatch. J Comput Mech, Power Syst Contr 2(3):19–27
40. Sarhrouni E, Hammouch A, Aboutajdine D (2012) Application of Symmetric Uncertainty and Mutual Information to Dimensionality Reduction and Classification of Hyperspectral Images. Int J Eng Technol. 4
41. Swamy SM, Rajakumar BR, Valarmathi IR (2013) Design of Hybrid Wind and Photovoltaic Power System using Opposition-based Genetic Algorithm with Cauchy Mutation. IET Chennai Fourth Int Conf Sustain Energy Intel Syst (SEISCON 2013), Chennai, India. https://doi.org/10.1049/ic.2013.0361
42. Wagh MB, Gomathi N (2019) Improved GWO-CS algorithm-based optimal routing strategy in VANET. J Netw Commun Syst 2(1):34–42
43. Wang HZ, Wang GB, Li GQ, Peng JC, Liu YT (2016) Deep belief network based deterministic and probabilistic wind speed forecasting approach. Appl Energy 182:80–93
44. Wang J, Deng Y, Wu L, Liu X, Yin Y, Xu N (2018) Estimation model of sandy soil liquefaction based on RES model. Arab J Geosci 11:565. https://doi.org/10.1007/s12517-018-3885-8
45. Zhang J, Wang Y (2021) An ensemble method to improve prediction of earthquake-induced soil liquefaction: a multi-dataset study. Neural Comput & Applic 33:1533–1546. https://doi.org/10.1007/s00521-020-05084-2
46. Zhang Y, Wang R, Zhang JM, Zhang J (2020) A constrained neural network model for soil liquefaction assessment with global applicability. Front Struct Civ Eng 14:1066–1082. https://doi.org/10.1007/s11709-020-0651-2
47. Zhang Y, Qiu J, Zhang Y, Xie Y (2021) The adoption of a support vector machine optimized by GWO to the prediction of soil liquefaction. Environ Earth Sci 80:360. https://doi.org/10.1007/s12665-021-09648-w
48. Zhang Y, Xie Y, Zhang Y, Qiu J, Wu S (2021) The adoption of deep neural network (DNN) to the prediction of soil liquefaction based on shear wave velocity. Bull Eng Geol Environ 80:5053–5060. https://doi.org/10.1007/s10064-021-02250-1
49. Zhang YG, Qiu J, Zhang Y et al (2021) The adoption of ELM to the prediction of soil liquefaction based on CPT. Nat Hazards 107:539–549. https://doi.org/10.1007/s11069-021-04594-z
50. Zhou X, Lin J, Zhang Z, Shao Z, Liu H (2019) Improved itracker combined with bidirectional long short-term memory for 3D gaze estimation using appearance cues. Neuro computing in press, corrected proof, available online 20

51. Zhou YG, Xia P, Ling DS, Chen YM (2020) A liquefaction case study of gently sloping gravelly soil deposits in the near-fault region of the 2008 Mw7.9 Wenchuan earthquake. Bull Earthq Eng 18:6181–6201. https://doi.org/10.1007/s10518-020-00939-4
52. Zuzulock ML, Taylor ODS, Maerz NH (2020) Soil fatigue hazard screening analyses framework for spacio-temporally clustered induced seismicity with examples of damage potential due to liquefaction. SN Appl Sci 2:1072. https://doi.org/10.1007/s42452-020-2878-x

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

OptNet-Fake: Fake News Detection in Socio-Cyber Platforms Using Grasshopper Optimization and Deep Neural Network

Sanjay Kumar, Akshi Kumar, Abhishek Mallik, and Rishi Ranjan Singh

***Abstract*— Exposure to half-truths or lies has the potential to undermine democracies, polarize public opinion, and promote violent extremism. Identifying the veracity of fake news is a challenging task in distributed and disparate cyber-socio platforms. To enhance the trustworthiness of news on these platforms, in this article, we put forward a fake news detection model, *OptNet-Fake*. The proposed model is architecturally a hybrid that uses a meta-heuristic algorithm to select features based on usefulness and trains a deep neural network to detect fake news in social media. The d-D feature vectors for the textual data are initially extracted using the term frequency inverse document frequency (TF-IDF) weighting technique. The extracted features are then directed to a modified grasshopper optimization (MGO) algorithm, which selects the most salient features in the text. The selected features are then fed to various convolutional neural networks (CNNs) with different filter sizes to process them and obtain the n-gram features from the text. These extracted features are finally concatenated for the detection of fake news. The results are evaluated for four real-world fake news datasets using standard evaluation metrics. A comparison with different meta-heuristic algorithms and recent fake news detection methods is also done. The results distinctly endorse the superior performance of the proposed *OptNet-Fake* model over contemporary models across various datasets.**

***Index Terms*— Convolutional neural network (CNN), fake news detection, feature selection, grasshopper optimization algorithm (GOA), term frequency inverse document frequency (TF-IDF).**

I. Introduction

OWING to the low-cost internet-enabled devices with easy and anytime access to the web, the use of social media platforms such as Facebook, Twitter, Instagram, and WhatsApp has grown fast and profound. The current global statistics reveal an active social media user pool of 4.55 billion [1]. While the amount of user-generated content is proliferating, the speed of diffusion is unusually striking, thus creating a quintessential "breeding ground" for posting and disseminating antagonistic content, which includes fake news, rumors, offensive, hateful, and bullying content [2]. Fake news stories have been afflicting countries globally. Some online information is blatantly fake or misleading, and some stories are subtly wrong. Fake news is not new, but new communication technologies such as social media have led to the propagation of fake news [3]. The recent pandemic is a witness to this contamination of information, where unconfirmed and falsified news on proven prevention, cures, and medication is being circulated, putting lives at risk. The World Health Organization (WHO) has called the spread of fake news about COVID-19 an infodemic. Countering falsehood and fact-checking the avalanche of information is a conscientious and time-consuming process, whereas masquerading (purposely creating false accounts) or automated bots can plague the digital media with alarming speed. Moreover, fake news could be in multiple forms, such as rumors, satire, false advertisements, cyber-bullying, and hate speech [4]. Such false or manipulated information not only creates a sense of distrust in online news and communication but also has a significant impact on our opinion and may lead to distrust and unrest in society.

Usually, fake information can be of two types—misinformation and disinformation. Misinformation is factually incorrect facts. However, disinformation is false information that is circulated to mislead the public leading to economic, political, and social impacts. The information is manipulated so that the reader feels it is correct. Fake news designers use social engineering and deception techniques on social media platforms and influence users' behaviors by persuading them, for example, to click on web links. The deception techniques involve a psychological process; in this way, fake news creators intend to achieve financial benefits. According to Kshetri and Voas [5], someone will engage in the creation and spreading of fake news based on the following mathematical equation:

$$M_b + P_b > I_c + O_{1c} + P_c + (O_{2c} \times \pi_{\text{arr}} \times \pi_{\text{con}}). \quad (1)$$

Here, the monetary and psychological gains or benefits reaped by fraudsters are represented by M_b and P_b, respectively. The direct investments, opportunity costs, and psychological costs associated with creating and managing fake news are represented by I_c, O_{1c}, and P_c, respectively. Also, O_{2c}, π_{arr},

Manuscript received 23 September 2022; revised 16 November 2022 and 30 January 2023; accepted 11 February 2023. *(Corresponding author: Akshi Kumar.)*

Sanjay Kumar and Abhishek Mallik are with the Department of Computer Science and Engineering, Delhi Technological University, New Delhi 110042, India (e-mail: sanjay.kumar@dtu.ac.in; abhishekmallik265@gmail.com).

Akshi Kumar is with the Department of Computing and Mathematics, Faculty of Science and Engineering, Manchester Metropolitan University, M15 6BH Manchester, U.K. (e-mail: akshi.kumar@mmu.ac.uk).

Rishi Ranjan Singh is with the Department of Electrical Engineering and Computer Science, IIT Bhilai, Sejbahar 492015, India (e-mail: rishi@iitbhilai.ac.in).

Digital Object Identifier 10.1109/TCSS.2023.3246479

and π_{con} represent the monetary costs of conviction, probability of arrest, and probability of conviction, respectively.

The multilingual and multimodal nature of the various social media platforms acts as an amplifier and rapid distribution channel for the propagation of fake news, making fake news detection a challenging task. It is a challenging task in the sense that fake news articles and texts often use modified slangs and languages along with terms which are not in the dictionary. Also, sometimes fake news creators use a mix of several languages which are difficult to detect. The extensive number of fake news spreaders further complicates the process of debunking. Furthermore, the constant adaptation of the design and presentation of content such that it appears more legitimate makes it nearly impossible to halt this widespread phenomenon in the socio-cyber world [6]. At the same time, not all the users can discern fake from actual news. Moreover, anyone with internet access can easily create malicious accounts, such as cyborg users and social bots leading to increased fake news. This makes it impossible to assess the never-ending online data manually. To minimize the circulation of fake news, several fact-checking projects such as the Google News Initiative and Verizon have been introduced. In general, fake news detection tasks can be accomplished with the help of feature extraction followed by model building. Few of the commonly used methods to detect fake news include content-based identification, feedback-based identification, and intervention-based solutions [7]. Most of the existing fake news detection methods use textual content, user responses, and computational methods [6]. However, these methods have several limitations, such as reduced feature relevance, redundant features, and increased feature correlation that lead to a low detection rate and accuracy of the model.

The proposed OptNet-Fake model is built by pairing the modified grasshopper optimization (MGO) meta-heuristic algorithm with a convolutional neural network (CNN) architecture. We start by preprocessing the textual data (news articles and social media posts) to achieve uniformity across the dataset. To create the initial feature matrix, we extract features using the TF-IDF technique. This gives a d-D feature set for every news article represented in a d-D feature space. The extracted features are then passed to the MGO algorithm, which selects the most relevant features from the generated feature space. As the classical grasshopper optimization is suitable for continuous tasks, we modify it to work on discrete tasks such as feature selection. The selected features are then fed to a deep CNN to process them and obtain the n-gram features from the text, which are further used to perform the final fake news detection. We examine the performance of the proposed OptNet-Fake model on four benchmark real-world datasets, namely, Kaggle Fake News Dataset [8], ISOT Fake News Dataset [9], COVID-19 Fake News Dataset [10], and WELFake Dataset [11] and evaluate several performance metrics for the task of fake news detection. The results of the MGO-CNN-based OptNet-Fake model are compared with several contemporary methods of fake news detection to obtain a comparative performance and utility of the introduced model. The main contribution of our work can be summarized as follows.

1) We propose a novel fake news detection model, OptNet-Fake, using an MGO and CNN.
2) We adopt a feature generation technique using TF-IDF to exploit the importance of occurrences of words in a text segment and across a text corpora.
3) We modify the classical Grasshopper Optimization algorithm (GOA) for feature selection to capture the most relevant and the least correlated features for fake news detection.
4) The proposed work uses a deep CNN architecture to extract the n-gram features to better characterize the news articles.
5) A comparative study involving various metaheuristics and contemporary fake news detection methods has been performed using four benchmarked real-world datasets.

The rest of the article is organized as follows. Section II presents a discussion on the already existing work in the field of fake news detection. Various datasets used by us for our experimental study are described in Section III. We illustrate our proposed work in detail in Section IV. The experimental results and analysis are presented in Section V. Finally, the concluding remarks are mentioned in Section VI.

II. Related Work

The proliferation of fake news on online social networks has gained much attention in the literature, and an increasing number of researchers have been exploring this field. We can broadly divide various fake news detection methods into three groups, namely, knowledge-graph-based, linguistic-based, and machine-learning and deep-learning-based techniques [12]. The knowledge-graph-based approach analyses network behavior and structure to bring out false news. This can be carried out in multiple ways, including knowledge graph analysis, which has an accuracy of 61%–95%, as claimed by Ciampaglia et al. [13]. They also studied the relationship between entities and put forward the theory of network effect variables. Another graph-based approach was used by Gangireddy et al. [14] in their study, where they identified misinformation with the help of label spreading, biclique identification, and feature vector learning. Their approach comprises three phases and achieves an accuracy near 80% for unsupervised detection of fake news.

Yang et al. [15] explored the source user's characteristics on Sina Weibo, China's popular social media platform. They examined a set of features and put forward a classifier to bring out false information. In general, the linguistic analysis methods are based on the n-gram approach, part-of-speech tags, and probabilistic context-free grammar (PCFG) [7]. The n-gram approach uses patterns of n continuous words within a text, consisting of words and phrases. Syntactic features such as part-of-speech tags are acquired by tagging every word according to a syntactic feature such as adjectives and nouns. The PCFG uses a CFG to denote a sentence's grammatical structure. The intermediate and terminal nodes represent syntactic constituents and words, respectively.

Zheng et al. [16] used supervised machine learning for effective and efficient spammer detection. They did this by collecting a dataset, classifying the users as spammers and non-spammers, and then using an SVM-based spammer detection algorithm. Verma et al. [11] introduced a WELFake model for fake news detection using a word embedding over linguistic features and machine learning classification. They preprocessed the dataset and validated the news content's veracity using linguistic features followed by a voting-based machine learning classifier. Karimi et al. [17] used a set of long short-term memory (LSTM) for multiclass and multisource fake news detection to discover the numerous degrees of fake news. Wu and Liu [18] presumed that intentionally spread false information is often manipulated to seem authentic. To bring out this falsified news, they used embeddings along with a combination of LSTM and recurrent neural networks (RNNs) to build a classifier based on propagation pathways in social media. A CNN-based study was given by Yang et al. [19]. They presented a model called TI-CNN, using latent and explicit features to analyze texts and images for incorrect information. Paka et al. [20] proposed a novel framework named Cross-SEAN for fake tweets' detection related to COVID-19. They introduced CTF, a large labeled Fake Tweets dataset. As part of Cross-SEAN, they proposed a cross-stitch-based semi-supervised end-to-end neural attention model. Shrivastava et al. [21] developed a model using differential equations to investigate the effect of user verification and blocking and the spread of messages on online social networks. Furthermore, the model presented the controlling mechanism for untrusted message propagation. The recently proposed susceptible-infected-recovered-anti-spreader (SIRA) model is based on the spread of epidemics and its applications to modeling the propagation and control of rumors in online social networks [22]. The proposed technique introduced a mechanism for the recovery of nodes in a situation where rumor propagation and rumor control happen simultaneously. Sengupta et al. [23] presented a blockchain-based model named ProBlock to perform the correctness of information propagated. The model used a secure voting system, where news reviewers can provide feedback on the news. A probabilistic mathematical model forecasts the news item's truthfulness based on the feedback received. Trueman et al. [24] proposed an approach for fake news detection and its classification into six subcategories. They presented an attention-based convolutional bi-directional LSTM framework.

III. Datasets

This section describes about the various datasets used in the study to gather the results. We have used various types of datasets for our study, and these include Kaggle Fake News Dataset [8], ISOT Fake News Dataset [9], COVID-19 Fake News Dataset [10], and WELFake Dataset [11]. Each dataset represents a different context and is suitable for our study. The composition of fake articles and real articles in the used datasets is given in Table I.

1) Kaggle Fake News: The Kaggle dataset contains both reliable and unreliable articles that context the 2016 U.S. presidential elections. It includes 20 800 IDs, 20 242 titles, 18 843 authors, 20 671 texts, and 20 800 labels. This dataset contains 10 413 real news and 10 387 fake news articles.

TABLE I
Composition of Fake and Real Articles in the Used Datasets

Dataset	Real	Fake
Kaggle Fake News	10413	10387
ISOT Fake News	21417	23481
COVID-19 Fake News	5600	5100
WELFake	35028	37106

2) ISOT Fake News: The dataset [9] contains fake and real news articles. The truthful articles were obtained from Reuters.com, and the fake articles were collected from various websites indicated by Politifact. The dataset contains articles related to political news, world news, government news, and regional news of the United States and the Middle East. This dataset contains 21 417 real news and 23 481 fake news articles.

3) COVID-19 Fake News: The dataset contains 10 700 social media posts based on COVID-19 news [10] collected from various platforms such as Facebook, Twitter, Instagram, Politifact, WHO, Indian Council of Medical Research (ICMR), and Centers for Disease Control and Prevention (CDC). It contains 5600 real and 5100 fake news articles.

4) WELFake: Word Embedding over Linguistic Features for Fake News Detection (WELFake) dataset [11] contains 72 134 news articles with a distribution of 35 028 real and 37 106 fake articles. The news articles were collected from four popular platforms: Kaggle, McIntire, Reuters, and BuzzFeed Political. The dataset contains four columns, namely, Serial Number, Title, Text, and Label, whether the article is fake, where 0 represents fake and 1 represents real.

A. Evaluation Metrics

In this section, we mention the evaluation metrics used in this study to measure the performance of the fake news detection model. We have used standard evaluation metrics: accuracy, precision, recall, and F1 score. Since all these metrics use information represented in the confusion matrix, we also computed the confusion matrix. Accuracy measures the total correctly identified samples out of all the samples. The measure of the ratio between the true positives and all the positives is known as precision. Precision helps us understand the model's performance to classify actual fake news articles as counterfeit among all the news articles classified as fake. However, the ability of the model to accurately identify the occurrence of a positive class instance is determined by recall. Recall helps us in deciding the number of actual fake news articles that are classified as fake. F1 score can be defined as the harmonic mean of precision and recall value. F1 score helps ascertain the model's performance in striking a balance between precision and recall when there is an imbalanced dataset.

IV. Proposed OptNet-Fake Model

This section illustrates the details of the proposed OptNet-Fake model for fake news detection using an MGO and CNNs techniques. The proposed model has four components: data processing, feature generation, feature selection

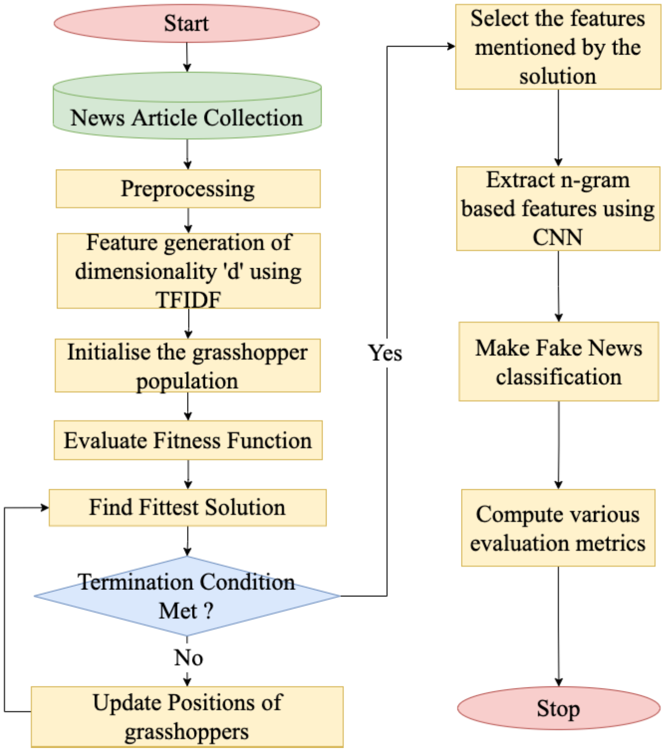

Start
News Article Collection
Preprocessing
Feature generation of dimensionality 'd' using TFIDF
Initialise the grasshopper population
Evaluate Fitness Function
Find Fittest Solution
Termination Condition Met ?
No
Update Positions of grasshoppers
Yes
Select the features mentioned by the solution
Extract n-gram based features using CNN
Make Fake News classification
Compute various evaluation metrics
Stop

Fig. 1. Flow diagram for the proposed model for fake news detection using MGO and CNN.

using MGO, and classification using CNN. In the first component, we start by cleaning and processing the input news articles to obtain uniformity across the news articles. Then in the second component, we represent each news article in a lower dimensional vector space (d). We perform feature selection using MGO to obtain more relevant features, reducing the dimensionality in component 3. Finally, in component 4, we process the selected features using CNNs to obtain n-gram features and classify each news article as real or fake based on its characteristics. Fig. 1 shows the overall architectural flow of OptNet-Fake.

A. Data Processing

The raw news articles present in the dataset or available on the internet have too many aberrations and are very noisy. Hence, we preprocess every news article to remove such words and obtain uniformity across the articles. First, we remove all the URLs, HTML tags, parentheses, slashes, dashes, and multiple white spaces from the news articles. Then we convert all the words of type “@Alice” and “#Bob” to “Alice” and “Bob.” Then we convert all the news articles into lower case and remove all the stopwords such as a, an, and the http://www.ranks.nl/stopwords. If a word has a character repeated more than three times consecutively, we replace that with a single occurrence of that character. For instance, “Wowwwwww” is replaced with “Wow.” We also replace acronyms with their full forms www.netlingo.com/acronyms.php. For example, “UN” changes to “United Nations.” Finally, we obtain a well-processed dataset with uniformity across the news articles.

B. Feature Generation

In general, for a classification task, a well-defined and uniform feature set for the data points and classes in which they will be classified is required. For our study of fake news detection, we consider the individual news articles as data points for our classification while real and fake act as our classes. We obtain the term frequency inverse document frequency (TF-IDF) values for every word in the article and then sort the words in descending order of their TF-IDF values. We then select the top d words based on their TF-IDF values to obtain a feature set in a predecided (d) dimensional vector space. The notion behind selecting the top d words and not all the words in the corpora is that articles may have a lot of words, and all the words are not equally relevant. Moreover, expanding the dimensionality of the vector space (d) to the number of different words in the corpora also tends to increase the computational complexity by a considerable margin. The obtained vector space acts as the feature space for news articles, and the corresponding vectors act as feature set for each news article. We vary the value of d to understand its impact on the performance of our proposed work and choose that value that yields the optimal results.

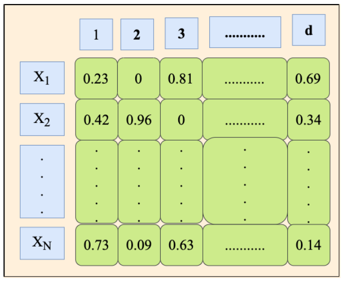
C. Feature Selection Using MGO

In recent years, metaheuristic algorithms have gained attraction for solving various optimization problems such as feature selection because of their ability to avoid local optima and search for a solution close to global optima. GOA is a recent swarm intelligence algorithm proposed by Saremi et al. [25] which mimics the grasshoppers’ foraging and swarming behavior. The life cycle of grasshoppers has two phases: the nymph phase and the adult phase. The nymph phase includes small steps and slow movements, while the adult phase includes long-range and abrupt movements. We introduce a modified version of the recently proposed GOA for feature selection as classical GOA being continuous in nature cannot be used for a discrete problem such as feature selection. Feature selection refers to selecting a subset of features from a feature space to yield the most optimal results with less computational cost. We start by generating a population of grasshoppers characterized by a d-D vector where d is the number of features with their values randomly initialized in the range of 0–1. Then based on the fitness function, we estimate the fitness value of all the grasshoppers and update the positions of all of them based on the fittest grasshopper. This step is repeated iteratively until the termination condition is met. After the process is terminated, we choose the attributes based on the fittest grasshopper. This is done by selecting the attributes for which the value of the grasshopper’s vector space is greater than 0.5. This helps us modify the classical GOA to suit the discrete task of feature selection. To better understand the algorithm, we also present a notation Table II. The various phases of the proposed MGO algorithm are described below.

1) Population Initialization: For any population-based algorithm, the first is to initialize a population. Let N be the size of the population and d be the dimensionality of the feature space. So for every grasshopper in the population, we generate

TABLE II
NOTATION TABLE FOR THE MGO ALGORITHM FOR FEATURE SELECTION

Notation	Description
N	Size of the population
d	Dimensionality of the feature space
X_i	Position of the i^{th} grasshopper
D_{ij}	Distance between the i^{th} & j^{th} grasshoppers
c	Comfort, attraction & repulsion zone decreasing coefficient
l	Current iteration
L	Maximum number of iterations
ub_k & lb_k	Upper and lower bounds of k^{th} dimension
S()	The social interaction function
T	The fittest grasshopper so far



	1	2	3	....	d
X_1	0.23	0	0.81	....	0.69
X_2	0.42	0.96	0	....	0.34
. . .	. . .	. . .	. . .	. . .	. . .
X_N	0.73	0.09	0.63	....	0.14

Fig. 2. Representation of the positions of the entire population of the grasshopper.

a d-D vector having random values in the interval [0, 1]. Hence, the position of every grasshopper can be represented as X_i, $(i = 1, 2, 3, \ldots, N)$. Each dimension of the grasshopper i can be represented as follows:

$$X_{ij} = \text{random}(), \quad j \in [1, d] \text{ and } i \in [1, N]. \tag{2}$$

Here, random() gives a random number in the range [0, 1]. Fig. 2 shows the representation of the positions of all the grasshoppers.

2) Fitness Function Calculation: After initializing the population, we need to evaluate the fitness function for every solution, i.e., we need to evaluate the fitness value of every grasshopper. For our work, we use the following fitness function for the ith grasshopper:

$$\text{Fitness}(X_i) = \text{errorRate} * \frac{\sum_{j=1}^{d} \text{round}(X_{ij})}{d}. \tag{3}$$

Here, round(X_{ij}) returns the rounded-off value of X_{ij}, i.e., for values greater than 0.5, it returns 1, while for values less than 0.5, it returns 0. The errorRate is the classification error using the selected features made by an artificial neural network classifier. For the ith possible solution, the jth feature is selected if the value of $X_{ij} > 0.5$. For this study, we try to minimize the fitness function, i.e., (3). This is done based on the notion that we try to reduce the classification error for any classification task while selecting the minimum number of features. The rationale behind using the fitness function mentioned in (3) is that it represents the product of classification error and the number of features selected by our algorithm. Here, the classification error refers to an error in classifying articles as real or fake using an artificial neural network. The chosen features represent the features of the grasshopper adopted by the proposed algorithm. Optimizing the fitness function involves minimizing the classification error while selecting the minimum number of features to obtain a robust model that can efficiently give good results.

3) Position Update: After calculating the fitness function for every grasshopper, next we go onto updating their positions by considering the social interaction operator (S_i), the gravity force operator (G_i), and the wind advection operator (A_i) as follows:

$$X_i = S_i + G_i + A_i. \tag{4}$$

To fit the feature selection task in a better way, we modify the above equation and ignore the effect of gravity operator (G_i) and assume that the direction of wind is always toward the target. The target is the fittest grasshopper. Therefore, the position update equation changes as follows:

$$X_{ik}^{t+1} = c\left(\sum_{j=1, j\neq i}^{N} c\frac{\mathrm{ub}_k - \mathrm{lb}_k}{2} S\left(\left|X_{jk}^t - X_{ik}^t\right|\right)\frac{X_j^t - X_i^t}{D_{ij}}\right) + T_k^t. \tag{5}$$

Here, X_{ik}^t represents the value of the kth dimension of the position of the ith grasshopper at time t, c is a decreasing coefficient to shrink the comfort zone, attraction zone, and repulsion zone. T_k^t denotes the fittest solution or fittest grasshopper at time t. The value of c is defined below

$$c = C_{\max} - l\frac{C_{\max} - C_{\min}}{L} \tag{6}$$

where $C_{\max}$ is the maximum value, $C_{\min}$ is the minimum value, l indicates the current iteration, and L is the maximum number of iterations. We use $C_{\max} = 1$ and $C_{\min} = 0.00001$. The upper and lower bounds of the kth dimension are denoted as ub_k and lb_k, respectively. Also, $S()$ is a function that defines the social forces and is defined as follows:

$$S(r) = fe^{\frac{-r}{l}} - e^{-r} \tag{7}$$

where f and l are constants representing the intensity of attraction and attractive length scale, respectively, while r is a real value. The distance between the ith and jth grasshopper is denoted as d_{ij} and it is calculated as $|X_j^t - X_i^t|$. The $(X_j^t - X_i^t)/D_{ij}$ is a unit vector from the ith to the jth grasshopper. The value of the kth dimension of the fittest solution or the fittest grasshopper so far is denoted by T_k^t. Equation (5) is used repeatedly and iteratively to update the position of the grasshoppers based on the position of other grasshoppers and the fittest grasshopper found so far. This process is carried out for a fixed and pre-stipulated number of maximum iterations, L.

4) Termination: Now, we mention the termination conditions for the proposed MGO algorithm. Our algorithm terminates after running for a fixed number of predecided iterations (L). After this, we select the grasshopper with the smallest value of the fitness function, or we choose the fittest solution. Then based on this solution, we choose the features whose corresponding dimension in the fittest solution is greater than 0.5. These features form our final feature set for which we make our final classification.

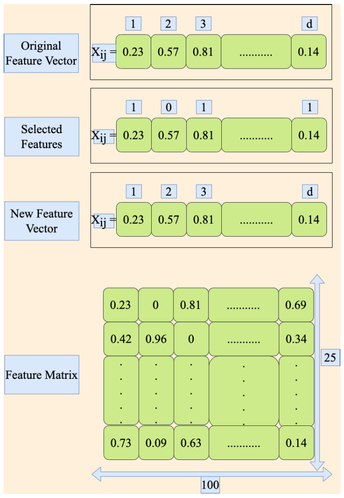


Fig. 3. Generation of 2-D feature matrix using the selected features from the feature vector of every news article.

D. Classification Using CNN

In this phase, we use deep CNNs to extract the n-gram-based features from the news articles and make the final fake news classification. We alter the feature vector of each news article using the TF-IDF of the features that are selected in the previous step and zero for the features that are not selected. To comply with the input requirements of the CNN, we convert the features of every news article into a 2-D matrix and then pass it to the CNN. The feature vector of each news article is converted into a 2-D matrix of dimension 25 × 100. Here, 25 and 100 are the dimension of our feature matrix, with 25 being the number of rows and 100 being the number of columns. This dimensionality is in accordance with the size of the feature vector, which is chosen to be 2500 based on the experimental analysis discussed ahead in Section V-A. Fig. 3 depicts the pictorial representation of generation of 2-D feature matrix using the selected features from the feature vector of every news article. We convert the feature vector into a feature matrix as an accepted input for our CNN architecture.

The feature matrix generated above is then passed through three different convolutional layers concurrently. The filter sizes of these convolutional layers are 2, 3, and 5, respectively. Different filter sizes are chosen to capture the details of the news articles based on the n-gram models. Therefore, the three convolutional layers select the 2-gram, 3-gram, and 5-gram features of the news article. These extracted n-gram features helps our model to incorporate the impact of a combination of words in signaling whether a news article is real or fake. Then we concatenate the output of these layers using a concatenation layer to generate a combined output containing the features extracted from all the convolutional layers. This output is then passed through a fully connected dense layer to finally classify the news articles as real or fake. Fig. 4 shows the process of extracting the n-gram features from the feature matrix of the news articles using a CNN which are then concatenated using the concatenation layer. The output of the concatenation layer is then passed through a dense layer and classified as real or fake depending on their characteristics.

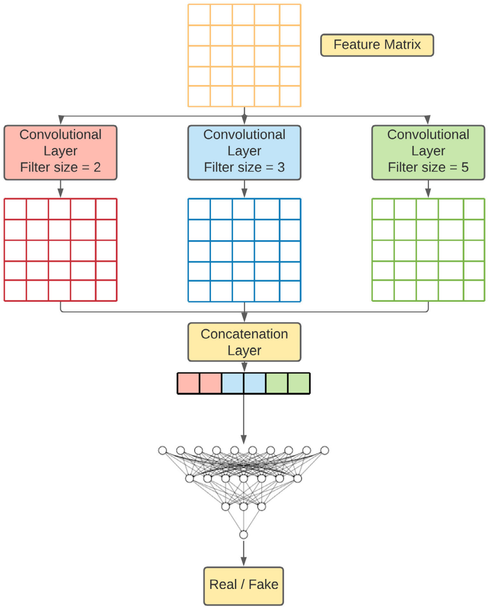

Feature Matrix
Convolutional Layer Filter size = 2
Convolutional Layer Filter size = 3
Convolutional Layer Filter size = 5
Concatenation Layer
Real / Fake

Fig. 4. Process of extraction of n-gram features from the feature matrix of the news articles using the convolutional layers and making classifying them as fake or real.

For training the classifier, we first split the entire dataset into training and testing datasets. The split is done in an 80:20 ratio, which is one of the common practices, with 80% of the dataset kept for training purposes while 20% of the dataset is reserved for testing purposes. The split has been done in such a manner so that we can prevent underfitting and overfitting. This also reduces the bias of the classifier toward a single output class. After training the classifier on the input news articles for the selected features, we test the efficiency of the proposed framework on test data.

V. Experimental Results and Analysis

In this section, we present the performed experimental result and analysis to validate the efficacy of the proposed fake news detection model. We obtained the results on all the datasets mentioned in Section III based on all the standard evaluation metrics such as accuracy, precision, recall, and F1 score. We have split the entire dataset into two parts for our experiments, namely, training and testing datasets. The split was done in an 80:20 ratio with 80% of the dataset being reserved for training while 20% of the dataset reserved for testing. The model parameters were fixed and used in similar settings across the datasets. For our MGO algorithm, we chose the number of grasshoppers to be 200 while the maximum number of iterations is chosen to be 300. For initial feature generation, as mentioned in Section IV-B, we picked the value of d to be 2500. The value of $C_{\max}$, $C_{\min}$, l, and f was taken to be 1, 0.00001, 1.5, and 0.5, respectively. For every dataset and every evaluation metrics, we run our experiments 100 times and the results were averaged out, to obtain more stable results that are free from statistical aberrations. The size of the filters for CNN is chosen to be 2, 3, and 5. We compare the performance of our algorithm with several

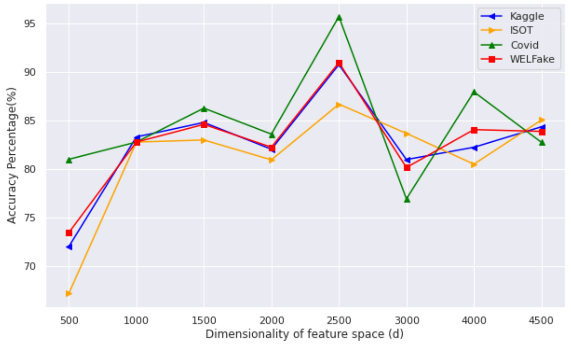


Fig. 5. Dimensionality of feature space (d) versus accuracy for various datasets.

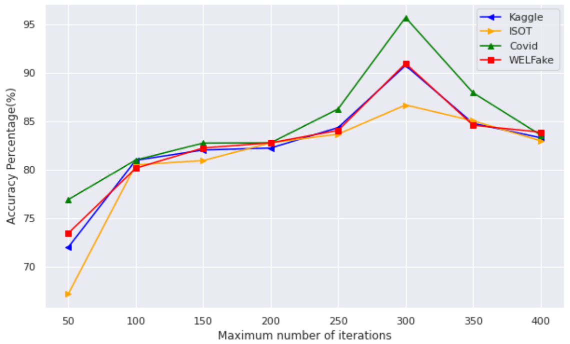


Fig. 6. Maximum number of iterations (L) versus accuracy for various datasets.

contemporary fake news detection algorithms such as CrossSEAN [20], C-BiLSTM [24], BerConvoNet [12], Semantics FND [26], and DeepFakE [27] to present the utility of our approach with other fake news algorithms. We also compare the performance of our proposed model with some of the classical, and contemporary metaheuristic algorithms such as dragonfly optimization (DGO) [28], gray wolf optimization (GWO) [29], particle swarm optimization (PSO) [30], firefly optimization (FO) [31], ant colony optimization (ACO) [32], and whale optimization (WO) [33]. This provides us an understanding of the performance comparison as opposed to other metaheuristic-based algorithms. The various simulations performed for the hyperparameters and evaluation metrics are discussed below.

A. Dimensionality of Feature Space (d) Versus Accuracy

Fig. 5 shows the effect of varying the dimensionality of feature space via TF-IDF as mentioned in Section IV-B. Here, we vary the dimensionality in steps of 500 starting from 500 and study its impact on accuracy across all the datasets. We observe three maxima at 1500, 2500, and 4000. But there is a global maximum at 2500, and the accuracy on all the datasets is maximum at 2500 only. Also, as the dimensionality of the feature space increases, the time it takes for the algorithm to run also increases. Therefore, we choose dimensionality (d) to be 2500 for the feature space of for the proposed model.

B. Parameter Setting

As part of the parametric study of the MGO algorithm, we studied the variations in the classification capability of the algorithm due to the variation in the maximum number of iterations and full population size. We evaluated the performance on all the datasets mentioned in Section III. The results obtained are as follows.

1) Maximum Number of Iterations (L) Versus Accuracy: Fig. 6 shows the impact on prediction accuracy of our proposed model based on the variation in the maximum number of iterations across all the datasets. From Fig. 6, we can observe that as the maximum number of iterations (L) increases from 50 to 100, there is a steep increase in the accuracy of our proposed MGO algorithm. But for the values of L between 100 and 250, there is a steady and almost horizontal growth. But when the maximum number of iterations is 300, we can see a peak in the performance of the proposed MGO algorithm, post which we see a decline in the performance. Moreover, as the number of iterations increases, the computational time also increases. This suggests we choose the maximum number of iterations L to be 300 as optimal. This gives us the justification for using the maximum number of iterations to be 300. It provides an optimal fake news detection accuracy across all the datasets while maintaining the computational feasibility of the process.

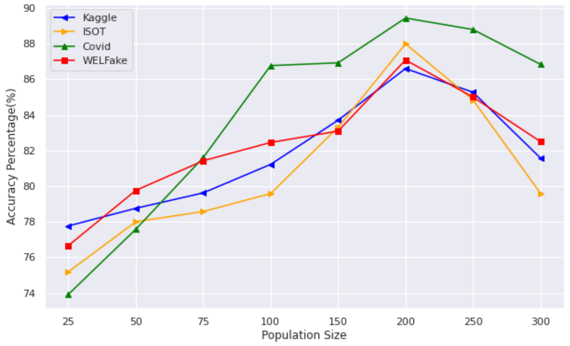


Fig. 7. Size of the population (N) versus accuracy for various datasets.

2) Size of the Population (N) Versus Accuracy: Fig. 7 depicts the impact of increasing population size on the accuracy achieved by the proposed OptNet-Fake model for fake news detection. It is evident that with an increase in the size of the population, the accuracy increases. But after reaching a threshold size of 200, the accuracy starts to drop for all the datasets. As the population size increases after 200, the number of probable solutions increases thereby increasing the complexity of the model and its performance degrades. This leads to a drop in accuracy for the model. Hence, the selected choice of population size offers an optimally efficient and effective balance in the parameters.

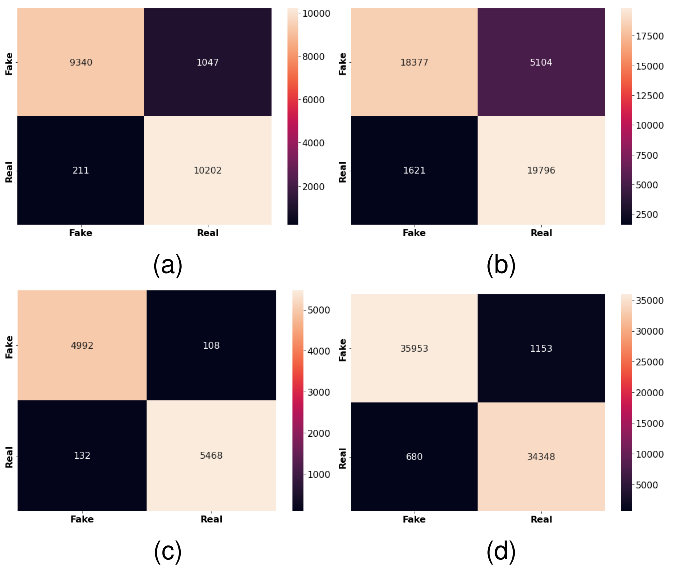
C. Confusion Matrix

Fig. 8 shows the confusion matrix obtained by our proposed algorithm for various datasets. It clearly shows the exemplary performance of the proposed OptNet-Fake model in classifying the fake news articles as fake and real news articles as real. We can see that the number of fake news articles classified

TABLE III

Performance Comparison of Our Proposed OptNet-Fake Model With Several Contemporary Fake-News Detection Methods in Terms of Accuracy (Acc.), Precision (Prec.), Recall (Rec.), and F1 Score (F1)

Methods	Kaggle				ISOT				Covid				WELFake			
	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1
Cross-SEAN	80.16	80.47	80.47	80.47	88.92	90.15	87.72	88.92	78.91	76.37	82.3	79.22	85.89	84.66	87.5	86.06
C-BiLSTM	74.6	73.98	73.98	73.98	80.73	74.8	89.81	81.62	84.22	84.34	84.34	84.34	90.2	97.91	91.04	94.35
Semantics FND	71.43	73.98	69.47	71.65	90.59	89.41	91.11	90.25	72.84	68.36	85.5	75.97	90.94	98.34	91.47	94.78
DeepFakE	75.79	72.26	86.15	78.6	70.56	65.35	92.11	76.46	89.23	86.96	90.61	88.75	93.55	95.78	97.1	96.44
BerConvoNet	96.85	82.79	82.79	82.79	69.35	61.65	94.92	74.75	89.07	86.99	91.77	89.32	93.43	99.13	93.52	96.24
OptNet-Fake	97.86	95.97	97.54	96.75	98.04	99.21	96.92	98.05	98.43	98.42	98.42	98.42	95.55	98.14	96.89	97.51



(a) (b) (c) (d)

Fig. 8. Confusion matrix comparisons of various algorithms on all the datasets chosen by us. (a) Kaggle. (b) ISOT. (c) COVID. (d) WELFake.

as fake is 9340, 18 377, 4992, and 35 953 for the Kaggle, ISOT, COVID, and WELFake datasets, respectively. Also, the number of news articles classified as real which is real by our proposed MGO algorithm is 10 202, 19 796, 5468, and 34 348 for the Kaggle, ISOT, COVID, and WELFake datasets, respectively. This shows that our model delivers good performance in detecting fake and real news articles. Fig. 8 also shows that our MGO algorithm has very few misclassifications. This is due to proper feature generation using TF-IDF and then an appropriate feature selection using the MGO algorithm.

D. Comparison With Contemporary Fake News Methods

In this section, we compare the performance of our proposed algorithm MGO-CNN with several recent contemporary fake news detection methods such as Cross-SEAN [20], C-BiLSTM [24], BerConvoNet [12], Semantics FND [26], and DeepFakE [27]. Table III shows the accuracy, precision, recall, and F1 score results obtained by various algorithms on all the different datasets used by us. Across all the datasets, we can see that the proposed model of fake news detection performs the best in terms of accuracy, precision, recall, and F1 score and gives a stable performance throughout. Semantics FND performs the worst for the Kaggle and Covid datasets. BerConvoNet and Cross-SEAN perform the worst for the ISOT and WELFake fake news datasets, respectively. It can be seen that for all the datasets, our method outperforms all the contemporary fake news detection methods by a considerable difference. This generates the utility of our proposed MGO-CNN algorithm as a benchmarked algorithm for the detection of fake news. Such excellent values of precision show the capability of our proposed approach to predict very few real news articles as fake. The high recall values also show that our proposed approach classifies a very less number of fake news articles as real. The superior performance obtained by our proposed algorithm can be understood due to the proper feature generation using TF-IDF and better exploration and exploitation capabilities of the MGO algorithm compared with other algorithms. Moreover, the proper use of a deep CNN to extract n-gram features from the text also helps our algorithm to extract the latent features from the news articles and classify the news articles optimally. The above discussion shows the utility of our proposed work in terms of fake news detection compared with other recent methods.

E. Comparison With Metaheuristic Optimization Methods

The proposed fake news detection model adopts the MGO algorithm as the feature selection method. In this section, we present the performance comparison of the adopted MGO algorithm for feature selection against some popular metaheuristic optimization algorithms in our study. For this analysis, we used all the strategies of our proposed model, namely, data processing, feature generation, and classification using CNN except the future selection methods. For future selection, we replaced the MGO in our proposed setup by some popular metaheuristic algorithms such as DGO [28], GWO [29], PSO [30], FO [31], ACO [32], and WO [33]. Table IV presents the results obtained for the various evaluation metrics obtained on all the datasets for all the chosen metaheuristic algorithms augmented in our model. From the obtained results, we can infer that our proposed fake news detection model using MGO algorithm performs the best across all the datasets for all the evaluation metrics except for precision in the WELFake dataset. In terms of precision, our model lags by a very slight margin from PSO, GWO, and ACO algorithms for the WELFake dataset. Overall, the performance of the proposed model is followed by the WO algorithm (WOA). All the other metaheuristic algorithms used follow thereby and perform closely to each other for fake news detection across the datasets and evaluation metrics. Overall, the superior results of our model compared with other metaheuristics optimization

TABLE IV

PERFORMANCE COMPARISON OF OUR PROPOSED OPTNET-FAKE MODEL WITH SEVERAL METAHEURITIC ALGORITHMS. HERE, THE MGO IS REPLACED BY OTHER METAHEURISTICS IN THE MODEL FRAMEWORK IN TERMS OF ACCURACY (ACC.), PRECISION (PREC.), RECALL (REC.), AND F1 SCORE (F1)

Methods	Kaggle				ISOT				Covid				WELFake			
	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1
DGO	75.79	72.26	86.15	78.6	70.56	65.35	92.11	76.46	62.32	62.32	82.16	70.88	90.2	97.91	91.04	94.35
GWO	81.35	78.52	85.48	81.85	69.35	61.65	94.92	74.75	72.84	68.36	85.5	75.97	85.33	98.52	84.96	91.24
PSO	80.16	80.47	80.47	80.47	80.73	74.8	89.81	81.62	78.91	76.37	82.3	79.22	93.43	99.13	93.52	96.24
FA	71.43	73.98	69.47	71.65	85.89	84.66	87.5	86.06	84.22	84.34	84.34	84.34	85.89	84.66	87.5	86.06
ACO	74.6	73.98	73.98	73.98	85.49	83.66	82.5	83.06	89.23	86.96	90.61	88.75	90.94	98.34	91.47	94.78
WOA	83.33	82.79	82.79	82.79	90.59	89.41	91.11	90.25	89.07	86.99	91.77	89.32	93.55	95.78	97.1	96.44
OptNet-Fake	97.86	95.97	97.54	96.75	98.04	99.21	96.92	98.05	98.43	98.42	98.42	98.42	95.55	98.14	96.89	97.51

algorithms justify our choice of using the MGO algorithm as a feature selection method in our model.

All the above experimental results' discussion shows that the proposed fake news detection model is the best performer across all the datasets and evaluation metrics presented in Section III. The generated feature vectors for the news articles using the TF-IDF approach, followed by proper hyperparameter tuning, enabled us to choose an optimal dimension (d) of feature vectors to capture all the important features of the news articles. Using the MGO algorithm with proper parameter tuning helped us to select the most relevant and the least correlated features from among the feature vectors. This can be attributed to the strong search space exploration and exploitation capabilities and robustness of the MGO algorithm. Also, using a deep CNN, we were able to extract the n-gram features of the news articles. Overall, we can say that the results obtained demonstrate the utility of the proposed OptNet-Fake model using MGO and CNN techniques for fake news detection.

VI. CONCLUSION

Fake news refers to false or misleading information that often leads to serious harm to individuals, organizations, and societies. In this article, we introduced a novel fake news detection model named OptNet-Fake using the MGO algorithm and CNN. We started by cleaning and processing each news article. Then we generated a d-D feature vector for every news article. The dimension of the feature vectors is chosen using proper hyperparameter tuning to capture all the essential characteristics from the news articles. The extracted features are then passed through the MGO algorithm to select the most relevant and the least correlated features. We modified the classical GOA to fit the problem of feature selection required for fake news detection. We ran experiments on four benchmarked fake news detection datasets and evaluated several popular performance metrics for the fake news detection task. We compared the performance of our model with several contemporary and metaheuristic algorithms to obtain a sense of comparative study. The obtained experimental results reveal the exemplary performance of the proposed model. As part of future work, we can explore multimedia datasets like image- or GIF-based fake news detection. For this, we can use some sophisticated deep CNN architectures to extract the features from the images or GIFs. These features can be fed to our model directly for making the classification.

REFERENCES

[1] D. Chaffey, *Global Social Media Statistics Research Summary*. 2022.

[2] W. Shahid et al., "Detecting and mitigating the dissemination of fake news: Challenges and future research opportunities," *IEEE Trans. Computat. Social Syst.*, early access, Jun. 6, 2022, doi: 10.1109/TCSS.2022.3177359.

[3] M. Babaei, J. Kulshrestha, A. Chakraborty, E. M. Redmiles, M. Cha, and K. P. Gummadi, "Analyzing biases in perception of truth in news stories and their implications for fact checking," *IEEE Trans. Computat. Social Syst.*, vol. 9, no. 3, pp. 839–850, Jun. 2022.

[4] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," *Inf. Sci.*, vol. 497, pp. 38–55, Sep. 2019.

[5] N. Kshetri and J. Voas, "The economics of 'fake news,'" *IT Prof.*, vol. 19, no. 6, pp. 8–12, Nov./Dec. 2017.

[6] P. Meel and D. K. Vishwakarma, "Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities," *Expert Syst. Appl.*, vol. 153, Sep. 2020, Art. no. 112986.

[7] K. Sharma, F. Qian, H. Jiang, N. Ruchansky, M. Zhang, and Y. Liu, "Combating fake news: A survey on identification and mitigation techniques," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 3, pp. 1–42, 2019.

[8] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *J. Econ. Perspect.*, vol. 31, no. 2, pp. 211–236, 2017.

[9] H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," *Secur. Privacy*, vol. 1, no. 1, p. e9, Jan. 2018.

[10] P. Patwa et al., "Fighting an infodemic: COVID-19 fake news dataset," in *Proc. Int. Workshop Combating Online Hostile Posts in Regional Lang. During Emergency Situation* (Communications in Computer and Information Science), vol. 1402. Springer, 2021, pp. 21–29.

[11] P. K. Verma, P. Agrawal, I. Amorim, and R. Prodan, "WELFake: Word embedding over linguistic features for fake news detection," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 4, pp. 881–893, Aug. 2021.

[12] M. Choudhary, S. S. Chouhan, E. S. Pilli, and S. K. Vipparthi, "BerConvoNet: A deep learning framework for fake news classification," *Appl. Soft Comput.*, vol. 110, Oct. 2021, Art. no. 107614.

[13] G. L. Ciampaglia, P. Shiralkar, L. M. Rocha, J. Bollen, F. Menczer, and A. Flammini, "Computational fact checking from knowledge networks," *PLoS ONE*, vol. 10, no. 6, Jun. 2015, Art. no. e0128193.

[14] S. C. R. Gangireddy, C. Long, and T. Chakraborty, "Unsupervised fake news detection: A graph-based approach," in *Proc. 31st ACM Conf. Hypertext Social Media*, 2020, pp. 75–83.

[15] F. Yang, Y. Liu, X. Yu, and M. Yang, "Automatic detection of rumor on sina Weibo," in *Proc. ACM SIGKDD Workshop Mining Data Semantics*, Aug. 2012, pp. 1–7.

[16] X. Zheng, Z. Zeng, Z. Chen, Y. Yu, and C. Rong, "Detecting spammers on social networks," *Neurocomputing*, vol. 159, pp. 27–34, Jul. 2015.

[17] H. Karimi, P. Roy, S. Saba-Sadiya, and J. Tang, "Multi-source multi-class fake news detection," in *Proc. 27th Int. Conf. Comput. Linguistics*, 2018, pp. 1546–1557.

[18] L. Wu and H. Liu, "Tracing fake-news footprints: Characterizing social media messages by how they propagate," in *Proc. 11th ACM international Conf. Web Search Data Mining*, 2018, pp. 637–645.

[19] Y. Yang, L. Zheng, J. Zhang, Q. Cui, Z. Li, and P. S. Yu, "TI-CNN: Convolutional neural networks for fake news detection," 2018, *arXiv:1806.00749*.

[20] W. S. Paka, R. Bansal, A. Kaushik, S. Sengupta, and T. Chakraborty, "Cross-SEAN: A cross-stitch semi-supervised neural attention model for COVID-19 fake news detection," *Appl. Soft Comput.*, vol. 107, Aug. 2021, Art. no. 107393.

[21] G. Shrivastava, P. Kumar, R. P. Ojha, P. K. Srivastava, S. Mohan, and G. Srivastava, "Defensive modeling of fake news through online social networks," *IEEE Trans. Computat. Social Syst.*, vol. 7, no. 5, pp. 1159–1167, Oct. 2020.

[22] A. Kumar, N. Aggarwal, and S. Kumar, "SIRA: A model for propagation and rumor control with epidemic spreading and immunization for healthcare 5.0," *Soft Comput.*, to be published, doi: 10.1007/s00500-022-07397-x.

[23] E. Sengupta, R. Nagpal, D. Mehrotra, and G. Srivastava, "ProBlock: A novel approach for fake news detection," *Cluster Comput.*, vol. 24, no. 4, pp. 3779–3795, Dec. 2021.

[24] T. E. Trueman, A. Kumar, P. Narayanasamy, and J. Vidya, "Attention-based C-BiLSTM for fake news detection," *Appl. Soft Comput.*, vol. 110, Oct. 2021, Art. no. 107600.

[25] S. Saremi, S. Mirjalili, and A. Lewis, "Grasshopper optimisation algorithm: Theory and application," *Adv. Eng. Softw.*, vol. 105, pp. 30–47, Mar. 2017.

[26] A. M. Braşoveanu and R. Andonie, "Integrating machine learning techniques in semantic fake news detection," *Neural Process. Lett.*, vol. 53, pp. 3055–3072, Oct. 2020.

[27] R. K. Kaliyar, A. Goswami, and P. Narang, "DeepFakE: Improving fake news detection using tensor decomposition-based deep neural network," *J. Supercomput.*, vol. 77, no. 2, pp. 1015–1037, Feb. 2021.

[28] S. Mirjalili, "Dragonfly algorithm: A new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems," *Neural Comput. Appl.*, vol. 27, no. 4, pp. 1053–1073, 2015.

[29] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014.

[30] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw. (ICNN)*, vol. 4, Nov./Dec. 1995, pp. 1942–1948.

[31] B. Selvakumar and K. Muneeswaran, "Firefly algorithm based feature selection for network intrusion detection," *Comput. Secur.*, vol. 81, pp. 148–155, Mar. 2019.

[32] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE Comput. Intell. Mag.*, vol. 1, no. 4, pp. 28–39, Nov. 2006.

[33] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016.

Perfectly Convergent Particle Swarm Optimization for Solving Combined Economic Emission Dispatch Problems with and without Valve Loading Effects

Devinder Kumar*
Department of Electrical Engineering,
G B Pant institute of Technology,
Delhi Skill and Entrepreneurship University, Okhla Phase III, New Delhi, India.
email:devdaksh@gmail.com
*Corresponding author

Narender kumar Jain
Department of Electrical Engineering,
Delhi Technological University,
Delhi, India.
email: vnarender84@yahoo.com

Nangia Uma
Department of Electrical Engineering,
Delhi Technological University,
Delhi, India.
email:uma_nangia@rediffmail.com

2023 International Conference on Power Electronics and Energy (ICPEE) | 978-1-6654-7058-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICPEE54198.2023.10060438

***Abstract*— The bulk of power is produced by carbon-fuelled thermal power stations, which discharge emissions like SO2, CO2, and NOx further into environment. Academics began concentrating their research work on many-objective load allocation. In order to resolve combined economic and multiple emissions dispatch scenarios with max-max price penalty component, this research introduces perfectly convergent particle swarm optimization (PCPSO) for addressing using quadratic functions, while considering the implications of emissions. Implementing this method on three different standard test systems, like the IEEE six-committed test unit system, ten generating test system, and forty generating real test system, and comparing the outcomes with other bio inspired algorithms, for the evaluation of this algorithm's effectiveness. To do this, we created a software in the MATLAB 2015a environment on hp lab-top with 4GB RAM. This technique has enhanced search tools with excellent convergence characteristics, optimizing the quadratic cost and quadratic emissions functions at diverse power demands with minimal transmission line losses. Various practical constraints are taken into account, like limits of ramp rate, restricted operating zone(s), power balancing restriction, and limits of committed system. Transmission losses taken into account when considering a multi - fuel system. This algorithm is quick, reliable, and efficient, and it requires less time to solve non-convex problems with excellent efficiency.**

***Keywords*— combined economic emission dispatch, Perfectly convergent Particle swarm optimization, price penalty factor, non-smooth cost function, quadratic emission**

I. INTRODUCTION

Fossil fuels are one of the most major part of power generation, make up a majority of global generation. Sulphur-di-oxide, nitrogen dioxide, carbon dioxide, ozone, and other hazardous pollutants and gases with particles are also released into the air, contributing to global warming. The Environmental Protection Agency (EPA) released the final Affordable Clean Energy regulation (ACE) in June 2019, repealing the Clean Power Plan for electrical power committed system. In order to save the environment from thermal power stations, a novel approach has been developed.

Numerous strategies have been put out and introduced to reduce the power system's economic dilemma. Linear programming, Lagrangian relaxation, and the Lagrange multiplier are some of the early methodologies that have been used. To enhance existing strategies, like genetic algorithm (GA)[5], evolutionary programming(EP)[5], particle swarm optimization(PSO)[1,5], Biogeography Based Optimization(BBO)[4] , harvest season artificial bee colony[7], differential evolution(DE)[6], Backtracking search algorithm(BSA)[26], Gravitational search algorithm(GSA)[23], epsilon-multi-objective genetic algorithm variable(ev-MOGA)[20], Flower pollination algorithm(FPA)[8], quasi oppositional teaching learning based optimization(QOTLBO)[10], modified artificial bee colony algorithm (MABC/D/Cat[9] , MABC/D/Log)[9] , Kernel search Optimization (KSO)[9] and more alternative generations composed by intelligent techniques have been developed. In comparison to other options, very few of the heuristic search algorithms has yet to be able to provide good enough performance to solve every optimization problems.

As a result, creating a in habitat-based heuristic search technique capable of preventing premature convergence while maintaining the rapid converging feature remains a difficulty. This problem motivated me to create such algorithm which is free from stagnation problems when used in multi-objectives problems with constraints.

Quadratic cost and Quadratic emission functions, which describe the right operational cost of producing units utilizing perfectly convergent Particle swarm optimization (PCPSO) [2], are considered in this research study. The use of this technique yielded good optimal results in a short amount of time. The following are the five sections of the paper: In section II, the formulation of the CEED optimization issue is covered. Section III discusses the PCPSO strategy. Section IV discusses the findings and debate. The conclusion is the fifth and last section.

II. Formulation of Combined Economic Emission optimization problem

The conceptual structure framework of the CEED challenge is outlined in this section, which includes the quadratic cost of fuel function model, quadratic pollutant model, and max-max price penalty function.

A. *Quadratic cost of fuel function*

The majority of the operational cost of fossil fuel-based power plants are stated as having a quadratic function because it is the system's primary goal and it also has requirements for equality and inequality.

$$Min\ F_{CF} = \sum_{i=1}^{n} a_i P_i^2 + b_i P_i + c_i + \left|\alpha_i \sin\left(\beta_i\left(P_{i,min} - P_i\right)\right)\right| \frac{\$}{h} \quad (1)$$

Liable to constraints:

Power balance restriction: The sum of entire power requirement and the losses in transmission equals the entire true power generation.

$$\sum_{i=1}^{n} P_i = P_D + P_L \quad (2)$$

Generator limit restriction: The following range must be met by the actual power output of the i_{th} committed generating unit.

$$P_{i\,min} \ll P_i \ll P_{i\,max} \quad (3)$$

Transmission loss restriction: George's equation given below, which states that the overall transmission loss $P_L$should be kept to a minimum.

$$P_L = \sum_{i=1}^{n} \sum_{j=1}^{n} P_i B_{ij} P_j \quad (4)$$

Where $F_{CF}$is the cost of fuel of entire committed units in \$/h,P_i is the true output power in MW of i_{th} unit,P_D , P_L are entire requirement and losses in transmission power in MW,$P_{i\,min}$, $P_{i\,max}$are the lower and upper power range of i_{th} unit, n is the committed units, the i_{th} committed units having coefficient of fuel cost curve are a_i, b_i, c_i, d_i and.$B_{ij}$is the coefficient of transmission loss of committed test system.

B. *Quadratic Pollution function*

Due to the combustion of fossil fuels, all thermal energy stations create hazardous gases such as SO2, NOx, and CO2, which add to the overall emissions and must be reduced individually. All three emissions are mathematically defined in this model using quadratic polynomials as follows:

$$E_{TP} = \sum_{i=1}^{n}(d_i P_i^2 + e_i P_i + f_i) + \gamma_i exp(\delta_i P_i)\ \ Kg/h \quad (5)$$

Whereas E_{TP} is the entire pollution with valve point loading effect in ton/h, d_i, e_{i,f_i} are pollution coefficients of i^{th} committed unit in ton/M $W^2h,$ton/MWh and ton/h , $\gamma_i\ and\ \delta_i$ are the valve point loading effect (VPL) pollution coefficient of i^{th} committed unit.

C. Price Penalty Factors (PPF)

The factors of Price penalty are calculated by dividing cost of fuel by pollution price and are being utilized to transform pollutant limits into comparable cost of fuel. The Max-Max price penalty factor, h_i employed in this study are listed below.

$$h_i = \frac{(a_i P_{i\,max}^2 + b_i P_{i\,max} + c_i) + \left|\alpha_i \sin\left(\beta_i(P_{i,min} - P_i)\right)\right|}{(a_i P_{i\,max}^2 + b_i P_{i\,max} + c_i) + \gamma_i exp(\delta_i P_i)} \quad (6)$$

D. Bi-objective CEED

The bi-objective CEED equations are shown below, which incorporate cost of fuel with each pollutant and are then transformed to a single outcome by multiplying a factor of price penalty for each of the pollutants independently.

$$F_T = \sum_{i=1}^{n}[(a_i P_i^2 + b_i P_i + c_i) + \left|\alpha_i \sin\left(\beta_i\left(P_{i,min} - P_i\right)\right)\right| + h_i\left(a_i P_i^2 + b_i\ P_i + c_i\ + \gamma_i exp(\delta_i P_i)\right)] \frac{\$}{h} \quad (7)$$

III. Particle Swarm Optimization:

Kennedy and Eberhart [1] established Particle swarm optimization without inertia weight in 1995, but for the first time in 1998, they developed it with constant inertia weight, and as a result it was dubbed as traditional PSO. Started with candidate solutions of particles moving through the problem area, every particle having a location and velocity and have latest updates as below:

$$x_j(k+1) = x_j(k) + v_j(k+1) \quad (8)$$

$$v_j(k+1) = \omega v_j(k) + c_1(K) r_1\left(p_j(k) - x_j(k)\right) + c_2(K) r_2\left(g(k) - x_j(k)\right) \quad (9)$$

Where, j =1, 2, 3 … n

$$\omega(K) = \omega_{max} - k \times (\omega_{max} - \omega_{min}) \div K_{max} \quad (10)$$

$$C_1(K) = 1.167 \times \omega(K)2 - 1.167 \times \omega(K) + 0.66 \quad (11)$$

$$C_2(K) = 3 - C_1(K) \quad (12)$$

$C_1(K), C_2(K)$ are non-linear dynamic changeable , asynchronous learning factors and are more prone of converging to best possible ideal optimal result.

k+1 stands subsequent epoch, k is the momentary epoch , v_j is the particle j's velocity ,x_j is particle j's location , factor of Inertia weight ω ,acceleration factors c_1 ,c_2, personal best particle p_j ,global best of the whole network swarm g, pseudo-random numbersr_1,$r_2$between 0 and 1. ω_{max} and ω_{min} are highest and lowest range value 0.9 and 0.4.

Postulated algorithm as Perfectly Convergent Particle Swarm Optimization (PCPSO)

One of the purpose of the proposed variant [2] in this scenario is to eliminate early convergence, because it contributes toward standstill and thus to substitute personal best particles for global particles because these allow for more exploration in search space. I've introduced an additional particle in this new form as with GCPSO [2], however it will seek area close to a global position for personal best position, while taking into consideration the current velocity update, lacking exploration and subjects to entrapment in many-modal situations with single or more local minima:

$$v_j'(k+1) = -x_j{}'(k) + pbest(k) + \omega v_j'(k) + \rho(k)(1 - 2r) \quad (13)$$

Other particles in the swarm, on the other hand, will adjust their velocity according to this new variant:

$$v_j'(k+1) = \omega(k)x_j{}'(k) + c_1 r_1\left(p_j(k) - x_j(k)\right) + c_2 r_2\left(-x_j(k)\right) \quad (14)$$

$$\omega(k) = \omega_{max} - k \times (\omega_{max} - \omega_{min}) \div K_{max}$$

$$C_1(k) = 1.167 \times \omega(k)2 - 1.167 \times \omega(k) + 0.66$$

$$C_2(k) = 3 - C_1(k)$$

$C_1(k)$ and $C_2(k)$ are non-linear dynamic changeable , asynchronous learning factors and are more prone of converging to best possible ideal optimal result..

In which the, $-x_j{}'(k) + pbest(k)$ part will conducts an investigation in personal best area, $\omega v_j'(k)$ will provide the momentum to conduct hunt in present trajectory, ρ(k) (1-2r) chooses a unique random check in the vicinity of personal best particle with only a distance equal to $2\rho(k)$, ρ(k) is the random stochastic hunt space diameter expressed below:

$$\rho(k+1) = \begin{cases} 2\rho(k) & Successes > sc \\ (0.5)\rho(k) & failure > fc \\ \rho(k) & otherwise \end{cases} \quad (15)$$

#successes (k+1)> #successes (k) => # failures (k+1) =0 and, # failures (k+1)> # failures (k) => #successes (k+1) =0

The threshold settings s_c= 15 & f_c=5 can finely tuned where #successes & #failures are the amount of subsequent unsuccessful attempts or successes.

This method uses an adaptable to choose the ideal sampling volume in its current iteration. The maximum range travelled in a single movement can be increased if a particular value of consistently produces a favourable outcome. The sampling volume needs to be reduced where, but at the other hand it delivers numerous failure in a row. If > 0 for all stages, at the end of the day, there won't be a halt.

This variant, in essence, enables every particles to compete, irrespective of if they're in exploratory phase or have a superior personal best than that of previous epoch or sit on the edge of global optimum, resulting in a true global technique. This technique gets over GCPSO's restrictions.

IV. PCPSO EXECUTION STEPS IN CEED PROBLEM

Step1. Specify the lower and upper limitations for every committed unit's output, as well as the demand in load restrictions.

Step2. Using the $i^{th}\, particle$ $P_i = [(P_{i1}^n, P_{i2}^n, P_{i3}^n \dots P_{iN}^n)]$ where i=1, 2...S. create particles at random in between min and max operating range of the size N for a population of size "S" in the j-th-dimensional space. S with 'r' is a random number which is uniformly distributed within 0 and 1 and must satisfy the generation limitations requirement (15).

$$P_{ij}^n = P_{min} + r\left(P_{ij\,max} - P_{ij\,min}\right) \quad (16)$$

Step3: Constraints imposed by restricted operating zones

The beginning population (or revised population) are adjusted and given output value close to the zone's (P_{ij}^{lower}) or higher (P_{ij}^{upper}) boundary, as indicated by the criteria, if any element P_{ij} is found to be inside the k^{th} forbidden operational zone. The centre of the k^{th} restricted area is $P_{mid,k}$.

$$P_{ij} = \begin{cases} P_{ij}^{lower} & if\ P_{ij}^{lower} \le P_{ij} < P_{mid,k} \\ P_{ij}^{upper} & if\ P_{mid,k} \le P_{ij} < P_{ij}^{upper} \end{cases} \quad (17)$$

Step4. Set particle velocity in the [$v_i^{min} v_i^{max}$] in N-dimensional space.

Step5. Utilize the equation (1, 5, 7.to assesses the fitness of each individually.

Step6: The parameters are iteratively changed to improve fitness. The parameters of PSO are updated using equations (8-15).

Step7: The evaluation function values for the changed particle positions are found. PCPSO assigns the new output value to pbest if it is good than prior one. The value of gbest's is also changed to show to that it is the optimal vector across pbest.

Step8. Stop criteria

Particles stop moving if equation (17) is below the standstill threshold of $\varepsilon = 1x10^{-6}$(if the position of the particles for the ideal solution is given as Gbest).

V. SIMULATION RESULTS AND DISCUSSION

CEED issues are solved using the proposed PCPSO methodology using three different test platforms and assessed on three different IEEE and real test unit generating stations with six, ten and forty units, including losses in variability transmission and other restrictions. The constants of the proposed PCPSO includes the particle quantity be 20, total epoch be 250, trails is 5, the linearly decreasing inertia weight with higher and lower inertia range w=0.9 to 0.4, acceleration constants are c1 = c2 =2, characteristics of suggested PCPSO.

Test case system 1

To highlight the efficiency of the PCPSO approaches for addressing the CEED optimization issue with line flow limits, the investigation and validation CEED issues with IEEE 30-bus, 6-generator system [3] at a demand load of 283.4 MW. Using the PCPSO approach including all system limitations, optimal generator scheduling was accomplished. All buses have a voltage range of 0.94 to 1.06 per unit, with a maximum voltage variation of 6%.The suggested PCPSO algorithm is contrasted with latest research papers algorithms BBO, GA, EP, PSO and DE with lowest F_{CEED} of 2004.30$/h, cost of fuel F_C of 838.15$/h and cost of emission is 335.10$/h as shown in table 1.The simulation results shows power loss P_L of 5.78MW with computational time of 6.08seconds resulting in excellent convergence characteristics.

Table1.Comparision of CEED results for six generating system with other techniques

	PCPSO	BBO [4]	GA [5]	EP [5]	PSO [5]	DE [5]
P_1	120.28	127.84	58.41	114.29	107.73	121.65
P_2	48.40	42.41	76.27	50.38	46.60	56.58
P_3	30.62	31.19	47.82	30.19	27.93	36.30
P_4	31.44	33.27	33.44	32.78	35.00	28.91
P_5	29.00	29.97	28.75	29.36	30.00	22.88
P_6	29.43	29.40	39.98	30.66	40.00	21.90
F_{CEED}	2004.30	2084.78	2107.19	2094.39	2109.47	2122.53

Test case system 2.

This scenario involves a thermal committed system with 10 units of generation along valve point loading effects. The cost of fuel coefficients matrix, generator constraint matrix, pollution coefficient matrix, and coefficient of transmission loss matrix are taken from [6]. Table 2 displays the outcomes of using PCPSO to solve CEED optimization problem for a 2000MW load demand and contrasts them along many more approaches. The proposed PCPSO algorithm shows the lower fuel cost with emission cost along with lowest combined economic emission dispatch of 216166.43 $/h which is lower than BSA, MODE, PDE, GSA, QOTLBO, NGPSO, FPA, ev-MOGA, ABCDP-LS, by 2948.85$/h, 2015.07$/h, 1701.25$/h, 1687.24$/h, 1624.31$/h, 4.11$/h, 2761.69$/h and 1848.61 $/h respectively in achieving optimal global minimum solution in low iteration and computing time. This shows the excellent performance of PCPSO without getting trapped in minima solution

Table II. Shows comparison of cost of fossil fuel, pollutant price and F_{CEED} of PCPSO with the other optimization methods.

	BSA [26]	MODE [6]	PDE [6]	GSA [27]	QOTLBO [10]	NGPSO [28]	FPA [8]	ev-MOGA [20]	PCPSO
P_1	55.00	54.9487	54.9853	54.9992	55.0000	55.00	53.188	54.1807	55.00
P_2	80.00	74.5821	79.3803	79.9586	80.0000	80.00	79.975	78.4981	80.00
P_3	86.53	79.4294	83.9842	79.4341	84.8457	81.2398	78.105	84.7653	78.2489
P_4	86.98	80.6875	86.5942	85.0000	83.4993	80.8334	97.119	81.3502	81.8443
P_5	129.15	136.89551	144.4386	142.1063	142.9210	160.00	152.740	138.0526	157.8900
P_6	146.92	172.6393	165.7756	166.5670	163.2711	235.0087	163.080	166.2667	231.8700
P_7	300.00	283.8233	283.2122	292.8749	299.8066	289.3507	258.610	295.466	290.0735
P_8	323.90	316.3407	312.7709	313.2387	315.4388	297.4542	302.220	326.7642	299.2467
P_9	435.99	448.5923	440.1135	441.1775	428.5084	401.5072	433.210	428.9338	403.6784
P_{10}	440.01	436.4287	432.6783	428.6306	430.5524	401.4275	466.070	429.6309	403.5242
F_C	112807.37	1.1348e5	1.1351e5	1.1349e5	113460	116179.6487	1.1337e5	113422.34	113360.46
E_C	4188.09	4124.9	4111.4	4111.4	4110.2	3939.2278	4147.17	4120.5204	3910.78
F_{CEED}	219115.28	218181.5	217867.68	217853.24	217790.74	216170.54	218928.12	218015.04	216166.43

Test case system 3.

The Tai power plant itself has 40 generating units and is a significant and diverse energy production system, with coal-fired, oil-fired, gas-fired, diesel, and combined cycle all being present [6]. The system's demand for load is 10500 MW, which is influenced by limits in ramp rate (RRL), banned operating zones (POZ), cost function of non-smooth nature with valve point effects, and emission function. The PCPSO's best simulation results as shown table 3. are compared to with MODE, PDE, MABC/D/Cat, MABC/D/Log, FPA, KSO, QOTLBO, GQPSO, SAIWPSO, PSOGSA , MA θ-PSO, HPSOGSA, IABC, GSA, NSGA-II , SPEA2, IABC-LS , TLBO , ev-MOGA , ABCDP , MBFA , DE-HS , MLTBO , RCCRA , BPO and OHS is lowest in fuel cost 12430.00 \$/h along with emission cost . The F_{CEED} is 220810.00 \$/h with low power loss P_L of 81.59MW which is also lower from the results in a very small computation time of 3.16 seconds in few iterations. Following table III contrasts the comparison of cost of fossil fuel and emission cost of all the recent algorithms

Table III. Cost of fuel F_C and pollution cost E_C of PCPSO is compared to other latest algorithms.

Algorithm	**Fuel Cost F_C**	**Emission Cost E_C**
PCPSO	12430.00	76610.00
QOTLBO[10]	125161.00	206490.00
GQPSO[11]	146121.50	270192.37
SAIWPSO[12]	121676.23	177276.36
PSOGSA[13]	129987.00	176678.00
MA θ-PSO[14]	129995.00	176682.00
HPSOGSA[15]	129997.00	176684.00
IABC[16]	129995.00	176682.00
GSA[17]	125782.00	210933.00
NSGA-II[18]	125825.00	210949.00
SPEA2[19]	125808.00	211098.00
IABC-LS[16]	12995.00	176682.00
TLBO[25]	125602.00	206648.00
ABCDP[16]	129995.00	176682.00
ABCDP-LS[16]	129995.00	176682.00
MBFA[21]	129995.00	176682.00
DE-HS[22]	129994.00	176682.00
MLTBO[23]	127283.87	99127.70
RCCRA[24]	124250.95	229395.90
BPO[25]	127335.40	97848.41
MODE [6]	12579.00	211190.00
PDE [6]	12573.00	211770.00
MABC/D/Cat[7]	12490.90	256560.67
MABC/D/Log[7]	12449.10	256560.00
FPA [8]	123170.00	208460.00
KSO [9]	125491.00	199591.00

VI. Conclusion

In order to solve CEED issues in power systems, PCPSO has been created in this study. The effectiveness of the PCPSO was evaluated for a number of test cases and contrasted with the recent research papers. It is confirmed that PCPSO is preferable an alternative algorithms for solving CEED issues, especially in large-scale power systems with valve point impact. Additionally, PCPSO shows avoiding to get struck in the premature convergence in local minima which results in better economic and emission impact, computational effectiveness, and its convergence feature. As a result, PCPSO optimization is a viable method for addressing challenging issues in power system networks. Future application of the suggested strategy to many-area power networks systems combined with solar power systems are part of the works for future application. The PCPSO excels at handling problems involving bi- and multi-objective power system optimization issues along outstanding outcomes in a shortest possible time and iterations.

References:

[1]J. Kennedy and R. Eberhart. “Particle swarm optimization”, in Proceedings of the IEEE International Conference on Neural Networks .IEEE Service Centre, Piscataway, NJ, IV, pp.1941-1948, 1995

[2]Devinder Kumar, N. k. Jain, Uma Nangia, ”Perfectly convergent Particle swarm optimization in multi-dimensional space”, International journal of Bio-inspired computation, Inder Science Publications, Vol 18,no 4,p.221228,2021.

[3] Venkatesh P, Gnanadass R and Padhy Narayana Prasad, “Comparison and application of evolutionary programming techniques to combined economic emission dispatch with line flow constraints”, IEEE Trans Power System, Vol. 18, Issue 2, pp. 688–697, 2003.

[4] Sunil Kumar Goyal, Neeraj Kanwar ,Jitendra Singh ,Manish Shrivastava ,Amit Saraswat, O. P. Mahela, Economic Load Dispatch with Emission and Line Constraints using Biogeography Based Optimization Technique”, International Conference on Intelligent Engineering and Management (ICIEM),PP.470-76,2020.

[5] I. Jacob Raglend, Sowjanya Veeravalli, Kasanur Sailaja, B. Sudheera and D.P. Kothari, “Comparison of AI techniques to solve combined economic emission dispatch problem with line flow constraints”, Electrical Power and Energy Systems, Vol. 32, pp. 592–598, 2010.

[6] M. Basu, ‘‘Economic environmental dispatch using multi-objective differential evolution,’’ Appl. Soft Computing., Vol. 11, no. 2, pp. 2845–2853, Mar. 2011.

[7] D. C. Secui, A new modified artificial bee colony algorithm for the economic dispatch problem, Int. J. Energy Convers. Management. Vol 89, pp. 43-62, 2015.

[8] A. Abdelaziz, E.Ali, and S.A. Elazim, ‘‘Combined economic and emission dispatch solution using flower pollination algorithm,’’ Int. J. Elect. Power Energy Syst., vol. 80, pp. 264–274, Sep. 2016.

[9] Ruyi Dong, Shengsheng Wang," New Optimization algorithm inspired by Kernel tricks for the Economic emission dispatch problem with valve point," Special section on AI technologies for electric power system, doi 10.1109/ACCESS.2020.2965725.

[10] P.K. Roy and S. Bhui, ''Multi-objective quasi oppositional teaching learning based optimization for economic emission load dispatch problem,'' Int. J. Elect. Power Energy Syst., vol. 53, pp. 937–948, Dec. 2013.

[11] L.D.S. Coelho, ''Gaussian quantum-behaved particle swarm optimization approaches for constrained engineering design problems,'' Expert Syst. Appl., vol. 37, no. 2, pp. 1676–1683, Mar. 2010.

[12] M. Taherkhani and R. Safabakhsh, ''A novel stability-based adaptive inertia weight for particle swarm optimization, '' Appl. Soft Computing. , vol. 38, pp. 281–295, Jan. 2016.

[13] S. Jiang, Z. Ji, and Y. Shen, ''A novel hybrid particle swarm optimization and gravitational search algorithm for solving economic emission load dispatch problems with various practical constraints,''Int. J. Electrical. Power Energy Syst., vol. 55, pp. 628–644, Feb. 2014.

[14]T. Niknam and H. Doagou-Mojarrad, ''Multi objective economic/emission dispatch by multi objective thetas-particle swarm optimisation,'' IET Generation., Transmission. Distribution, vol. 6, no. 5, pp. 363–377, May 2012.

[15] S. Jiang, Z. Ji, and Y. Shen, ''A novel hybrid particle swarm optimization and gravitational search algorithm for solving economic emission load dispatch problems with various practical constraints, ''Int. J. Electrical. Power Energy Syst., vol. 55, pp. 628–644, Feb. 2014.

[16] D. Aydin, S. Özyön, C. Yaşar, and T. Liao, ''Artificial bee colony algorithm with dynamic population size to combined economic and emission dispatch problem, ''Int. J. Electrical. Power Energy Syst., vol.54, pp.144–153, Jan. 2014.

[17] M. A. Ajzerman, E. M. Braverman, and L. I. Rozonoehr, ''Theoretical foundations of the potential function method in pattern recognition learning,'' Automat. Remote control, vol. 25, no. 6, pp. 821–837, Jan. 1964.

[18] B. Qu, J. Liang, Y. Zhu, Z. Wang, and P. Suganthan, ''Economic emission dispatch problems with stochastic wind power using summation based multi-objective evolutionary algorithm,'' Inf. Sci., vol. 351, pp. 48–66, Jul. 2016.

[19] M. Basu, ''Economic environmental dispatch using multi-objective differential evolution,'' Appl. Soft Computing., vol. 11, no. 2, pp. 2845–2853, Mar. 2011.

[20] E. Afzalan and M. Joorabian, ''Emission, reserve and economic load dispatch problem with non-smooth and non-convex cost functions using epsilon-multi-objective genetic algorithm variable,'' Int. J. Electrical. Power Energy Syst., vol. 52, pp. 55–67, Nov. 2013.

[21] P. Hota, A. Barisal and R. Chakrabarti, ''Economic emission load dispatch through fuzzy based bacterial foraging algorithm,'' Int. J. Elect. Power Energy Syst., vol. 32, no. 7, pp. 794–803, Sep. 2010.

[22] S. Sayah, A. Hamouda, and A. Bekrar, ''Efficient hybrid optimization approach for emission constrained economic dispatch with non-smooth cost curves, ''Int. J. Electrical. Power Energy Syst., vol.56, pp.127–139, Mar.2014.

[23] Sharmila Deve Venkatachalam, Keerthivasan Krishnamoorthy, Ahmed Said Ahmed Al-Shahri, Balachander Kalappan, Modified Teaching Learning Based Optimization Technique to Achieve the Best Compromise Solution of Economic Emission Dispatch Problem", Journal of Green Engineering (JGE) ,Vol-10, Issue-5,pp. 2458–2482, 2020.

[24] Kuntal Bhattacharjee, Aniruddha Bhattacharya, Sunita Halder nee Dey, "Solution of Economic Emission Load Dispatch problems of power systems by Real Coded Chemical Reaction algorithm", Electrical Power and Energy Systems, Vol. 59, pp.176–187, 2014.

[25] Rajasomashekar S, Aravindhababu S, "Biogeography based optimisation technique for best compromise solution of economic emission dispatch", Swarm and Evolutionary Computation, Vol. 7, pp.47-57, 2012.

[26] Mostafa Modiri-Delshad, Nasrudin Abd Rahim, "Multi-objective backtracking search algorithm for economic emission dispatch problem", Applied Soft Computing, Vol 40,pp. 479-494,2016.

[27] Güvenc, Ugur, Y. U. S. U. F. Sönmez, Serhat Duman, and Nuran Yörükeren. "Combined economic and emission dispatch solution using gravitational search algorithm." Scientia Iranica 19, no. 6, pp. 1754-1762.2012.

[28] Dexuan Zou, Steven Li, Zongyan Li, Xiangyong Kong, "Ä new global particle swarm optimization for the economic emission dispatch with or without transmission losses", Energy conversion and management, Vol 139,pp.49-70,2017.

Performance Analysis of Solar PV Modules with Dust Accumulation for Indian Scenario

Komal Singh
Department of Electrical Engineering
Delhi Technological University
Delhi, India
komalsingh2581998@gmail.com

M. Rizwan
Department of Electrical Engineering
Delhi Technological University
Delhi, India
rizwan@dce.ac.in

2022 IEEE 10th Power India International Conference (PIICON) | 978-1-6654-5930-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/PIICON56320.2022.10045127

Abstract— **Solar energy has proven to be an assured front runner among renewable energy sources since it is clean, cost-effective and environment friendly. The output power and lifespan of a photovoltaic module are governed by a variety of factors such as incident solar radiation intensity, cell temperature, cloud and other shading effects, dust accumulation, weather conditions, geographical location, module orientation, etc. This work examines the impact of dust accumulation on the performance of the 5 kW photovoltaic system installed on the rooftop of the laboratory at Delhi Technological University. Performance analysis of the 5 kW photovoltaic system is carried out over 62 days such that the panels were left naturally uncleaned for the first 31 days and then cleaned on a regular basis for the following 31 days. Performance parameters such as performance ratio, capacity factor, system energy yield and reference energy yield are derived. The performance analysis results of the practical 5 kW system were later compared with the PVsyst software results.**

Keywords— Grid Interactive Photovoltaic System, Dust Accumulation, Performance Parameters, PVsyst Software

I. Introduction

Despite the severe disruptions brought on by the global pandemic and the ensuing GDP collapse, solar and wind capacity expanded by a colossal 238 GW in 2020 – almost double its previous highest annual increase [1]. This growth trend is in line with the decarbonization goals promoted by Agenda 2030. There are a variety of factors that contribute to soiling, including sand and dust particles, bird droppings, snow etc; that reduce the efficiency of solar panels [2]. The energy production by photovoltaic systems is severely conditioned by the abnormal operating conditions caused by dust depositions on the surface of PV panels.

In order for photovoltaics to be efficient, a considerable amount of unalterable and alterable factors must be taken into account; dust is one such unalterable factor that significantly reduces PV module efficiency [3]. It is imperative that PV panels needs to be cleaned frequently depending on their site location, as the Saharan region has strong solar energy potential, but is plagued with sand and dust accumulation, wind as well as high temperatures [4].

In [5], the effect of dust on the solar panel was analyzed for different modular technologies; the impact of dust on voltage, power and current of the respective modules was studied and quantitative performance degradation was observed. Power performance can be significantly reduced up to 60-70% due to the deposition of dust on the surface of PV panel [6]. The size and shape of dust particles accumulated vary with location [7] and hence type of dust soiling can significantly have a varying effect on PV performance [8].

The deposit of coal dust on PV panels causes a significant performance drop compared to that caused by fly ash [9], gypsum, and fertilizer industry dust [6].

IEC-61724-3 provides guidelines on evaluation methods and data collection for performance of long-term capacity and short-term system, it also outlines guidance for evaluation of performance over the full range of operating conditions [10]. IEC 61724-2 provides the evaluation of power output during reference conditions (a few relatively sunny days) [11]. Performance ratio (PR) measures the quality of PV system and hence is also popular as quality factor in the solar energy sector [10]. Performance analysis can solely be used for making the right decisions for current and future installations [12]. In [13] authors calculated the performance ratio of the 6 kW PV system for four months. Performance ratio (PR) is a worldwide recognised indicator used by countries like Australia, the US and European countries to judge the performance of photovoltaic plants. With the help of such an analysis, these countries were able to improve the efficacy of their PV plants by identifying the deficiency and thus planned for smarter investment choices. The Performance ratio has a better predictive value than the PV plant's final yield because it closely accounts for actual solar radiation [14]. EU performance report signifies that a well performing system should have a performance ratio of 0.8 and higher [15]. As per SM, performance ratio measures the percentage of the energy that can be exported to the grid, after subtracting arbitrary energy losses and consumption [16]. Weather variability is a strong determinant of PR regardless of the plant's location or system size. [17]-[19] presents the performance ratio overview for different countries. In [20], the authors reviewed the effect of dust on the performance of photovoltaic panels in the North Africa, Middle East, and the Far East region. Due to limited water resources and high dust accumulation rates throughout the year, most countries in these regions have always struggled with cleaning their PV panels [21]. Henceforth, location plays a crucial role in analysing the effect of dust on the performance of PV panels.

Capacity factor (CF) can also be used as a performance indicator for grid-connected photovoltaic systems, but due to its lack of consideration for threshold irradiation, it cannot offer a realistic representation of PV plant performance. Moreover, it does not take into account environmental factors, grid availability, and system faults and hence crippling its performance analyzing potential [14]. Sometimes, capacity factor is used by investors to estimate the return on investment of their solar PV systems.

PVsyst software allows the study, sizing, modelling and analysis of photovoltaic systems. PVsyst features a database of solar system component data and diverse meteorological

978-1-6654-5930-3/22/$31.00 ©2022 IEEE

data sources such as meteonorm, NREL etc. [22]-[24] presents PVsyst simulation of photovoltaic system for different locales. It is typically used to identify the optimal tilt angle, azimuth angle, and PV module and inverter for improving PV plant performance and minimizing losses [24]. This study presents the performance analysis of the 5 kW photovoltaic system installed on the rooftop of the laboratory at Delhi Technological University for May and June, 2022. The panels were left naturally dirty in May whereas regularly cleaned in June. System energy (kWh), peak power (Wp) and hours of operation of the 5 kW photovoltaic system were collected daily from the inverter unit in May and June. A few days of April and July were also added to May and June data respectively due to the PV system being inoperable for 5 days in May. The reference energy yield (Y_R), system energy yield (Y_S), performance ratio (PR), capacity factor (CF) and peak power (W) are compared for both months. An identical 5 kW photovoltaic system was also simulated in PVsyst, and the performance analysis results were compared to the practical PV system.

Further, this paper consists of different sections; section II describes the analytical approach used for performance analysis and PVsyst analysis, ratings of modules and inverters are defined in section III, followed by performance analysis results and PVsyst results in section IV, and finally conclusion is drawn in section V.

II. Methodology

A. Performance Analysis

The performance ratio (PR) is independent of location and is often described as the quality factor since it efficiently measures the quality of the PV plant. The relationship between the desired and actual energy outputs from a PV system is described by the performance ratio (PR). PV plant energy is heavily dependent on insolation from the sun, and performance ratio accounts for this global incident irradiation hence it becomes a powerful performance assessing tool.

$$Y_R = \frac{\text{Measured insolation in kWh/m}^2}{\text{Reference irradiation in kW/m}^2} \tag{1}$$

$$Y_S = \frac{\text{AC Energy output in kWh}}{\text{Nameplate rated output in kW}} \tag{2}$$

$$PR = \frac{\text{System Energy Yield } (Y_S)}{\text{Reference Energy Yield } (Y_R)} \tag{3}$$

Capacity factor (CF) for the photovoltaic system is defined as the ratio of energy produced (kWh_{AC}) from the PV plant divided by the theoretical peak energy production of the PV plant throughout a specific period. But this factor does not serve as an ideal performance judging factor since it does not include the insolation from the sun.

$$CF = \frac{\text{Actual Energy output of system in kWh}}{24 * \text{ Nameplate rated output in kW}} \tag{4}$$

The performance ratio and capacity factor can also be expressed as percentages. For solar PV systems, PR values typically range from 60-90% and capacity factor values between 10-25%.

B. PVsyst Analysis

PVsyst software is a comprehensive solar design tool used by thousands of researchers and engineers across the globe. PVsyst is becoming the standard for large as well as utility-scale photovoltaic installations. It is sometimes also used to study the performance degradation of the already installed PV systems. The main components of the software are project design, simulation and utilities.

Project design comprises of grid-connected, standalone and pumping system. The software includes main parameters such as system, orientation, and detailed losses; as well as optional parameters such as near shading and economic evaluation. The system provides users with the ability to configure PV system power and choose appropriate PV modules, inverters, and batteries.

The simulation is performed over a full year in hourly steps and produces a comprehensive report that includes graphs, tables and diagrams.

Utilities include databases and tools. Databases consist of monthly and hourly climatic data. Climate data that is compatible with PVsyst can also be imported from external sites. Datasheets of unlisted PV modules, inverters and batteries can also be added. With the help of tools, the behaviour of a PV installation can be quickly estimated and visualized and it also provides access to compare PV installations and simulation by importing measured data of existing PV installations.

Solar radiation data can be imported into PVsyst, or there is built-in Meteonorm, NASA, Solcast, and NREL data. A library of PV modules, inverters, batteries, pumps, and generators from various manufacturers is built in. If unlisted, there is a provision to import or edit the existing datasheet of PV modules, inverters etc.

III. SYSTEM DESCRIPTION

The MPPT-based inverter can provide us with the daily system energy (kWh), daily peak power (W) and daily hours of operation of the 5 kW photovoltaic system.

The 5 kW photovoltaic system consists of 20 modules, each with a rated output of 250 Wp connected in two parallel strings, with 10 modules in each string. Table II below show the ratings of the 250 Wp module at STC.

TABLE I. 250 Wp MODULE RATINGS AT STC

Cell type	Polycrystalline
Rated output (P_{mpp})	250 Wp
Rated voltage (V_{mpp})	30.48 V
Rated current (I_{mpp})	8.21 A
Open Circuit Voltage (V_{oc})	37.47 V
Short circuit current (I_{sc})	8.81 A
Module Efficiency	15.4 %
Number of Cells	60
Module size	1640*992*35 mm
Number of diodes	03

TABLE II. 5.5 KW_{AC} INVERTER RATINGS

Max. AC output active power	5.5 kW
MPP Voltage Range	200-900 V
Maximum Input Voltage	1000 V
Maximum Input Current	2*11 A
I_{SC} PV (absolute maximum)	2*16.5 A
Maximum Continuous Output Current	3*8.5 A
Rated Grid Voltage	380/400 V

The Table I shows the ratings of the 250 Wp module. Table II defines the ratings of the 5.5 kWp inverter. Datasheets of PV module and inverter as shown in Table I and Table II respectively along with tilt angle and azimuth angle as shown below in Fig. 1 were imported into the PVsyst for simulation. Delhi Technological University is located at 28.7495^0 N and 77.1184^0 E. The latest NASA daily global irradiation data of the above location is used for performance ratio (PR) calculation.

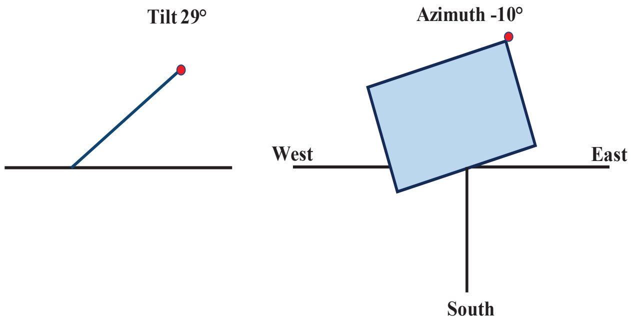


Fig. 1. Tilt angle and Azimuth angle in PVsyst

TABLE III. PVSYST PARAMETERS

Field Type	Fixed-tilted plane
Number of 250 Wp modules	20
Module area	33 m^2
Nominal PV power	5 kWDC
Nominal AC power	5.5 kWAC
Modules in series	10
Number of Strings	2
Site Latitude	28.7495^0 N
Site Longitude	77.1184^0 E
Altitude	300 m

PVsyst parameters after importing the datasheets of PV module and inverter are shown in the above Table III.

IV. RESULTS AND DISCUSSION

The data is gathered from the inverter unit over 62 days and the 5 kW PV system is working daily for 10-14 hours. The panels were left naturally dirty in May, whereas the panels were regularly cleaned in June and since the PV system was inoperable for 5 days in May, hence the data includes some dates from April and July. The below Figs. 2-7 depicts the comparison of reference energy yield (Y_R), system energy yield (Y_S), performance ratio (PR), capacity factor (CF) and peak power (W).

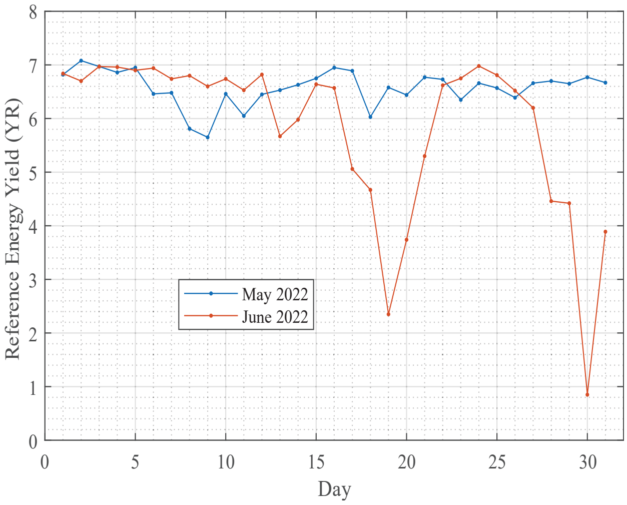


Fig. 2. Comparison of Reference Energy Yield of both months

Fig. 2 above illustrates the comparison of the reference energy yield of both months and it can be observed that due to frequent rainfall, there was a sudden decrement in reference energy yield for a few days in June and consequently panels were naturally getting cleaned and their temperature also dropped down. It can also be observed that May month has a higher potential for solar energy as compared to June.

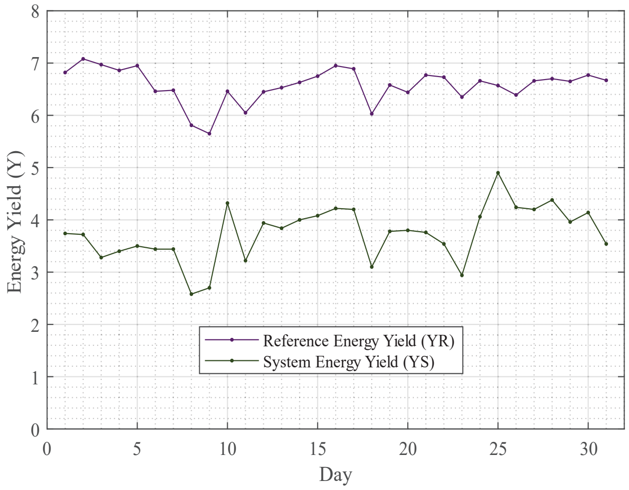


Fig. 3. Comparison of Reference Energy Yield (Y_R) and System Energy Yield (Y_s) in May 2022

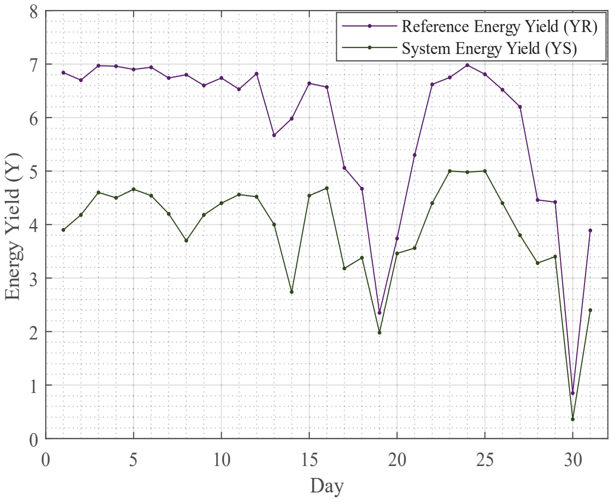


Fig. 4. Comparison of Reference Energy Yield (Y_R) and System Energy Yield (Y_s) in June 2022

Fig. 3 and Fig. 4 present the comparison of reference Energy yield (Y_R) and system energy yield (Y_S) in May and June respectively and it is evident that in June, the system energy yield approaches the reference energy yield more than in May, implying that more power is extracted from the PV system.

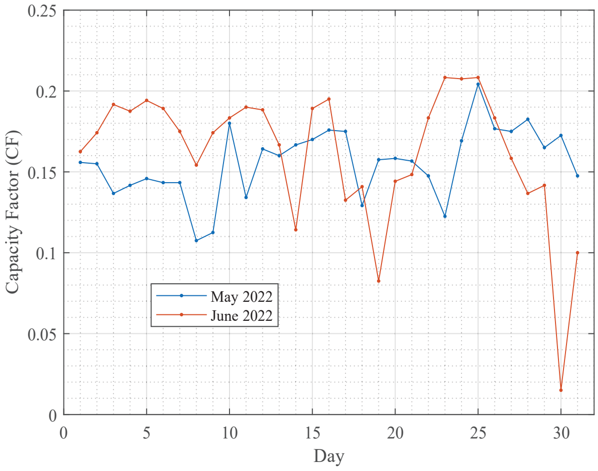


Fig. 5. Comparison of Capacity Factor per Day of both months

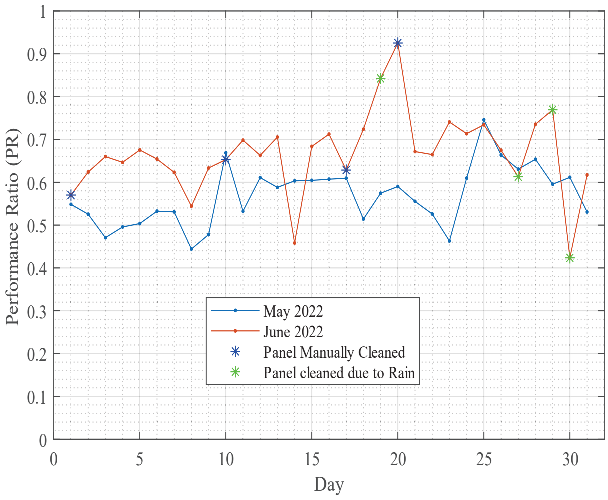


Fig. 6. Comparison of Performance Ratio per Day of both months

Fig. 5 compares the daily capacity factor of both months, but since the capacity factor does not account for environmental factors such as irradiance hence it fails to provide any conclusion whereas Fig. 6 depicts the performance ratio comparison of both months and it can be observed that except for 5 days, the PR of June is exceeding the PR of May. Due to regular cleaning of panels, maximum PR of 0.925 has been observed in June and due to frequent cloudy days, minimum PR of 0.42 also occurred in the same month. The marked points indicate the days on which the panels were either manually cleaned or were naturally washed due to rain. Fig. 6 also depicts that the 5 kW photovoltaic system performs better in June than in May.

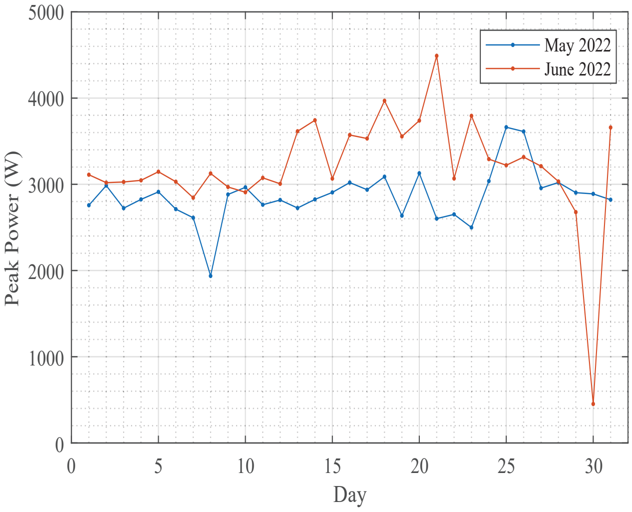


Fig. 7. Comparison of Peak Power (W) per day of both months

It can be perceived from Fig. 7 that on most days, the daily peak power of June is exceeding that of May as a consequence of regular cleaning of panels. In June, highest peak power of 4490 W was recorded and due to frequent cloudy days lowest peak power of 452 W was also observed.

TABLE IV. COMPARISON OF MONTHLY AVERAGE VALUES OF BOTH MONTHS

Month **Average**	**May 2022**	**June 2022**
Reference Energy Yield	6.57	5.839
System Energy Yield	3.741	3.886
Capacity Factor	0.156	0.162
Performance Ratio	0.568	0.667
Peak Power	2864.80 W	3203.484 W

Above Table IV shows the monthly average values of various quantities for both months and it can be observed that the monthly reference energy yield of June is lesser than that of May whereas system energy yield, capacity factor, performance ratio and peak power of June exceed the corresponding May values. May has a higher potential for solar energy but June has better performance results. In June, there has been a remarkable 10% improvement in the performance ratio as compared to May. In June, the peak power also significantly improved as can be seen in Fig. 7.

TABLE V. TOTAL SYSTEM ENERGY COMPARISON OF BOTH MONTHS

Month Energy	May 2022	June 2022
Total Practical System Energy (kWh)	579.8	602.4
PVsyst Total System Energy (kWh)	797.762	715.608

After simulating the 5 kW photovoltaic system in PVsyst, the daily system energy of the practical PV system and simulation is compared and plotted in Fig. 8 for 62 days. June's practical system energy was higher than May's, relative to PVsyst energy and on one significant day, the practical system energy exceeds the PVsyst energy. The above Table V shows that regular cleaning of the panel considerably improves the monthly total system energy. The higher practical system energy for June is primarily attributable to the regular cleaning of panels, despite the higher PVsyst daily system energy for May. In contrast to the overall PVsyst system energy of 1513.37 kWh, the total practical system energy for the 62 days is 1182.2 kWh. Total system losses for 62 days relative to PVsyst energy are 331.7 kWh, with 217.96 kWh occurring in May and 113.21 kWh occurring in June. Due to regular PV panel cleaning, energy losses were reduced by 104.75 kWh in June.

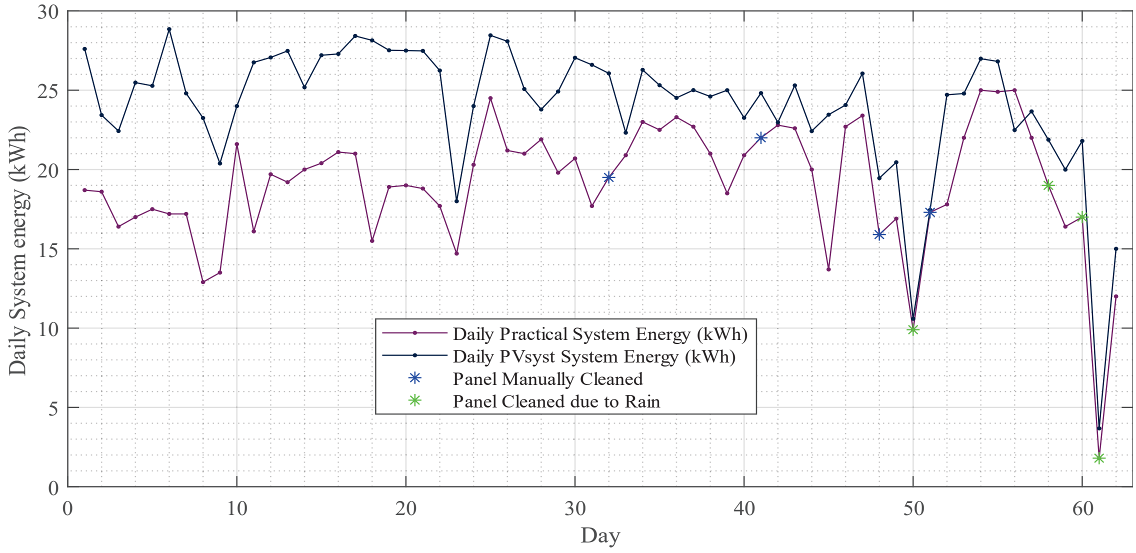


Fig. 8. Daily system energy comparison of Practical 5 kW PV system and PVsyst 5 kW PV system over 62 days

TABLE VI. COMPARISON OF PR FOR PRACTICAL SYSTEM AND PVSYST SYSTEM FOR BOTH MONTHS

Performance Ratio Month	Practical 5 kW PV System	PVsyst 5 kW PV System
May 2022	0.568	0.784
June 2022	0.667	0.778

Table VI shows the performance ratio comparison of simulation and practical system for both months. Performance ratio is a globally acceptable indicator for judging the efficiency of the PV system and it can be observed that there has been a significant 10% improvement in performance ratio in June.

V. CONCLUSION

A 62-day performance analysis of the 5 kW photovoltaic system is conducted. For the former 31 days, the panels were left naturally dirty whereas regularly cleaned for the remaining 31 days. May has a higher potential for solar energy but the performance ratio is high for June as a result of regular cleaning of panels. PV panels perceive rainfall as a significant factor while analysing performance; although cloudy days block the sun's insolation, rainfall cleans the panels and reduces their temperature, leading to better output on sunny days after rainfall. As a result of routinely cleaning the panels, there was a considerable 10% improvement in performance ratio in June. According to PVsyst analysis, June's practical system energy was higher than May's, relative to PVsyst energy and on one significant day, practical system energy outperforms the PVsyst output. PVsyst software has great potential in analysing PV systems and should be explored further. The frequency of the cleaning of PV panels is also another important aspect for performance improvement and thus can be included in future research.

REFERENCES

[1] BP, “Statistical Review of World Energy 2021,” London, UK, 2021.

[2] P. G. Kale, K. K. Singh and C. Seth, "Modeling Effect of Dust Particles on Performance Parameters of the Solar PV Module," 2019 Fifth International Conference on Electrical Energy Systems (ICEES), 2019, pp. 1-5.

[3] N.W. Alnaser, M.J. Al Othman, A.A. Dakhel, I. Batarseh, J.K. Lee, S. Najmaii, A. Alothman, H. Al Shawaikh, W.E. Alnaser, “Comparison between performance of man-made and naturally cleaned PV panels in a middle of a desert”, Renewable and Sustainable Energy Reviews,Volume 82, Part 1, 2018, pp. 1048-1055.

[4] Z. Abderrezzaq, M. Mohammed, N. Ammar, S. Nordine, D. Rachid and B. Ahmed, "Impact of dust accumulation on PV panel performance in the Saharan region," 2017 18th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), 2017, pp. 471-475.

[5] E E. M. Gnedi and I. M. Saleh, "Evaluating the Performance of Different PV Modules Technology Due to Dust Accumulation in Tripoli Region," 2021 IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA, 2021, pp. 559-564

[6] M. A. Elias, M. S. M. Resali, N. Muda, and R. B. Ramli, “Effects of coal and fly ash dust deposition of photovoltaic panel performance: A photovoltaic system at coal-fired power plant case study,” J. Phys. Conf. Ser., vol. 1358, no. 1, 2019, pp. 012038.

[7] N. Sakarapunthip, D. Chenvidhya, S. Chuangchote and T. Chenvidhya, "Dust Accumulation and Its Effect on PV Performance in Tropical Climate and Rice Farm Environment," 2021 IEEE 48th Photovoltaic Specialists Conference (PVSC), 2021, pp. 1848-1854.

[8] A. Hussain, A. Batra, and R. Pachauri, “An experimental study on effect of dust on power loss in solar photovoltaic module,” Renewables, vol. 4, no. 1, 2017, pp. 1–13.

[9] Y. Andrea, T. Pogrebnaya, and B. Kichonge, “Effect of industrial dust deposition on photovoltaic module performance: Experimental measurements in the tropical region,” Int. j. photoenergy, vol. 2019, 2019, pp. 1–10.

[10] IEC TS 61724-3, "Photovoltaic system performance – Part 3: Energy evaluation method," Edition 1.0, 2016-07

[11] IEC TS 61724-2, "Photovoltaic system performance – Part 2: Capacity evaluation method," Edition 1.0, 2016-07

[12] K. P. Satsangi, G. S. Sailesh Babu, D. B. Das and A. K. Saxena, "Performance Evaluation of Grid Interactive Photovoltaic System," 2018 International Conference on Computing, Power and Communication Technologies (GUCON), 2018, pp. 691-695.

[13] M. Adar, Z. Khaouch, M. Mabrouki, A. Benouna and A. Chebak, "Performance Analysis of PV Grid-Connected in Fours Special Months of the Year," 2017 International Renewable and Sustainable Energy Conference (IRSEC), 2017, pp. 1-5.

[14] A. M. Khalid, I. Mitra, W. Warmuth, ad V. Schacht, “Performance ratio – Crucial parameter for grid connected PV plants,” Renew. Sustain. Energy Rev., vol. 65, 2016, pp. 1139–1158.

[15] Performance Monitoring Guidelines for Photovoltaic Systems. PERFORMANCE project funded by the European Commission 6th Framework Programme, Contract no: SES-019718.

[16] Performance Ratio-Quality Factor for the PV plants-SMA, Web. 22 Jan 2015

[17] H. S. Huang, J. C. Jao, K. L. Yen and C. T. Tsai, "Performance and Availability Analyses of PV Generation Systems in Taiwan," Interantional Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering, vol. 5, no. 6, 2011, pp. 731-735.

[18] J J. Leloux, L. Narvarte, and D. Trebosc, “Review of the performance of residential PV systems in Belgium,” Renew. Sustain. Energy Rev., vol. 16, no. 1, 2012, pp. 178–184.

[19] D. D. Milosavljević, T. M. Pavlović, and D. S. Piršl, “Performance analysis of A grid-connected solar PV plant in Niš, republic of Serbia,” Renew. Sustain. Energy Rev., vol. 44, 2015, pp. 423–435.

[20] R. Shenouda, M. S. Abd-Elhady, and H. A. Kandil, “A review of dust accumulation on PV panels in the MENA and the Far East regions,” J. Eng. Appl. Sci., vol. 69, no. 1, 2022, pp. 1–29.

[21] R. Shenouda, M. S. Abd-Elhady, and H. A. Kandil, “Influence of seasonal effect on dust accumulation on Photovoltaic panels that operate light posts,” Energy rep., vol. 8, 2022, pp. 1275–1284.

[22] A. Soualmia and R. Chenni, "Modeling and simulation of 15MW grid-connected photovoltaic system using PVsyst software," 2016 International Renewable and Sustainable Energy Conference (IRSEC), 2016, pp. 702-705.

[23] M. Satish, S. Santhosh and A. Yadav, "Simulation of a Dubai based 200 KW power plant using PVsyst Software," 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN), 2020, pp. 824-827.

[24] P. Yadav, N. Kumar and S. S. Chandel, "Simulation and performance analysis of a 1kWp photovoltaic system using PVsyst," 2015 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC), 2015, pp. 0358-0363.

Taylor & Francis
Taylor & Francis Group

International Journal of Parallel, Emergent and Distributed Systems

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gpaa20

Position control of a ball balancer system using Particle Swarm Optimization, BAT and Flower Pollination Algorithm

Ajit Kumar Sharma & Bharat Bhushan

INTERNATIONAL JOURNAL OF PARALLEL, EMERGENT AND DISTRIBUTED SYSTEMS
https://doi.org/10.1080/17445760.2023.2190972

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Position control of a ball balancer system using Particle Swarm Optimization, BAT and Flower Pollination Algorithm

Ajit Kumar Sharma and Bharat Bhushan

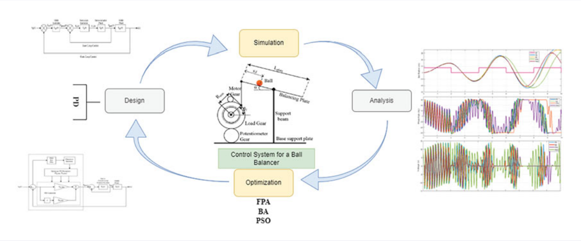
Department of Electrical Engineering, Delhi Technological University, Delhi, India

ABSTRACT

The design and control of the 2DoF Ball Balancer system is presented in this work. The ball balancer is a feedback-based underactuated system that is nonlinear, multivariate, and electromechanical. The proportional derivative (PD) controller is optimized by using Bat Algorithm, Particle Swarm Optimization, and Flower Pollination Algorithm in this research. By regulating the plate inclination angle, the suggested controller accomplishes self-balancing control for a ball on the plate. The modelling of the ball balancer system is accomplished using a 2DoF ball balancer system. In addition, Bat Algorithms, Particle Swarm Optimization, and the Flower Pollination Algorithm are used to analyze the state of a process autonomously. The system's model is created using MATLAB/Simulink approaches, and the results present the system with a steady and controllable output for ball balancing and plate angle control.

Graphical abstract

The author control the position of the ball balancer by using the PD controller and optimized the parameter of the controller through FPA, BA, and PSO.



ARTICLE HISTORY
Received 7 April 2022
Accepted 10 March 2023

KEYWORDS
Ball balancer system; Bat Algorithms (BA); Particle Swarm Optimization (PSO); Flower Pollination Algorithm (FPA); proportional derivative controller

1. Introduction

Nonlinear systems with underactuated actuators are approximated with intelligent-control and autonomous decision development methods [1]. They arose in a variety of contexts [2] and were attempted in a variety of ways. Researchers have investigated the behaviour of numerous controllers aiming at achieving self-balancing control and steady-state operation because of the structural complexity of these systems. The inverted pendulum [3], the twin-rotor-multi-input–multi-output system (TRMS) [4], the ball & beam system [5], the hovercraft [6], the furuta pendulum [7], and the ball &

CONTACT Ajit Kumar Sharma sharmaajit01@gmail.com Department of Electrical Engineering, Delhi Technological University, Delhi, India

plate system [8] are all used as benchmark examples in the majority of the studies. In general, linear controllers make closed-loop control for such systems simple to implement [9], but their complex nonlinear dynamics limit the control rules for all generalized applications. This gained the attention of a number of nonlinear control techniques [10], but these controllers have difficulties when it comes to dealing with external load and lagging caused by additional feedback. The literature has developed feedback-linearization [11] and partial feedback-linearization [12] for mechanically underactuated systems in order to satisfy these needs. However, challenges arising from a lack of resilience have limited their use in a variety of disciplines. In addition, [13] proposes a passivity-based control at the selected equilibrium point, proposed a strategy for passivating the system with a storage function. With differential feedback, this has the drawback of being unable to magnify measurement noises. These drawbacks also prevented the successful development of a control mechanism capable of achieving steady-state operation. These objectives are met by exemplifying the position control and path tracking for an underactuated ball and plate benchmark problem in this study. Different proportional–integral–derivative (PID) controller-based approaches for system control on a point-to-point basis have previously been investigated [1415]. Disturbance rejection controllers [16] and different optimization algorithms [17] are also used to demonstrate the desired tracking performance for the ball and plate systems. Sliding mode control has been extensively explored for achieving self-balancing control [18], as well as the development of fractional-order sliding mode control [19] to eliminate the chattering phenomenon of classic SMC with greater efficiency [20]. Because of their advantages with time-varying reference systems, model predictive controllers were often utilised with ball balancer systems [21]. The fundamental problem of these classic approaches, however is that they produce a long settling time and peak overshoot. In addition to self-balancing control, hybrid and intelligent controllers such as fuzzy [22], fuzzy cerebellar model articulation controller [23], and particle swarm optimization-based fuzzy-neural controller [24] are utilized to control the position and trajectory of the ball and plate system. The PID controller is widely utilised in practical engineering applications, despite the fact that there are various control algorithms for establishing self-balancing control with balancer systems in the literature. The PID controller provides a number of benefits, including a simple design, high dependability, and exceptional stability. On the other hand, traditional PID controllers have a severe problem with parameter tuning.

In many engineering and industrial design applications, we must try to discover the best solution to a problem while dealing with exceedingly complex limitations. Such limited optimization problems are frequently highly nonlinear, and finding the best solution can be time-consuming. For issues involving nonlinearity and multimodality, traditional optimization does not produce good solutions. To solve such tough problems, the current tendency is to use nature-inspired metaheuristic algorithms, which have been demonstrated to be unexpectedly efficient. As a result, the metaheuristics literature has grown dramatically in the recent two decades [25–28]. Researchers have only used a few natural properties so far, and there is still room for more algorithm improvement. There are a variety of strategies for tweaking PID parameters that may be found in the literature. Various intelligent approaches, such as fuzzy [29] neural network [30] self-tuning algorithms [31] genetic [32] and evolutionary algorithms [33] are used in these techniques. On the other hand, these strategies strive to retain superior generations, resulting in a local rather than global optimum. Furthermore, issues with intelligent controller weight adjustment, low memory, premature convergence, weak local search, and high computational effort for genetic and other evolutionary algorithms have led to the development of an optimal multi-objective approach for solving combinatorial optimization problems [34,35], such as Particle Swarm Optimization (PSO), BAT Algorithm (BA), and Flower Pollination Algorithm (FPA). To construct a nonlinear algorithm that can try to address multimodal optimization difficulties, the attraction mechanism was integrated with light intensity fluctuations. Swarm Intelligence (SI) refers to a type of interpretive capability that appears in processing unit communication [36]. Where the theory of intelligence indicates that the analytical ability is successful in some ways [37,38], the theory of swarm describes stochastic manner, multiplicity, messiness, and unpredictability. SI is inspired by human behaviour as well as insects such as termites, ants, wasps, and bees, as well as other social animal groupings such as flocks of birds and schools of fish [39,40]. Individuals in the swarm can be described as simple solutions, yet they have a strong ability to work together to solve complex non-linear problems [41].

PSO was established in 1995 by Kenndy and Eberhart for the purpose of training neural networks and solving non-linear optimization issues. Human cognition of natural behaviour, such as how human learning is influenced by their surroundings, how they interact with others, and how they encode their patterns into their learning methods, are simple findings in PSO. PSO uses this learning phenomenon to find an optimal solution. PSO has become increasingly natural for dealing with non-linear complex optimization problems, especially in a wide range of fields. A swarm in PSO is a population of vector solutions that is probing new search areas while hunting for food, resembling the evolution of a school of fish. To find the global optimum, all particles in the swarm translate information and follow each other's best experiences as well as their own past best experiences [42]. Each particle must adhere to the basic rule of determining the location of its prior best or neighbour.

Researchers have developed a novel data clustering technique (named FPAB) that simulates flower pollination by bees [43]. Following that, a flower pollination algorithm (FPA) was devised, which resembles a broader notion of the flower reproduction process [44]. Due to its effective application to real-world problems, FPA has recently gained a lot of attention. The FPA has been utilised to manage a range of optimization problems in a variety of real-world scenarios due to its efficiency and versatility.

Xin-She Yang created the bat algorithm (BA) in 2010, based on the echolocation features of microbats [25]. During foraging, BA employs frequency tuning in conjunction with changes in loudness and pulse emission rates. All of these algorithms can be categorized as swarm intelligence heuristic algorithms since they optimise social interactions and biologically inspired rules [26,27].

In this paper, the authors present simulation findings of controllers on the control of a ball balancer. Due to its inherent complexity, the ball balancer system faces issues such as (1) balancing the ball on a plate and (2) point stabilisation control, which allows the ball to be moved to a precise location and held there while minimising tracking error and time. This study contributes to mathematical modelling, optimal parameter selection, and exemplary controller design for the ball balancer system to solve the existing issues. The concept of using metaheuristic algorithms to control a ball balancer system is a new one. The goal of this study is to compare several metaheuristics control strategies used on a ball balancer. The focus of the article is on utilising Simulink to develop and implement controllers for ball balancer setup. Adapting a metaheuristics algorithm for optimising controllers proved to be an innovative adaptation, as evidenced by the Ball Balancer findings. Finally, the comparison is performed using various control algorithms. The remainder of the paper is divided into the following sections: The ball balancer setup modelling and its operating phenomenon are described in Section II. The third section deals with the algorithms used for observing the ball's position on the plate. The results of comparative simulations are presented in Section 4 to verify the proposed methodology. Finally, in Section 5, the conclusion is presented, followed by references.

2. 2 DOF ball balancer model

A ball balancer is a typical benchmark problem for achieving, balancing control, ball position tracking, and visual servo control. The goal of creating a controller for a two-degree-of-freedom ball balancer is to keep the ball in place on the balance plate. This requires controlling the position of the rotating servos linked to the plate's bottom using the X-Y position of the ball as measured by the overhead camera. A ball and plate mechanism is depicted schematically in Figure 1.

The 2 DOF Ball Balancer module comprises a free-moving plate on which a ball can be placed. Two DOF gimbals connect two actuation units to the sides of the plate. The plate can be rotated in both the X- and Y-axis directions. The servo motors are wired together in actuating units, which are controlled by a potentiometer. To balance the ball at a specific planar position, the tilt angle of the plate can be adjusted by adjusting the position of the servo load gears. The Faulhaber series DC micromotor [45] is utilised to balance the system in both directions using the rotational motion of the plate.

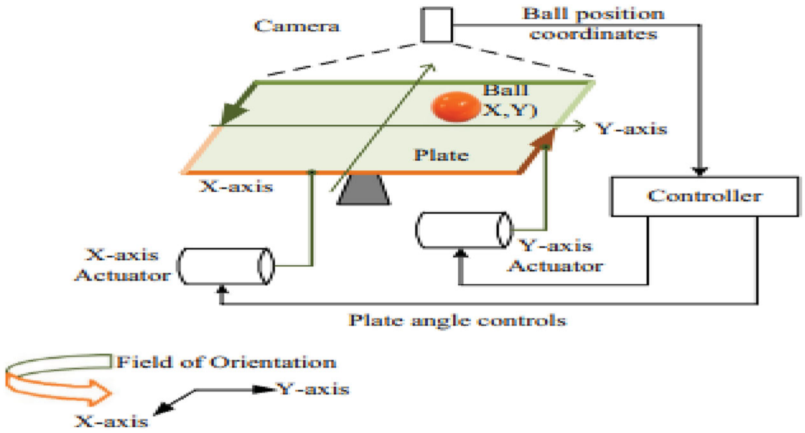


Figure 1. Schematic representation of ball and plate system.

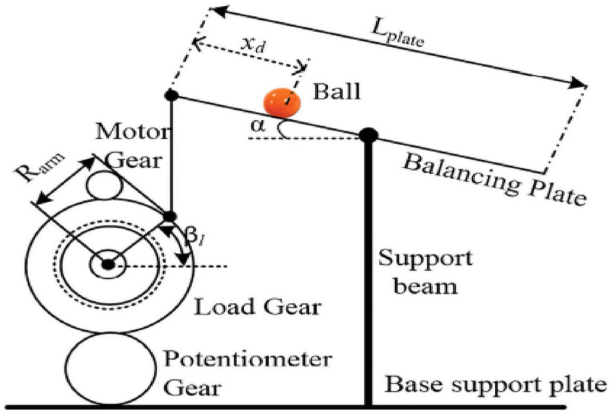


Figure 2. 2 DOF ball balancer free body diagram.

2.1. Mathematical modeling of ball balancer

The dynamics of each axis are considered to be the same in the 2 DOF Ball Balancer, which uses two symmetrical servo motor units. With the premise that the angle of the x-axis servo solely effects ball motion in the x direction, the 2 DOF Ball Balancer is represented as two de-coupled 'ball and beam' systems. The equation of motion represents the ball's motion along the x-axis in relation to the plate angle. The free body diagram of the Ball and Beam system is shown in Figure 2.

A positive voltage servo motor rotates the gear in the positive direction, causing the beam to rise and the ball to roll in the positive direction. The forces exerted on the ball as it goes down the beam will be, according to Newton's first law of motion:

$$m_{ball}\ddot{x}(t) = F_{x,t} - F_{x,r} \tag{1}$$

where m_{ball} is the ball's mass, $x(t)$ is its displacement, $F_{x,r}$ is its inertia force, and $F_{x,t}$ is its gravitational translational force. When the momentum force and gravitational force are both equal, the ball is considered to be in equilibrium.

The ball inertia force $F_{x,r}$ is given as:

$$F_{x,r} = m_{ball}g\sin\alpha_{beam} \tag{2}$$

And gravitational translational force $\boldsymbol{F_{x,t}}$ is equal to,

$$\boldsymbol{F_{x,t}} = \frac{\boldsymbol{J_{ball}\ddot{x}(t)}}{\boldsymbol{r^2{}_{ball}}} \tag{3}$$

The non linear equation of motion of ball beam is given as.

$$\boldsymbol{m_{ball}\ddot{x}(t)} = \boldsymbol{m_{ball}g}\sin\boldsymbol{\alpha_{beam}} - \frac{\boldsymbol{J_{ball}\ddot{x}(t)}}{\boldsymbol{r^2{}_{ball}}} \tag{4}$$

The acceleration is given as

$$\boldsymbol{\ddot{x}(t)} = \frac{\boldsymbol{m_{ball}g}\sin\boldsymbol{\alpha_{beam}r^2{}_{ball}}}{\boldsymbol{m_{ball}r^2{}_{ball}} + \boldsymbol{J_{ball}}} \tag{5}$$

The beam angle $\boldsymbol{\alpha}$, is affected by the position of the ball on the plate, which is further influenced by the servo gear angle. The following is the relationship between gear angle and beam angle:

$$\sin\boldsymbol{\theta_{gear}} = \frac{\sin\boldsymbol{\alpha_{beam}l_{plate}}}{2\boldsymbol{r_{arm}}}$$

The nonlinear equation for ball motion in terms of gear angle is:

$$\boldsymbol{\ddot{x}(t)} = \frac{2\boldsymbol{m_{ball}gr_{armr^2{}_{ball}}}}{\boldsymbol{l_{plate}}(\boldsymbol{m_{ball}r^2{}_{ball}} + \boldsymbol{J_{ball}})}\sin\boldsymbol{\theta_{gear}} \tag{6}$$

At zero angle the linearized equation of motion of the ball is given as –

$$\boldsymbol{\ddot{x}(t)} = \frac{2\boldsymbol{m_{ball}gr_{armr^2{}_{ball}}}}{\boldsymbol{l_{plate}}(\boldsymbol{m_{ball}r^2{}_{ball}} + \boldsymbol{J_{ball}})}\boldsymbol{\theta_{gear}} \tag{7}$$

In addition, the transfer function for controlling ball position for input $\boldsymbol{\theta_{gear}}$ and output $\boldsymbol{x}$ is as follows:

$$\boldsymbol{S_b(s)} = \frac{\boldsymbol{x(s)}}{\boldsymbol{\theta_{gear}(s)}} = \frac{\boldsymbol{K_b}}{\boldsymbol{s^2}} \tag{8}$$

where, $\boldsymbol{K_b} = \frac{2\boldsymbol{m_{ball}gr_{armr^2{}_{ball}}}}{\boldsymbol{l_{plate}}(\boldsymbol{m_{ball}r^2{}_{ball}}+\boldsymbol{J_{ball}})}$

Similarly, the servo motor’s control of plate angle is expressed as a transfer function.

$$\boldsymbol{S_s(s)} = \frac{\boldsymbol{\theta_{gear}}}{\boldsymbol{V_m(s)}} = \frac{\boldsymbol{K_g}}{\boldsymbol{s}(1+\boldsymbol{s\tau})} \tag{9}$$

The overall transfer function of the servo motor and ball balancer module cascaded connection is as follows:

$$\boldsymbol{S(s)} = \boldsymbol{S_s(s)S_b(s)} = \frac{\boldsymbol{x(s)}}{\boldsymbol{V_m(s)}} = \frac{\boldsymbol{K_bK_g}}{\boldsymbol{s^3}(1+\boldsymbol{s\tau})} \tag{10}$$

The state-space representation of the above equation is given as,

$$\begin{bmatrix} \boldsymbol{\dot{x}}(t) \\ \boldsymbol{\ddot{x}}(t) \\ \boldsymbol{\dot{\theta}_{gear}}(t) \\ \boldsymbol{\ddot{\theta}_{gear}}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \boldsymbol{K_b} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1/\tau \end{bmatrix} \begin{bmatrix} x(t) \\ \boldsymbol{\dot{x}}(t) \\ \boldsymbol{\theta_{gear}}(t) \\ \boldsymbol{\dot{\theta}_{gear}}(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \dfrac{\boldsymbol{K_bK_g}}{\tau} \end{bmatrix} \boldsymbol{u(t)}$$

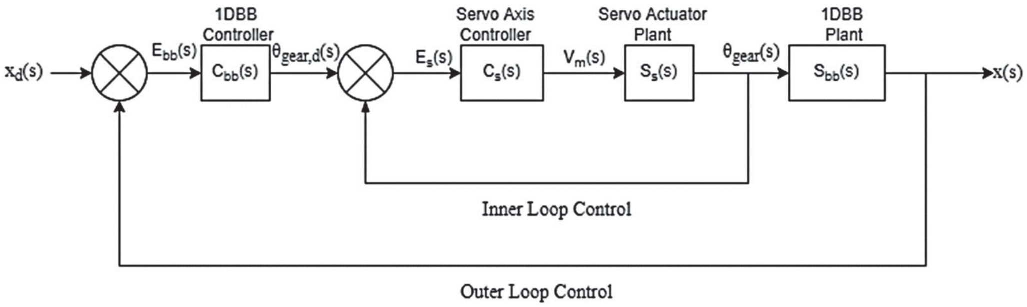


Figure 3. Block diagram of closed loop ball balancer system.

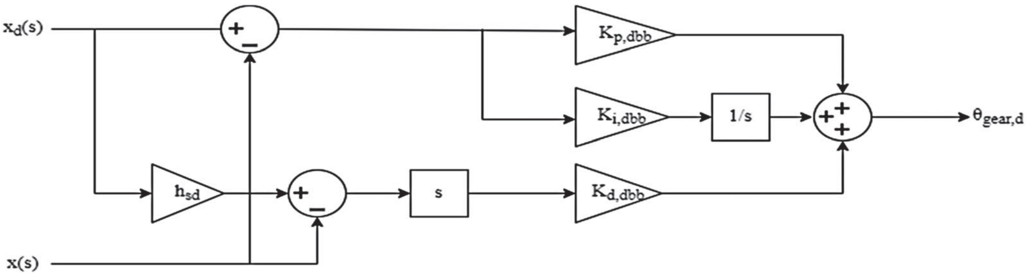


Figure 4. PID compensator with derivative set point weight.

3. Designing structure of ball balancing controller

Figure 3 shows the SRV02's x-axis control model, which is combined with the ball balancer mechanism. The ball balancer block diagram depicts two loops of control. The SRV02 motor model is the first loop, and the 1D ball balancer is the second loop. The inner loop's goal is to control the servo motor's position and estimate the voltage in order to calculate the load's desired angle.

The following optimization strategies were developed and tested for the PD controller to balance and regulate the ball balancer system:

- BAT Algorithm (BA)
- Particle Swarm Optimization (PSO) Algorithm
- Flower Pollination Algorithm (FPO)

In order to calculate the initial operating gains, the 1DBB controller must be represented as a PID controller in the time domain.

$$\theta_{gear,d}(t) = K_{p,dbb}(x_d(t) - x(t)) + K_{d,dbb}\left(h_{sd}\left(\frac{d}{dt}x_d(t)\right) - \frac{d}{dt}x(t)\right) + K_{i,dbb}\int(x_d(t) - x(t))d(t) \tag{11}$$

where, $K_{p,dbb}$, $K_{d,dbb}$ and $K_{i,dbb}$ is proportional gain, derivative gain and velocity gain respectively. h_{sd} is a velocity weight parameter that is included by a controller to compensate for the derivative error as shown in Figure 4.

The closed-loop equation of outer loop system when servo dynamics are neglected of a ball balancer.

$$\theta_{gear}(s) = \left(K_{p,dbb} + \frac{K_{i,dbb}}{s}\right)(x_d(s) - x(s)) + K_{d,dbb}s(h_{sd}x_d(s) - x(s)) \tag{12}$$

when the ball rotates along x-axis of the plate, the estimated and needed gear load are equal ($\theta_{gear,d} = \theta_{gear}$), substituting the outer loop controller with the one-dimensional ball balancer system yields the closed-loop equation as follows:

$$\frac{x(s)}{x_d(s)} = \frac{K_{ball}(K_{p,dbb}s + K_{i,dbb} + K_{d,dbb}s^2h_{sd})}{s^3 + K_{bb}K_{p,dbb}s + K_{bb}K_{i,dbb} + K_{bb}K_{d,dbb}s^2} \tag{13}$$

where, K_{ball} is a ball balancer constant.

To calculate the PID constant, the third order prototype equation is given as:

$$(s^2 + 2\zeta\omega_n s + {\omega_n}^2)(s + p_0) \tag{14}$$

where ω_n is the system's natural frequency, ζ is the damping ratio, and p_0 is the pole location.

The above third-order characteristic equation becomes

$$s^3 + (2\zeta\omega_n s + p_0)s^2 + ({\omega_n}^2 + 2\zeta\omega_n p_0)s + {\omega_n}^2 p_0 \tag{15}$$

The closed-loop equation's third-order characteristic equation is:

$$s^3 + K_{ball}K_{p,dbb}s + K_{ball}K_{i,dbb} + K_{ball}K_{d,dbb}s^2 \tag{16}$$

Equating equation no. (15) and (16), the following observations are made:

$$K_{ball}K_{p,dbb} = 2\zeta\omega_n + p_0$$

$$K_{ball}K_{i,dbb} = {\omega_n}^2 p_0$$

$$K_{ball}K_{d,dbb} = {\omega_n}^2 + 2\zeta\omega_n p_0$$

Further, the PID control gains can be calculated as follows:

$$K_{p,dbb} = \frac{2\zeta\omega_n s + p_0}{K_{ball}}$$

$$K_{i,dbb} = \frac{{\omega_n}^2 p_0}{K_{ball}}$$

$$K_{d,dbb} = \frac{{\omega_n}^2 + 2\zeta\omega_n p_0}{K_{ball}}$$

To meet the specifications of proportional derivative gain, the pole location is adjusted at origin, i.e, $p_0 = 0$.

Hence the control gains of PD controller is given as:

$$K_{p,dbb} = \frac{2\zeta\omega_n}{K_{ball}}$$

$$K_{d,dbb} = \frac{{\omega_n}^2}{K_{ball}}$$

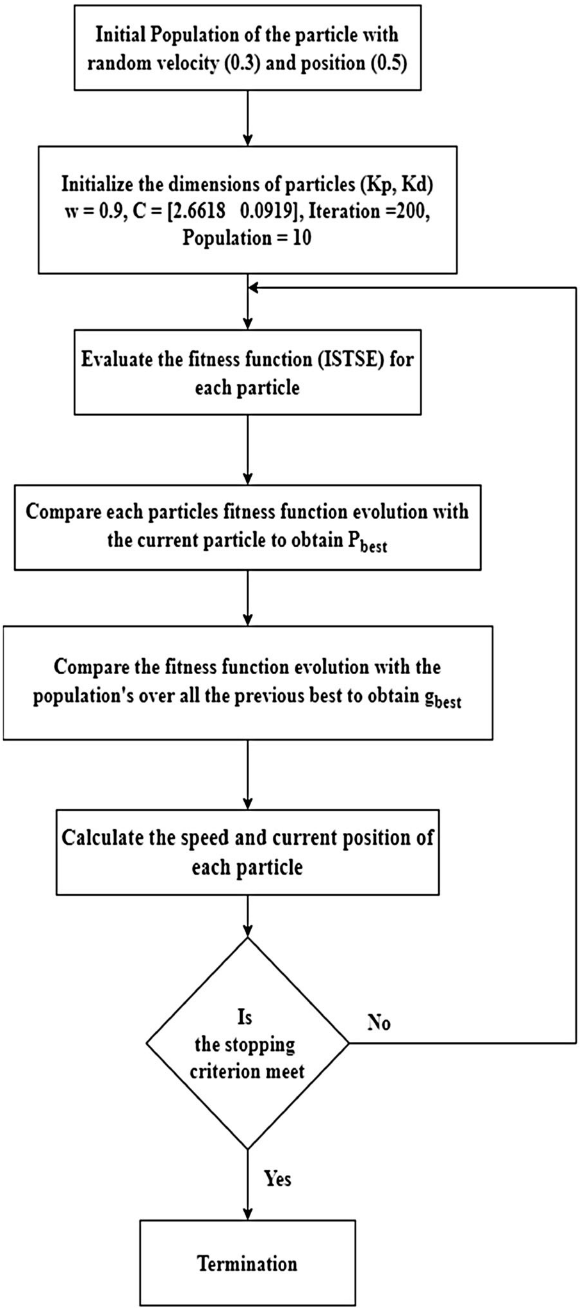

Initial Population of the particle with random velocity (0.3) and position (0.5)
Initialize the dimensions of particles (Kp, Kd) w = 0.9, C = [2.6618 0.0919], Iteration =200, Population = 10
Evaluate the fitness function (ISTSE) for each particle
Compare each particles fitness function evolution with the current particle to obtain P_{best}
Compare the fitness function evolution with the population's over all the previous best to obtain g_{best}
Calculate the speed and current position of each particle
Is the stopping criterion meet
No
Yes
Termination

Figure 5. Flow chart of PSO algorithm with initial parameters.

3.1. Particle Swarm Optimization

Particle Swarm Optimization (PSO), often known as swarm intelligence, is a type of intelligence inspired by flocks of birds [36]. Birds foraging for food and interacting with one another as they fly are both optimization strategies.

The PSO is made up of a population of particles, each of which represents a possible solution to the problem $\boldsymbol{K_p}$ and $\boldsymbol{K_d}$ in our situation. Each particle can be represented by an object having a position vector and a vector velocity, with the location relative to the search space and the velocity guiding the particle position during the process execution (Figure 5).

The basic PSO algorithm consists of the equation of velocity and position, respectively:

$$\boldsymbol{v_i}(\boldsymbol{k}+1) = \boldsymbol{w}.\boldsymbol{v_i}(\boldsymbol{k}) + \boldsymbol{c}_1\boldsymbol{r}_1(\boldsymbol{pbest_i} - \boldsymbol{x_i}(\boldsymbol{k})) + \boldsymbol{c}_2\boldsymbol{r}_2(\boldsymbol{gbest} - \boldsymbol{x_i}(\boldsymbol{k}) \tag{17}$$

$$\boldsymbol{x_i}(\boldsymbol{k}+1) = \boldsymbol{x_i}(\boldsymbol{k}) + \boldsymbol{\Delta k}.\boldsymbol{v_i}(\boldsymbol{k}+1) \tag{18}$$

The population size is given by i $=$ 1 … .n. ***pbest*** (personal best) and ***gbest*** (global best) are the best positions achieved by a particle in a given position and the entire population in a given neighbourhood, respectively; w is the inertia constant; $\boldsymbol{c}_1$ is a social factor; $\boldsymbol{c}_2$ is the factor cognitive; $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are random numbers generated using a uniform distribution in the interval [0,1]; and t $=$ 1. A social factor of 1.2 and a cognitive factor of 0.12 were employed in the simulation findings. The inertia constant, w, is set to 0.9. Figure 5 shows the flowchart of PSO with its initial parameters.

3.2. BAT algorithm

The Bat Algorithm is a metaheuristic based on some bat species' night-flight echolocation technique. A set of bats stored in the form of a vector, each representing a candidate solution, is generated in this computational model. The goal is to go to the prey, which is the best approach for minimising the cost function.

Initially, all n bats $\boldsymbol{x_i}(\boldsymbol{i} = 1, 2, \ldots .\boldsymbol{n})$ are initialized with the following parameters: pulse rate $\boldsymbol{r_i}$ velocity $\overrightarrow{\boldsymbol{v_i}} = 0$, amplitude $\boldsymbol{A_i}$, frequency $\boldsymbol{f_i}$ and position $\vec{\boldsymbol{x}}_{\boldsymbol{i}}$. For each instant the velocity and position are updated, respectively. The steps of bat algorithim is discussed in flow chart in Figure 6.

$$\boldsymbol{v_i^j}(\boldsymbol{t}) = \boldsymbol{v_i^j}(\boldsymbol{t}-1) + [\boldsymbol{x^i_{cgbest}} - \boldsymbol{x_i^j}(\boldsymbol{t}-1)]\boldsymbol{f_i} \tag{19}$$

$$\boldsymbol{x_i^j}(\boldsymbol{t}) = \boldsymbol{x_i^j}(\boldsymbol{t}-1) + \boldsymbol{v_i^j}(\boldsymbol{t}) \tag{20}$$

The variable $\boldsymbol{\beta} \in [0, 1]$ is a random number generated from a uniform distribution and is used to update and weight $\boldsymbol{f_i} \in [\boldsymbol{f_{min}}, \boldsymbol{f_{max}}]$. The variable $\boldsymbol{x^i_{cgbest}}$ denotes the current global best solution for a decision variable d, which is determined by comparing all of the solutions offered by n bats. In order to explore the domain of candidate solutions to the problem, the algorithm executes a local search in the form of a random walk: $\boldsymbol{x_i^{new}} = \boldsymbol{x_i^{old}} + \boldsymbol{\varepsilon m}$, where $\boldsymbol{m}$ is the mean of the amplitude of all bats at time t, and $\boldsymbol{\varepsilon}$ is a random value derived from a uniform distribution. The algorithm comes to stop when $\boldsymbol{r_i}$ hits a predetermined minimum value or when the maximum number of iterations is reached, which are known as stopping conditions. An amplitude of 0.5 and an initial pulse rate of 0.5 were used in this job. At maximum and minimum frequencies, 2 and 0 are formed, respectively. The algorithm's population is similar to the PSO in that each bat represents a $\boldsymbol{K_p}$ and $\boldsymbol{K_d}$.

3.3. Flower Pollination Algorithm

The Flower Pollination Algorithm (FPA) is a method for pollinating flowers that was proposed by [17]. Because pollinators may fly great distances, they are classified as global pollinators, and the Lévy probability distribution can be used to describe their behaviors. Two key rules govern the implementation of the method utilizing the Lévy distribution: 1 – The direction of travel must be random. A uniform distribution can be used to generate a direction; however, the creation of steps must follow the Lévy distribution.

The following rules [17] were devised by Yang: 1 – Biotic pollination and crossover are regarded as a global pollination process, with pollen transporters flying from Levy; 2 – Abiotic pollination and self-pollination are called local pollination; 3 – Loyalty to a flower can be thought of as having a probability of reproduction proportionate to the similarity of the two plants involved; 4 – Local and global pollination are governed by a probability $\boldsymbol{p} \in [0, 1]$.Local pollination can play a substantial role in overall pollination activities due to physical proximity and other factors such as wind.

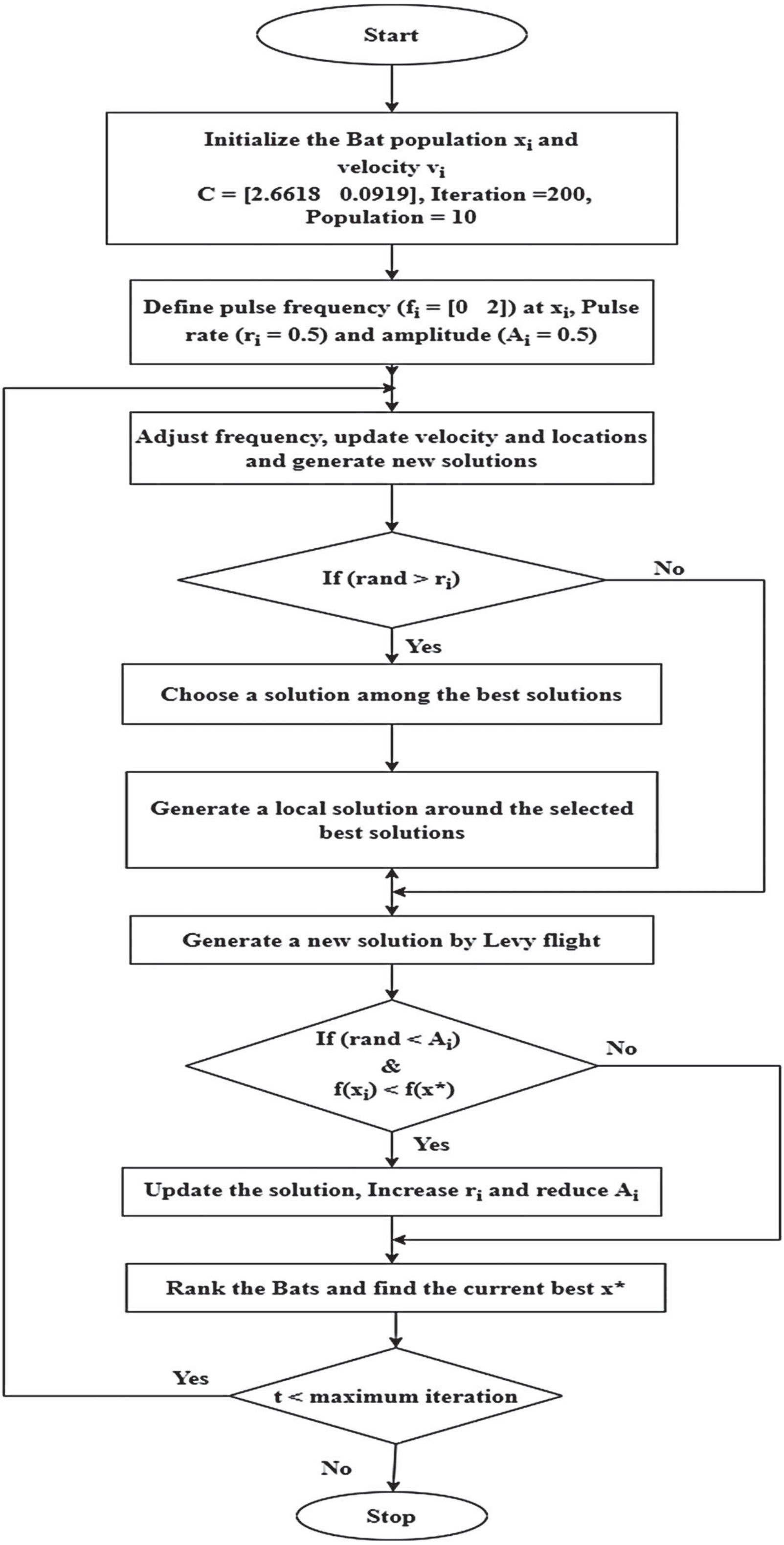

Start
Initialize the Bat population x_i and velocity v_i
C = [2.6618 0.0919], Iteration =200, Population = 10
Define pulse frequency (f_i = [0 2]) at x_i, Pulse rate (r_i = 0.5) and amplitude (A_i = 0.5)
Adjust frequency, update velocity and locations and generate new solutions
If (rand > r_i)
No
Yes
Choose a solution among the best solutions
Generate a local solution around the selected best solutions
Generate a new solution by Levy flight
If (rand < A_i) & $f(x_i) < f(x^*)$
No
Yes
Update the solution, Increase r_i and reduce A_i
Rank the Bats and find the current best x*
Yes
t < maximum iteration
No
Stop

Figure 6. Flow chart of Bat Algorithm with initial parameters.

The best individual represented by $\boldsymbol{g_*}$. The first rule, along with a flower's loyalty, may be expressed mathematically as $\boldsymbol{x_i^{t+1}} = \boldsymbol{x_i^t} + \boldsymbol{L}(\boldsymbol{x_i^t} - \boldsymbol{g_*})$ where $\boldsymbol{x_i^t}$ is the pollen $\boldsymbol{i}$ in the vector of solutions $\boldsymbol{x_i}$ at iteration t, and L is the pollination strength, whose value is determined by the Lévy distribution. The insects can move a long distance with just a few steps away, and this can mimic with a Lévy flight. That is $\boldsymbol{L} > 0$ from a Lévy distribution.

$$\boldsymbol{L} \sim \frac{\lambda \boldsymbol{r}(\lambda)\boldsymbol{sen}(\pi\lambda/2)}{\pi}\frac{1}{\boldsymbol{s}+\lambda}(\boldsymbol{s} \gg \boldsymbol{s}_0 > 0) \tag{21}$$

where $\coprod(\lambda)$ is the Gamma function and $\boldsymbol{s}_0$ a minimum step.

Local pollination (rule 2) and flower loyalty are represented by the equation $\boldsymbol{x_i^{t+1}} = \boldsymbol{x_i^t} + \epsilon(\boldsymbol{x_j^t} - \boldsymbol{x_k^t})$ where $\boldsymbol{x_j^t}, \boldsymbol{x_k^t}$ are pollens from separate plants of the same species in the same iteration. This is similar to a flower's allegiance in a small neighbourhood. If $\boldsymbol{x_j^t}, \boldsymbol{x_k^t}$ were of the same species or from the same population, a tour local random walk (local random walk) would be created by selecting $\coprod$ from a uniform distribution. To switch between undertaking global pollination and enhancing local pollination, an exchange probability, or probability of closeness p, is employed according to rule 4. The reason for this characteristic is that the majority of pollination actions are carried out by bees. It might happen on a local or global level. In terms of practicality, flowers that are close by or not too far away from the neighbourhood are more vulnerable to pollination than those that are farther away. $p = 8$ in this article, and each flower represents a $\boldsymbol{K_p}$ and $\boldsymbol{K_d}$. Figure 7 shows the algotithim of flower pollination.

4. Simulation analysis

The numerical simulation of the 2DoF ball balancer model described in Section 2 was created using MATLAB/ Simulink software. The action of one servo unit's controller has an impact on the action of the second servo unit's controller because the plate on the two servo units is symmetrical. Regardless of the fact that both controllers are developed in a decoupled context, they operates them in a connected environment. The technology is set up to manage the ball's square trajectory on a plate. As the reference trajectory, PD is used to control the ball balancer by providing it a square input signal with a frequency of 0.08 Hz and an amplitude of 5 volt. The PD controller's values are originally determined using the method explained in Sect. 3.1. In addition, the parameters of the PD controller are optimised using three different optimization methodologies (PSO, BA, and FPA), and the difference between the desired and measured ball position is measured as shown in figure 8. The results of PSO, BA, and FPA on the PD controller are compared to the action of the typical PD controller on the same simulation running for the same trajectory to assess their performance.

It is identified that by measuring the error between the reference trajectory and captured ball position coordinates, the plate angle can be controlled. Hence the choice of the objective function is to optimize the operation of a controller which should base on the error between measured and desired trajectories of the ball position. Generally, an error can be termed as an objective function by formulating it as an integral of the square, time, and absolute. All these functions express error as an objective function by evaluating its integral over a fixed interval of time. The optimal solution is connected with an objective function $\boldsymbol{'J'}$ during the optimization process. Conventionally, the integral square error, integral absolute error, and integral time absolute error were used for formulating the objective function. But due to its slow response and large oscillation time here we use time-weighted indices, integral of squared time-multiplied square of the error (ISTSE) to improve error performance in optimization of controller (Figure 8). The objective function $\boldsymbol{'J'}$ is given as:

$$\boldsymbol{J} = \int_{\boldsymbol{o}}^{\boldsymbol{T}} \boldsymbol{t}^2\boldsymbol{e}^2(\boldsymbol{t})\boldsymbol{dt} \tag{22}$$

Figure 9 contains a detailed description of the ball balancer system's ball position, servo angle, and voltage optimizations for PSO, BA, and FPA. Figure 9a presents the comparison of the PSO, BA, and FPA algorithms for the ball's position on the x-axis. The results demonstrate that the x-axis position of

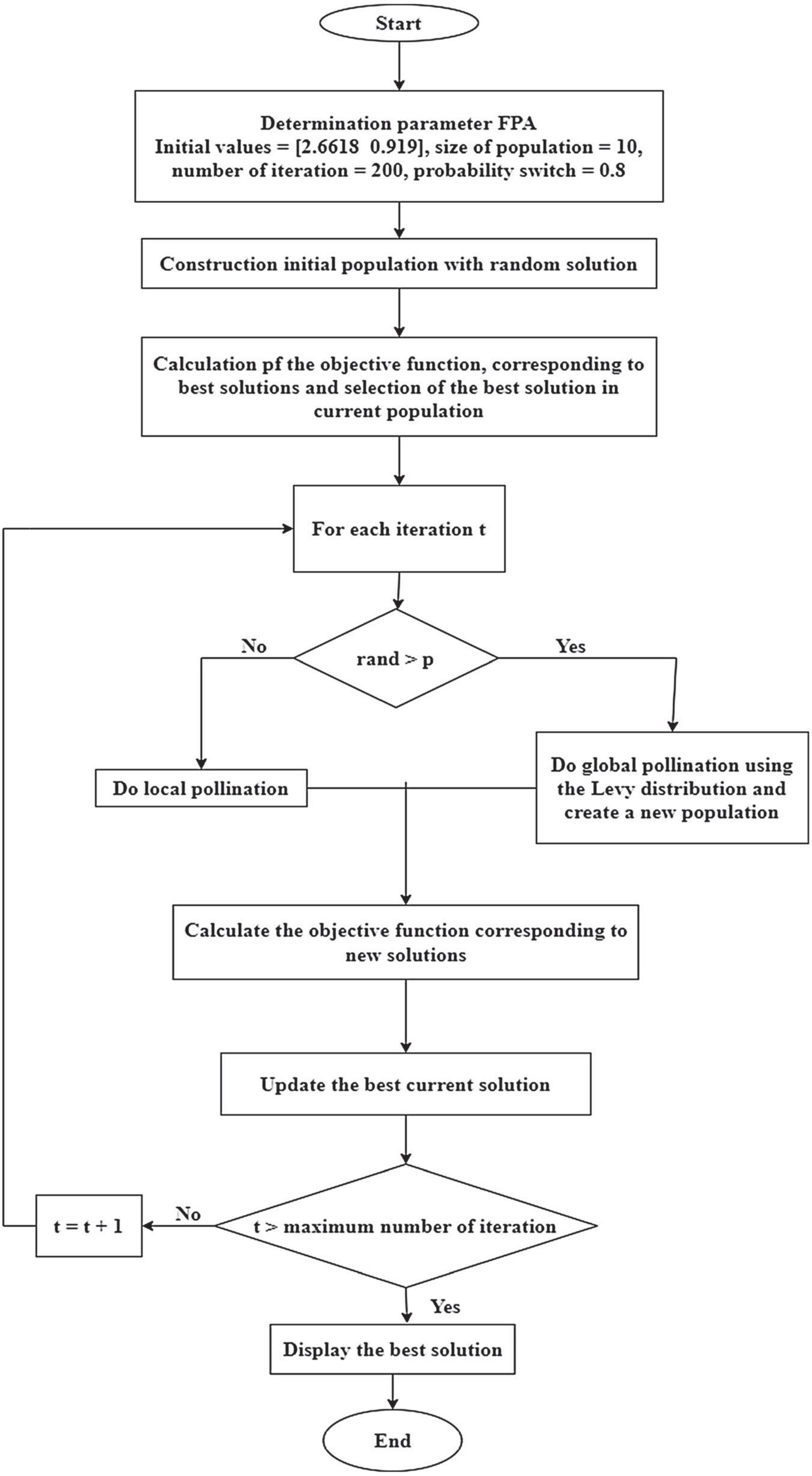

Start
Determination parameter FPA
Initial values = [2.6618 0.919], size of population = 10,
number of iteration = 200, probability switch = 0.8
Construction initial population with random solution
Calculation pf the objective function, corresponding to best solutions and selection of the best solution in current population
For each iteration t
rand > p
No
Yes
Do local pollination
Do global pollination using the Levy distribution and create a new population
Calculate the objective function corresponding to new solutions
Update the best current solution
t > maximum number of iteration
No
t = t + 1
Yes
Display the best solution
End

Figure 7. Flow chart of Flower Pollination Algorithm with initial parameters.

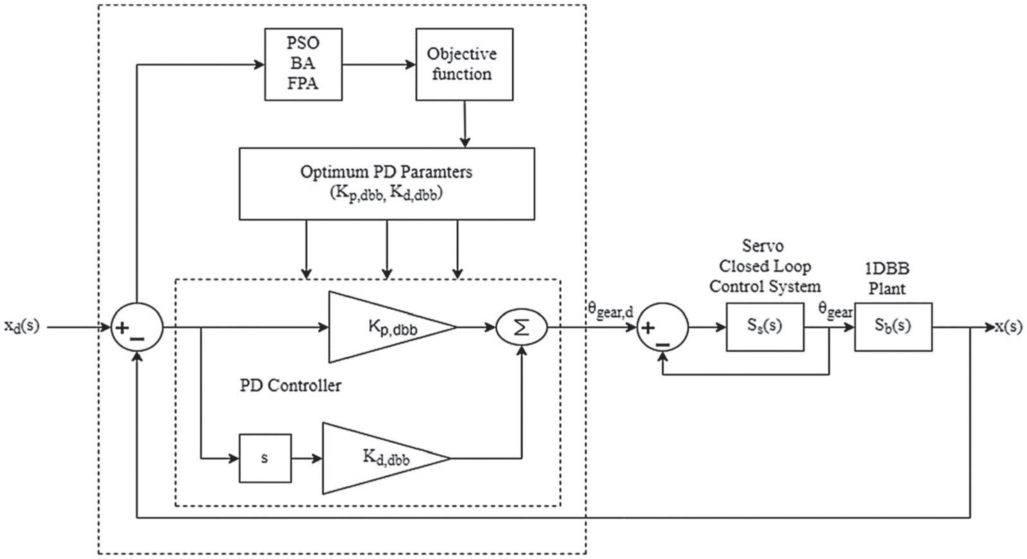


Figure 8. Position Control using PSO, BA, FPA.

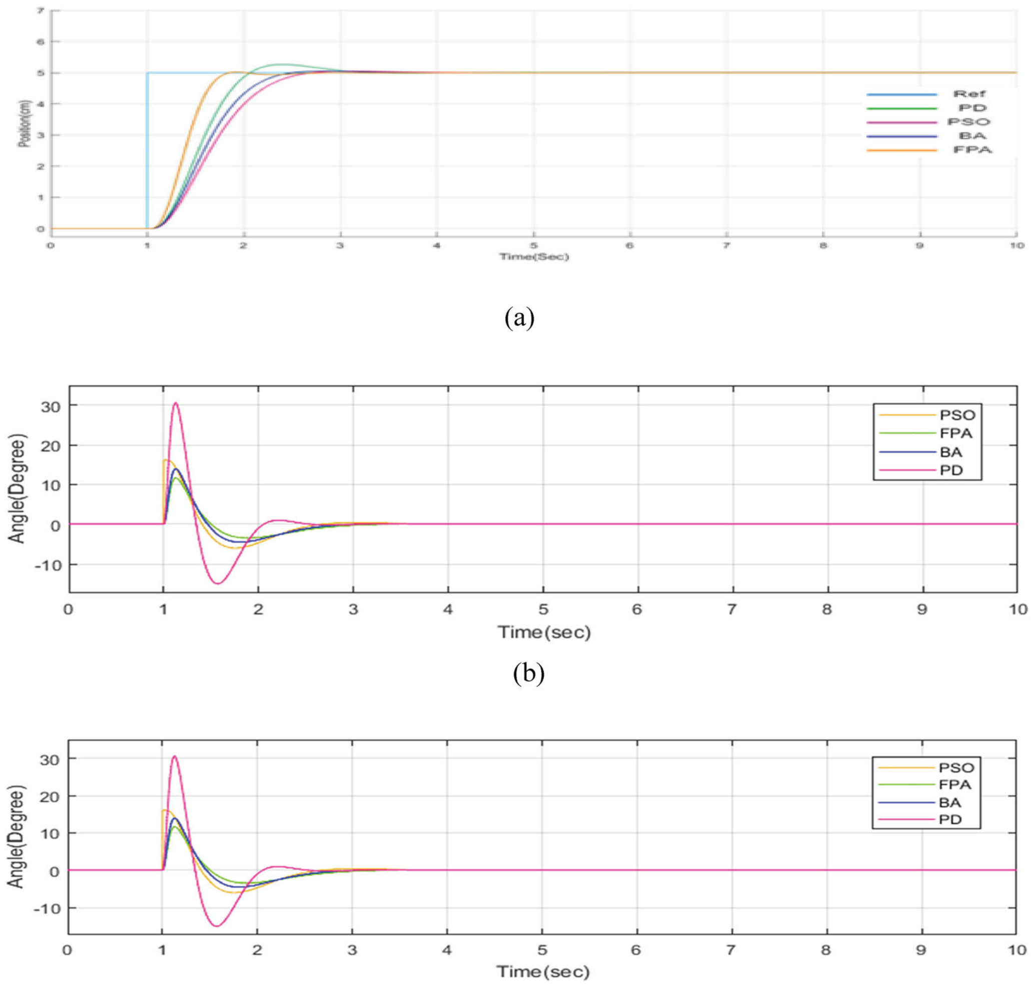
Table 1. Performance parameters for various control using PSO, BA and FPA on ball balancer system using Simulink.

Controllers	Peak time (tp) (s)	Settling time (ts) (s)	Peak overshoot (Mp) (%)
PSO	1.57	2.21	30.2
BA	1.45	2.178	25.1
FPA	1.43	2.16	20.8

the ball is within a defined range. As a result, the controller's effectiveness is demonstrated by the least difference between the initial and final positions. The FPA has a minimum final position and receives the target value in a short amount of time in this case, then the BA algorithm delivers the minimum position, and finally the PSO algorithm holds the minimum position. In addition, figure. 9b shows the ball's servo angle reaction on the x-axis, exhibiting the servo motor's control angle fluctuation. The minimum control angle determines a controller's precision in achieving balancing control for a ball balancer system. The FPA has a lower control angle than the BA and PSO controls in this circumstance, but the BA provides a better result than the PSO. In the FPA algorithm, the plate moves slowly while balancing the ball, which helps the system achieve a constant response. Figure 9c depicts the servo input voltage fluctuation as a function of the controller action. The servo units for the FPA optimization control are noted for running at a lower voltage and settling down earlier than the other controller. Because the increased FPA achieves the smallest position control and balancing angle, the speed of the servo motors is reduced to the smallest value possible, ensuring that the ball remains positioned on the plate while following the appropriate path. As a result, the FPA performs better than the BA and PSO when compared

In addition, time domain specifications are derived to examine the performance of PSO, BA, and FPA approaches, with the results displayed in Table 1. The peak overshoot of PSO is 30.2 percent, causing massive oscillations and making it impossible to balance the ball on the plate, according to the results. In contrast, as seen in the graph, FPA has a good response to peak overshoot of 20.8 percent and exhibits perfect ball-on-plate balance with less oscillation.

The integral of squared time-multiplied square of the error (ISTSE) value of PSO, BA, and FPA optimizations is also determined during the operation of the ball balancer for square trajectory in terms of ball position. The results are tabulated as shown in Table 2.



(a)

(b)

(c)

Figure 9. A position of the ball on the x-axis, b servo angle response of ball on the x-axis, c input voltage applied to the servo motor for the x-axis.

Table 2. Integral of squared time-multiplied square of the error (ISTSE) for position during simulation results.

	ISTSE
Controller	Ball position
PSO	45.337023
BA	45.446932
FPA	44.375275

PSO, BA, and FPA all produced outcomes that were near to one other when compared using the ISTSE performance measure. The FPA, on the other hand, achieved a better outcome and provides great position control of the ball on the plate in the ball balancer system.

5. Conclusion

This work uses three different optimal strategies to set the parameters of proportional derivative control to achieve self-balancing and position control of a two-degree-of-freedom balancer system: PSO, BA, and FPA are three different types of PSO. Simulation findings show that the developed strategy improves performance significantly within the context of the standard control structure. On the basis of time response analysis, the outcomes of the established control approaches are validated. On the ball balancer system, the provided controller has adaptability and good control performance. According to the findings, the FPA optimised technique performs BA and PSO in terms of ISTSE, settling time, peak time, and peak overshoot.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Ajit Kumar Sharma http://orcid.org/0000-0002-5678-7856

References

[1] Murray RM, Astrom KJ, Boyd SP, et al. Future directions in control in an information-rich word. IEEE Control Syst Mag. 2003;23:20–33. doi:10.1109/MCS.2003.1188769.

[2] Nelles O. Nonlinear system identification. Berlin, Heidelber: Springer; 2001.

[3] Boubaker O. The inverted pendulum: A fundamental benchmark in control theory and robotics. Int Conf Educ e-Learning Innov. 2012.

[4] Chalupa P, Přikryl J, Novák J. Adaptive control of Twin ROTOR MIMO system. Proc 2015 20th Int Conf Process Control. PC 2015. 2015: 314–319. doi:10.1109/PC.2015.7169982.

[5] Nowopolski K. Ball-and-beam laboratory system controlled by Simulink model through dedicated microcontrolled-Matlab data exchange protocol. Comput Appl Electr Eng. 2013;11:310–320.

[6] Aranda J, Chaos D, Dormido-Canto S, et al. Benchmark control problems for a non-linear underactuated hovercraft: A simulation laboratory for control testing. IFAC Proc. 2006;39:463–468. doi:10.3182/20060621-3-ES2905.00080

[7] Acosta JA. Furuta's Pendulum: A conservative nonlinear model for theory and practise. Math Probl Eng. 2010;2010. doi:10.1155/2010/742894.

[8] Awtar S, Bernard C, Boklund N, et al. Mechatronic design of a ball-on-plate balancing system. Mechatronics (Oxf). 2002;12:217–228. doi:10.1016/S0957-4158(01)00062-9.

[9] Aguilar-Avelar C, Moreno-Valenzuela J. New feedback linearization-based control for arm trajectory tracking of the furuta pendulum. IEEE/ASME Trans Mechatronics. 2016;21:638–648. doi:10.1109/TMECH.2015.2485942.

[10] Rudra S, Barai RK, Maitra M. Block backstepping control of the underactuated mechanical systems. In: Block backstepping design of nonlinear state feedback control Law for underactuated mechanical systems. Singapore : Springer Singapore; 2007. p. 31–52.

[11] Moreno-Valenzuela J, Aguilar-Avelar C. Feedback Linearization Control of the Furuta Pendulum. Presented at the (2018).

[12] Spong MW. Partial feedback linearization of underactuated mechanical systems. In: proceedings of IEEE/RSJ international conference on intelligent robots and systems (IROS'94). p. 314–321. IEEE.

[13] Ortega R, Spong MW, Gomez-Estern F, et al. Stabilization of a class of underactuated mechanical systems via interconnection and damping assignment. IEEE Trans Automat Contr. 2002;47:1218–1233. doi:10.1109/TAC.2002.800770.

[14] Sun S, Li L. The study of ball and plate system based on non-linear PID. Appl Mech Mater. 2012;187:134–137. doi:10.4028/www.scientific.net/AMM.187.134.

[15] Mochizuki S, Ichihara H. Generalized kalman-YakubovichPopov lemma based I-PD controller design for ball and plate system. J Appl Math. 2013;2013:1–9. doi:10.1155/2013/854631.

[16] Pinagapani AK, Mani G, K R C, et al. Composite disturbance rejection control for ball balancer system. Procedia Comput. Sci. 2018;133:124–133.

[17] Ali HI, Jassim HM, Hasan. Optimal nonlinear model reference controller design for ball and plate system. Arab J Sci Eng. 2019;44:6757–6768. doi:10.1007/s13369-018-3616-1.

[18] Bang H, Lee. Implementation of a ball and plate control system using sliding mode control. IEEE Access. 2018;6:32401–32408. doi:10.1109/ACCESS.2018.2838544.

[19] Das A, Roy P. Improved performance of cascaded fractionalorder smc over cascaded smc for position control of a ball and plate system. IETE J Res. 2017;63:238–247. doi:10.1080/03772063.2016.1258336.

[20] Kao S-T, Ho M-T. Second-order sliding mode control for ball-balancing system. In: 2018 IEEE conference on control technology and applications (CCTA). p. 1730–1735. IEEE (2018).
[21] Bang H, Lee Y. Embedded model predictive control for enhancing tracking performance of a ball-and-plate system. IEEE Access. 2019;7:39652–39659. doi:10.1109/ACCESS.2019.2907111.
[22] Zhang Z, Yuan D. Modelling and control scheme of the ball–plate trajectory-tracking pneumatic system with a touch screen and a rotary cylinder. IET Control Theory Appl. 2010;4:573–589. doi:10.1049/iet-cta.2008.0540.
[23] Marco A, Moreno-Armendariz CAP-O. Floriber to Ortiz Rodrıguez, E.R.: Indirect hierarchical FCMAC control for the ball and plate system. Neuro Comput. 2010;73:2454–2463. doi:10.1016/j.neucom.2010.03.023.
[24] Dong X, Zhao Y, Xu Y, et al. Design of PSO fuzzy neural network control for ball and plate system. Int J Innov Comput Inf Control. 2011;7:7091–7103.
[25] Yang XS. A new metaheuristic bat-inspired algorithm. In: Nature-inspired cooperative strategies for optimization (NICSO 2010). Studies in computational intelligence, Vol. 284. Berlin: Springer; 2010. p. 65–74.
[26] Altringham JD. Bats: biology and behaviour. Oxford: Oxford University Press; 1998.
[27] Bell WJ. Searching behaviour: the behavioural ecology of finding resources. London: Chapman & Hall; 1991.
[28] Yang XS. Nature-Inspired metaheuristic algorithms. Luniver Press; 2008.
[29] Wang Y, Jin Q, Zhang R. Improved fuzzy PID controller design using predictive functional control structure. ISA Trans. 2017;71:354–363. doi:10.1016/j.isatra.2017.09.005.
[30] Wang J, Zhu Y, Qi Ret al. Adaptive PID control of multi-DOF industrial robot based on neural network. J Ambient Intell Humaniz Comput. 2020. doi:10.1007/s12652-020-01693-w.
[31] Abdo MM, Vali AR, Toloei AR, et al. Stabilization loop of a two axes gimbal system using self-tuning PID type fuzzy controller. ISA Trans. 2014;53:591–602. doi:10.1016/j.isatra.2013.12.008.
[32] Zhang J, Zhuang J, Du H, et al. Self-organizing genetic algorithm based tuning of PID controllers. Inf Sci (NY). 2009;179:1007–1018. doi:10.1016/j.ins.2008.11.038.
[33] Hou Y-Y. Design and implementation of EP-based PID controller for chaos synchronization of Rikitake circuit systems. ISA Trans. 2017;70:260–268. doi:10.1016/j.isatra.2017.04.016.
[34] Chang Y-H, Chang C-W, Tao C-W, et al. Fuzzy sliding-mode control for ball and beam system with fuzzy ant colony optimization. Expert Syst Appl. 2012;39:3624–3633. doi:10.1016/j.eswa.2011.09.052.
[35] Angelov PP. Handbook on computational intelligence. Singapore: World Scientific; 2016.
[36] Li X, Clerc M. Swarm intelligence, In: Handbook metaheuristics. Cham, Switzerland: Springer; 2019. p. 353–384.
[37] Kennedy J. Swarm intelligence, In: Handbook of nature-inspired and innovative computing. Springer; 2006. p. 187–219.
[38] Gao K, Cao Z, Zhang L, et al. A review on swarm intelligence and evolutionary algorithms for solving _exible jobshop scheduling problems. IEEE/CAA J Automatica Sin. 2019;6(4):904–916.
[39] Engelbrecht A. Fundamentals of computational swarm intelligence. Hoboken (NJ): Wiley; 2006.
[40] Bahel V, Peshkar A, Singh S. Swarm intelligence-based systems: A review, In: Proc. Int. Conf. Comput. Sci. appl. Singapore: Springer; 2020. p. 149–156.
[41] Manne JR. Swarm intelligence for multi-objective optimization in engineering design, In: Advanced methodologies and technologies in artificial intelligence, computer simulation, and human-computer interaction. Hershey, PA, USA: IGI Global; 2019. p. 180–194.
[42] Kennedy J, Eberhart R. Particle swarm optimization, In: Proc. IEEE ICNN, Vol. 4.; 1995. p. 1942–1948.
[43] Kazemian M, Ramezani Y, Lucas C, et al. Swarm clustering based on flowers pollination by artificial bees. In: Swarm intelligence in data mining. Berlin: Springer; 2006. p. 191–202.
[44] Yang XS, Gandomi AH, Talatahari S, et al. (2012). Metaheuristics in water, geotechnical and transport engineering.
[45] Electronics M. Faulhaber DC-Micromotors Series 2338.

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-14512-z

Probabilistic intuitionistic fuzzy c-means algorithm with spatial constraint for human brain MRI segmentation

Rinki Solanki[1] · Dhirendra Kumar[2]

Received: 9 August 2021 / Revised: 14 July 2022 / Accepted: 31 January 2023

Abstract

Segmentation of brain MRI images becomes a challenging task due to spatially distributed noise and uncertainty present between boundaries of soft tissues. In this work, we have presented intuitionistic fuzzy set theory based probabilistic intuitionistic fuzzy c-means with spatial neighborhood information method for MRI image segmentation. We have investigated two well known negation functions namely, Sugeno's negation function and Yager's negation function for representing the image in terms of intuitionistic fuzzy sets. The proposed approach takes leverage of intuitionistic fuzzy set theory to address vagueness and uncertainty present in the data. The spatial neighborhood information term in the segmentation process is included to dampen the effect of noise. The segmentation performance of the proposed method is evaluated in terms of average segmentation accuracy and Dice score. Further, the comparison of the proposed method with other similar state-of-art methods is carried out on two publicly available brain MRI dataset which shows the significant improvements in segmentation performance in terms of average segmentation accuracy and Dice score. The proposed approach achieves on average 91% average segmentation accuracy in the presence of noise and intensity inhomogeneity on BrainWeb simulated dataset, which outperformed the state-of-art methods.

Keywords Intuitionistic fuzzy c-mean algorithm · Clustering · Magnetic resonance imaging · Image segmentation · Unsupervised machine learning

✉ Dhirendra Kumar
dhirendrakumar@dtu.ac.in

Rinki Solanki
rinkisolanki21@gmail.com

[1] Times Internet, Gurugram, India

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

Published online: 07 March 2023

1 Introduction

In the recent past, diagnostics have been revolutionized with the advancement of many medical imaging modalities such as positron emission tomography (PET), magnetic resonance imaging (MRI), computed tomography (CT), Mammogram, X-rays, Ultrasound etc. These modalities help in delineating the human anatomy for disease diagnosis. Among all, MRI [26] is the frequently used modality for capturing the soft tissues present in the human brain such as gray matter(GM), white matter(WM) and cerebrospinal fluids (CSF). The image sequences [14] are captured in MR images by applying an appropriate setting of pulse parameters such as repetition time (TR), echo time (TE), spin-echo, gradient-echo, inversion-recovery etc. TE and TR are the two key parameters for obtaining different image contrast. Due to this, the MRI machines can delineate the multi-spectral image with high contrast. Nowadays, these diagnostic machines are easily accessible which produce huge amount of medical data for disease diagnosis. Manual analysis of these images for disease diagnosis requires the expert radiologist. This being a time consuming process and may involve human error. There is a requirement of analyzing these MRI images in less time for faster diagnosis. The computer aided diagnosis [8] may help the expert radiologist in faster analysis of medical images. In some situations, the quantification and localization of different normal and abnormal tissues are required for brain related diseases using MRI modality. For this, these MRI images need to be segmented in different similar regions. The manual segmentation of MRI images is a challenging task as images are likely to have artifacts during the delineation process. The main factors affecting the quality of MRI segmented images includes (a) a non-uniform intensity variation is introduced in the MRI images. This variation is due to radio frequency utilized in the MRI, termed as bias field effect or intensity in-homogeneity (IIH) or intensity non-uniformity (INU) [1]; (b) noise; (c) partial volume effect. The presence of such artifacts adversely affect segmentation as well as visual evaluation based on absolute pixel intensities [13].

Machine learning (ML) based techniques are the most extensively used for segmenting brain MR images. These techniques are further classified into supervised and unsupervised techniques. The supervised segmentation techniques are fully automatic and effective segmentation approaches [2, 10, 16, 29, 42, 47, 48]. Although the segmentation accuracy is improved by the supervised ML techniques by incorporating prior knowledge, the major drawbacks of supervised techniques are as follows [2]: (a) training classifier with the same training set for a large number of MR images may often lead to biased results due to physiological variability between different subjects; (b) several parameters are required by the classifiers to be trained, thus necessitating the requirement of fast processing devices with large amount of main memory.

Unsupervised segmentation techniques [46] can be described as partitioning the image into different groups or regions, each having alike features such as texture, color, etc. Clustering is one of the popular unsupervised techniques to explore and analyze the structural information associated with the unlabeled data. The conventional way of obtaining clusters is the Hard c-means (HCM) clustering method, which results in c-crisp partitions of the data set [39]. 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 it's interpretability for many real world applications.

Fuzzy c-means (FCM) [5] overcomes this problem by assigning membership values to each data point to c number of clusters where each cluster is represented by fuzzy sets. FCM [6] is the most widely used clustering algorithm for segmenting brain MR images [8, 9, 23].

The reason for wide acceptance of FCM for MRI image segmentation is its ability to handle (a) uncertainty present in image boundaries/regions; (b) imprecise gray levels in images; (c) vagueness in defining class. The performance of FCM degrades in presence of imaging artifacts because it does not consider any spatial information [52]. In the past, many research work has been done by incorporating the local spatial information to the FCM clustering algorithm [1, 11, 13, 33, 36, 37, 41, 45, 48, 50, 56, 62, 63]. Several other research work related to brain MRI segmentation also reported in [4, 31, 32, 49] etc.

The methods discussed so far are dependent on selection of optimal parameter values and lose their fine image details. Krindis et al. [33] addressed this issue by proposing a fuzzy local information clustering method (FLICM) to tackle the problem of noise in image segmentation. This method is similar to FCM_S [1] as it uses the neighboring pixels deviation from centroid's intensity, weighted by a fuzzy factor and spatial distance of neighbours. The FLICM doesn't take into account any parameter but calculates the local information term for each iteration and hence makes it a time consuming segmentation method. The literature reports that the objective value is not minimized further by FLICM rather converging the fuzzy partition matrix only. Guo et al. proposed an Adaptive fuzzy c-means (NDFCM) [22] method, which is based on local noise detection. In this method, the spatial parameters for each pixel were dependent on the noise level in a given immediate neighbourhood. In spite of being the noise adaptive algorithm, NDFCM has a high computational complexity because it depends on the three input parameters which are required to be fine tuned for good performance. Recently a fast and robust fuzzy c-means algorithm (FRFCM) was proposed by Lei et al., which gave magnificent results with significantly low time complexity [35]. The pre-processing step in FRFCM employed morphological reconstruction operation, which made it robust to a variety of noises. The post processing step uses membership filtering for avoiding the heavy computation in measuring the distance between the neighbour pixels and centroids to handle noisy pixels. The FRFCM performs well for several noise varieties, but shows its poor performance on high noise samples because the sharp edges and shapes are not preserved. In another research work Deviation-sparse fuzzy c-means with neighbor information constraint (DSFCMN) algorithm [60] is proposed, which modeled the deviation between the original pixel values and measured noisy pixels value as residual and incorporated this value in the optimization function. The residual term in DSDCMN is sparse matrix and uses the L1 norm distance measure in objective function as a constraint over residuals. However DSFCMN did not show good results when tested on a dataset with high noise. Further, Wang et. al. proposed Weighted Residual fuzzy c-means (WRFCM) [55], which uses weighted L2-norm measure for residual estimation and showed satisfactory performance compared to the previous research methods.

In order to deal with non-linear structure present in any image, many research methods have been reported in literature that utilize the kernel distance measure. The research work [61] proposed a kernel generalized fuzzy c-means (KGFCM) clustering with spatial information for image segmentation. Most kernel based methods are dependent on optimal selection of input parameters values for satisfactory segmentation performance. The grid search method is mostly used to find the optimal values of these parameters which is a time consuming process. Gong et al. [21] 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. The performance of KWFLICM method is better in comparison to the FLICM method but still it inherits the problem of FLICM method.

The membership values in variants of FCM depend on the distance between cluster centroids and image pixels. In some situations, the image acquisition process leads to uncertainty due to imprecise pixel intensity value. Hence, calculation of membership values of a given pixel to different clusters is imprecise [44]. Therefore to handle such problems, an intuitionistic fuzzy set (IFS) introduced by Atanassov [3] that deals with imprecise and vagueness in defining the membership value [12, 28]. For this, IFS set includes non-membership and hesitancy components along with membership value. The introduction of IFS theory into the clustering process increases the segmentation accuracy. Further, it makes the segmentation method robust and faster in comparison to FCM algorithm [27]. The research work [57], suggested a fuzzy clustering of data represented in terms of IFS which utilizes the Euclidean distance measure [51] defined for IFS. Chaira [12] introduced the concept of IFS theory to incorporate hesitation in defining the membership value in FCM algorithm. The research work [12] increases the significant data points in a given cluster. The problem of variations in pixel intensities is studied in the research work [18] which utilizes the IFS theory to represent the MRI images in terms of IFSs and further these data are clustered for image segmentation. PIFCM [40] is a recently proposed clustering algorithm which uses probabilistic Euclidean distance measure (PEDM) in the objective function. The presence of PEDM in the PIFCM have shown following advantages over conventional IFCM algorithms: (1) It is an adaptive algorithm, as it uses probabilistic weights; (2) reduced number of iterations for convergence; (3) lower sensitivity towards the fuzzy factor *m*, therefore, leads to higher stability. Further, the research work [53] suggested an improved Probabilistic Intuitionistic Fuzzy c-Means Clustering Algorithm. The improved PIFCM uses the min-max normalization as a membership function which minimizes the matrix computation of the original PIFCM. The PIFCM and Improved PIFCM handle the uncertainty in the dataset very well but are susceptible to the noisy dataset as in the case of MRI images. The performance of IFS theory based clustering method for image segmentation process deteriorates in presence of noise. To handle noise, the incorporation of local spatial information is advocated in literature.

The research work [25] proposed neighborhood information based IFCM algorithm with genetic algorithm (NIFCMGA) for automatic optimal parameter selection. It reduces the effect of noise and outliers in medical images segmentation but consumes more time as it utilizes genetic algorithm. The research work [54] suggested improved IFCM (IIFCM) to handle noise which combines both local spatial and grey level information together for MRI segmentation. Their algorithm is free from parameter tuning but have considerably higher running time. The research work [34] proposed IFCM with spatial neighborhood information (IFCM-SNI). The spatial neighborhood information (SNI) term is incorporated in the objective function of IFCM algorithm and is capable of dealing with noise without losing the fine image details. Their model gives better results on highly noisy MRI images.

From the above discussion, it is evident that noisy pixels can be correctly classified by incorporating spatial neighborhood information in the image segmentation process. The performance of the PIFCM [40] method is not giving promising results for image segmentation in presence of noise. To address this issue, we have proposed a intuitionistic fuzzy clustering that uses probabilistic Euclidean distance measure with spatial constraints (PIFCM_S). The proposed PIFCM_S method utilizes a spatial regularization term in the optimization problem for obtaining the clusters. This spatial regularization term utilizes the mean filtered image to dampen the effect of noise with a regularization parameter. The spatial regularization parameter sets a trade off between the level of noise and the segmentation performance.

Higher the noise in the image, the value of this regularization parameter should be high. Further, we have investigated two well known intuitionistic fuzzy generators, namely, Sugeno's negation function and Yager's negation function for representing the image in terms of IFS. To validate the performance of the proposed method, we have utilized two publicly available brain MRI image dataset. Further, the performance of the proposed method is compared with several state-of-the-art methods in terms of average segmentation accuracy and Dice score.

The rest of the paper is organized as follows: preliminaries and related works are included in Section 2. The PIFCM_*S* algorithm and its formulation is discussed in Section 3. Section 4 discusses experimental setup and results. Finally, conclusion is included in Section 5.

2 Preliminaries and related works

The description of notations and related work used throughout the paper are discussed in the section.

The fuzzy set (FS) F, is defined by using membership function $\mu_F(x),\ x \in X$ and $\mu_F(x) \in [0, 1]$

Intuitionistic Fuzzy Set (IFS) [3], A is defined using membership function $\mu_A(x)$ and non-membership function $\nu_A(x)$ and is represented as:

$$A = \{\langle x, \mu_A(x), \nu_A(x)\rangle | x \in X\} \tag{1}$$

Here $\mu_A : X \rightarrow$ [0,1] and $\nu_A : X \rightarrow$ [0,1] simultaneously assigns membership value and non-membership value respectively to each element $x \in X$ with respect to A, if

$$0 \leq \mu_A(x) + \nu_A(x) \leq 1. \tag{2}$$

For every $x \in X$ in A, If $\nu_A(x) = 1 - \mu_A(x)$, then set A reduces to fuzzy set.

In an IFS, the hesitancy value, $\pi_A(x)$ defines the uncertainty in definition of membership function and is calculated as:

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x),\ \text{where } 0 \leq \pi_A(x) \leq 1. \tag{3}$$

Hence, due to presence of hesitancy value in IFS, the membership value lies in the interval $[\mu_A(x), \mu_A(x) + \pi_A(x)]$.

2.1 Construction and representation of intuitionistic fuzzy sets for gray images

The image acquisition process involves conversion of energy response received on sensing devices to gray levels. This introduces the imprecise estimation of gray levels for many of the pixels in the image which in turn includes uncertainty in representing the gray levels in the image. This issue is resolved by converting the medical image into an intuitionistic fuzzy domain. In this way, a given gray level corresponding to a pixel is represented using membership value, non-membership value and hesitancy value. The membership value for a given pixel in the gray image is obtained by normalizing in the range [0 1]. The non-membership value and hesitancy value for the pixel is calculated using the membership value through intuitionistic fuzzy generators (discussed below). We have used two intuitionistic fuzzy generator functions namely, Yager negation function [58] and Sugeno negation function [43] for our study.

An intuitionistic fuzzy generator [12] is a function $g : [0, 1] \rightarrow [0, 1]$ satisfying the following properties :

1. $g(\mu) \leq 1 - \mu$ for all $\mu \in [0, 1]$,
2. $g(0) = 1$ and $g(1) = 0$

If g is continuous, decreasing (increasing) then the intuitionistic fuzzy generator is called continuous, decreasing (increasing). The non-membership function $NM(\mu)$ for a given generating function g(.) is defined as:

$$NM(\mu) = g^{-1}(g(1) - g(\mu)) \tag{4}$$

where, $g^{-1}(.)$ is inverse of generating function $g(.)$.

- Yager's negation function (YNF) [58, 59]: The Yager's generating function $g_Y(\mu)$ with negation parameter β is given as follows:

  $$g_Y(\mu) = \mu^{\beta} \tag{5}$$

  Its inverse $g_Y^{-1}(\mu)$ is given by:

  $$g_Y^{-1}(\mu) = \mu^{\frac{1}{\beta}} \tag{6}$$

  Yager's negation function calculates the non-membership value using (4), (5) and (6) which is given by:

  $$\nu_A(x) = NM(\mu_A(x)) = (1 - \mu_A(x)^{\beta})^{\frac{1}{\beta}}, \ \beta > 0 \tag{7}$$

  where $\mu_A(x)$ represents membership value of IFS A.

- Sugeno's negation function (SNF) [43]: The Sugeno's generating function $g_S(\mu)$ with negation parameter β is given as:

  $$g_S(\mu) = \frac{1}{\beta}\log(1 + \beta\mu), \ \ \beta > 0 \tag{8}$$

  Its inverse $g_S^{-1}(\mu)$ is given by:

  $$g_S^{-1}(\mu) = \frac{1}{\beta}(\exp(\beta\mu) - 1), \ \ \beta > 0 \tag{9}$$

  Sugeno's negation function calculates the non-membership value using (4), (8) and (9) which is given by:

  $$\nu_A(x) = NM(\mu_A(x)) = \frac{1 - \mu_A(x)}{1 + \beta\mu_A(x)}, \ \beta > 0 \tag{10}$$

  where $\mu_A(x)$ represents membership value of IFS A.

The intuitionistic fuzzy generator defined above is used to construct the intuitionistic fuzzy data for gray image. Let X be the set of p number of pixel and x_i represent the pixel intensity value corresponding to i^{th} pixel in X, where $i \in \{1, 2, \ldots p\}$. Therefore, each pixel in an image can be represented by an IFS as $X^{IFS} = \{\langle \mu_X(x_i), \nu_X(x_i), \pi_X(x_i)\rangle \mid i = 1, 2, \ldots, p\}$, where $\mu_X(x_i)$ is membership value obtained by normalization of image in range [0 1] and $\nu_X(x_i)$ is non-membership value calculated using negation function described in (7) and (10) corresponding to Yager's negation function and Sugeno's negation function respectively.

The probabilistic intuitionistic fuzzy distance measure between i^{th} element X_i^{IFS}= $\langle \mu_X(x_i), \nu_X(x_i), \pi_X(x_i)\rangle$ and j^{th} element X_j^{IFS}= $\langle \mu_X(x_j), \nu_X(x_j), \pi_X(x_j)\rangle$ of IFS X^{IFS} can be defined as [40]

$$\tilde{d}_2(X_i^{IFS}, X_j^{IFS}) = \left[\frac{1}{2}\Big(p_{ij}(\mu_X(x_i) - \mu_X(x_j))^2 + q_{ij}(\nu_X(x_i) - \nu_X(x_j))^2 + \rho_{ij}(\pi_X(x_i) - \pi_X(x_j))^2\Big)\right]^{1/2} \quad (11)$$

Here the probabilistic weights p_{ij}, q_{ij} and ρ_{ij} corresponding to the membership value, non-membership value and hesitancy value respectively are data driven. The weight ρ_{ij} corresponding to the hesitancy value is computed using the following formula of correlation coefficient.

$$\rho_{ij} = 1 - \frac{\omega}{3(1+\omega)} \quad (12)$$

where $\omega = |\mu_X(x_i) - \mu_X(x_j)| + |\nu_X(x_i) - \nu_X(x_j)| + |\pi_X(x_i) - \pi_X(x_j)|$.

2.2 Fuzzy clustering with spatial constraints

An approach was proposed in the research work [1] to increase the robustness of FCM to noise by an addition of a penalty term in the FCM objective function. The penalty term makes the smoothing of a pixel within its specified neighborhood. The modified objective function of FCM_S algorithm [1] is given as:

$$J_m(U, V : X) = \sum_{i=1}^{p}\sum_{j=1}^{c} u_{ij}^m \left\| x_i - v_j \right\|^2 + \frac{\alpha}{N_R}\sum_{i=1}^{p}\sum_{j=1}^{c} u_{ij}^m \sum_{r\in N_i} \left\| x_r - v_j \right\|^2 \quad (13)$$

Here $X = \{x_1, x_2, \ldots, x_p\}$ are p pixels, m $(1 < m < \infty)$ is the fuzzification factor, c $(1 < c < p)$ represents the number of clusters which are fixed, u_{ij} $(0 \leq u_{ij} \leq 1)$ represents the membership degree for i^{th} pixel in j^{th} cluster, N_i denotes the number of neighboring pixels around the center pixel x_i and N_R is cardinality of N_i. The parameter α controls the trade-off effects of the neighboring pixel. The optimization problem (13) can be solved by the Lagrange method of undetermined multipliers. Membership value and cluster centroid are given as [1]:

$$u_{ij} = \frac{\left(\left\| x_i - v_j \right\|^2 + \frac{\alpha}{N_R}\sum\limits_{r\in N_i} \left\| x_r - v_j \right\|^2\right)^{-\frac{1}{m-1}}}{\sum\limits_{k=1}^{c}\left(\| x_i - v_k \|^2 + \frac{\alpha}{N_R}\sum\limits_{r\in N_i} \| x_r - v_k \|^2\right)^{-\frac{1}{m-1}}} \quad (14)$$

$$v_j = \frac{\sum\limits_{i=1}^{p} u_{ij}^m \left(x_i + \frac{\alpha}{N_R}\sum\limits_{r\in N_i} x_r\right)}{(1+\alpha)\sum\limits_{i=1}^{p} u_{ij}^m} \quad (15)$$

The value $\frac{1}{N_R}\sum\limits_{r\in N_i} x_r$ in (15) represents the mean value of gray-level around the pixel x_i within a specified window. However, FCM_S algorithm have high computation time. In order to decrease computation time of FCM_S algorithm, a variant of FCM_S algorithm, named the FCM_S1 is proposed in [13]. The mean filtered image in FCM_S1 consists of its neighbor average gray values around each pixel within a window. The objective function of FCM_S1 algorithm is given as:

$$J_m(U, V : X) = \sum_{i=1}^{p}\sum_{j=1}^{c} u_{ij}^{m} \left\| x_i - v_j \right\|^2 + \alpha \sum_{i=1}^{p}\sum_{j=1}^{c} u_{ij}^{m} \left\| \bar{x_r} - v_j \right\|^2 \tag{16}$$

where $\tilde{x}_r$ represents the mean value of neighboring pixels around the pixel x_r and is computed in advance. The optimization problem (16) can be solved by the Lagrange method of undetermined multipliers. Membership value and cluster centroid are given as [13]:

$$u_{ij} = \frac{\left(\left\| x_i - v_j \right\|^2 + \alpha \left\| \bar{x_r} - v_j \right\|^2 \right)^{-\frac{1}{m-1}}}{\sum\limits_{k=1}^{c} \left(\| x_i - v_k \|^2 + \alpha \| \bar{x_r} - v_k \|^2 \right)^{-\frac{1}{m-1}}} \tag{17}$$

$$v_j = \frac{\sum\limits_{i=1}^{p} u_{ij}^{m} \left(x_i + \alpha \bar{x_r} \right)}{(1 + \alpha) \sum\limits_{i=1}^{p} u_{ij}^{m}} \tag{18}$$

The neighborhood term of the FCM_S algorithm is simplified in FCM_S1 algorithm. FCM_S is suitable for images which are contaminated by Gaussian noise. The parameter α controls the trade-off effect between the mean filtered image and original image. If the parameter α is set to zero, then both FCM_S and FCM_S1 reduce to the FCM algorithm. The outline of FCM_S1 algorithms [13] is given in Algorithm 1.

```
Input  : fuzzy factor ($m$); number of centroids (c); spatial regularization parameter
         ($\alpha$); tolerance level ($\epsilon$)
Output: fuzzy partition $U$, centroid $v_j$.
1. Compute the mean filtered image.
2. Initialize fuzzy partition matrix $U^{(k=0)}$.
3. $k \leftarrow 1$
Repeat
4. Update the cluster centroids $v_j^{(k+1)}$ using (18).
5. Update fuzzy partition matrix $U^{(k)} = [u_{ij}^{(k)}]_{p \times c}$ using
   (17).
6. $k \leftarrow k + 1$
   Until
7. $\|U^{(k+1)} - U^{(k)}\| < \epsilon$
8. Return $U^{(k+1)}$, $v_j^{(k+1)}$
```

Algorithm 1 FCM_S1 algorithm.

2.3 Probabilistic intuitionistic fuzzy C-Means algorithm

Probabilistic Intuitionistic Fuzzy C-Means (PIFCM) [40] is an adaptive IFS based clustering algorithm. It incorporates the advantage of IFS for handling uncertainty which arises due to imprecise and incomplete information. The peculiarity of PIFCM is that it assigns weights p_{ij}, q_{ij} and ρ_{ij} corresponding to membership, non-membership and hesitancy value respectively in the objective function (19) directly from the dataset. Therefore, this algorithm gives weightage to each data point in every cluster. PIFCM algorithm divides p data points into c clusters by optimizing the objective function through continuous updation of the centroid (v_j^{IFS}) and membership degree (u_{ij}) until the termination condition is achieved. The objective function of PIFCM was formulated as follows:

$$J_m(U, V^{IFS} : X^{IFS}) = \sum_{i=1}^{p}\sum_{j=1}^{c} u_{ij}^m \tilde{d}_2(X_i^{IFS}, v_j^{IFS})$$

$$\text{subject to, } \quad \sum_{j=1}^{c} u_{ij} = 1, \quad 1 \leq i \leq c \tag{19}$$

Here m is a fuzzy parameter, $X = \{x_i^{IFS}\}_{p\times 1}$ represents the image in terms of IFS, and the i^{th} element $X_i^{IFS} = \langle \mu_X(x_i), \nu_X(x_i), \pi_X(x_i)\rangle$, $U = [u_{ij}]_{p\times c}$ is the fuzzy partition matrix in which each entry u_{ij} represents the membership value of i^{th} data point into the j^{th} cluster, $V = \{v_j^{IFS}\}_{c\times 1}$ denotes cluster centroid and PEDM $\tilde{d}_2(X_i^{IFS}, v_j^{IFS})$ computes the distance between image pixel X_i^{IFS} and centroid pixel v_j^{IFS}. The weights p_{ij}, q_{ij} and ρ_{ij} is obtained using Algorithms 2, 3 and 4 respectively. The solution of the optimization problem given in (19) can be obtained using Lagrange method of undetermined multiplier which is given as:

$$u_{ij} = \left\{ \sum_{k=1}^{c} \left(\frac{\tilde{d}_2(X_i^{IFS}, v_j^{IFS})}{\tilde{d}_2(X_i^{IFS}, v_k^{IFS})} \right)^{\frac{2}{m-1}} \right\}^{-1} \tag{20}$$

$$\mu_V(v_j) = \frac{\sum_{i=1}^{p} p_{ij} u_{ij} \mu_X(x_i)}{\sum_{i=1}^{p} p_{ij} u_{ij}}, \quad \forall \ 1 \leq j \leq c \tag{21a}$$

$$\nu_V(v_j) = \frac{\sum_{i=1}^{p} q_{ij} u_{ij} \nu_X(x_i)}{\sum_{i=1}^{p} q_{ij} u_{ij}}, \quad \forall \ 1 \leq j \leq c \tag{21b}$$

$$\pi_V(v_j) = \frac{\sum_{i=1}^{p} \rho_{ij} u_{ij} \pi_X(x_i)}{\sum_{i=1}^{p} \rho_{ij} u_{ij}}, \quad \forall \ 1 \leq j \leq c \tag{21c}$$

The outline of PIFCM method is depicted in Algorithm 5 [40].

3 Probabilistic intuitionistic fuzzy c-means with spatial constraint (PIFCM_S)

The acquisition process in an image gives rise to noise, which may bring variation in the pixel intensity value. Hence, the noisy pixels show an anomalous behaviour in its adjacency which leads to incorrect segmentation of image. The PIFCM algorithm does not incorporate

```
**Input**  : $M = [mu_i]$ and $O = [pi_i]$
**Output**: $P = [p_i]$
    $mu'$:=Compute the minimum value of M(:).
    $sum$:=Compute the sum of $mu'$.
        If $sum$:$\neq 0$
          $temp = mu'/sum$.
        else
          $temp$ =1.
        End if
    min:=Compute the sum of $temp. * mu'$.
    $temp'$:=Compute the sum of $temp. * pi_i$.
    max:=$min + temp'$.
    $p_i$:=Compute the mean of min and max.
End for
```

Algorithm 2 Weight matrix P for membership values.

```
**Input**  : $N = [nu_i]$ and $O = [pi_i]$
**Output**: $Q = [q_i]$
    $nu'$:=Compute the minimum value of N(:).
    $sum1$:=Compute the sum of $mu'$.
        If $sum1$:$\neq 0$
          $temp1 = mu'/sum1$.
        else
          $temp1$ =1.
        End if
    min 1:=Compute the sum of $temp1. * mu'$.
    $temp$”:=Compute the sum of $temp1. * pi_i$.
    max 1:=$min1 + temp$”.
    $q_i$:=Compute the mean of min 1 and max 1.
End for
```

Algorithm 3 Weight matrix Q for non-membership values.

```
**Input**  : $M = [mu_i]$, $N = [nu_i]$ and $O = [pi_i]$
**Output**: $R = [\rho_i]$
    $temp1$ := Compute absolute difference of columns of M from each other.
    $temp2$ := Compute absolute difference of columns of N from each other.
    $temp3$ := Compute absolute difference of columns of R from each other.
    $temp$:= $temp1 + temp2 + temp3$ .
    $temp'$:=Compute the sum of $temp/(1 + temp)$.
    $\rho_i$ = 1-(1/3).*$temp'$.
End for
```

Algorithm 4 Weight matrix R for hesitancy values.

any spatial information in its objective function (19) to handle such noises and results in

```
Input  : Dataset D with $p \times 1$ dimensions; fuzzy factor $(m)$; intuitionistic fuzzy
         parameter $(\beta)$; number of centroids (c); tolerance level $(\epsilon)$
Output: Intuitionistic fuzzy partition $U$, centroid $v_j^{IFS}$.
1. Intuitionistic fuzzification of data.
2. Initialize intuitionistic fuzzy centroid $V^{IFS(k=0)}$.
3. $k \leftarrow 1$
   Repeat
4. Update intuitionistic fuzzy partition matrix
     $U^{(k)} = [u_{ij}^{(k)}]_{p \times c}$ using (20).
5. Update weight matrices P, Q and R (after normalization) using
     Algorithms 2 - 4.
6. Update the cluster centroids using (21)
     $v_j^{IFS(k+1)} = \left\langle \mu_V(v_j)^{(k+1)}, \nu_V(v_j)^{(k+1)}, \pi_V(v_j)^{(k+1)} \right\rangle$.
7. $k \leftarrow k + 1$
    Until
8. $\|v_j^{IFS(k+1)} - v_j^{IFS(k)}\| < \epsilon$
9. Return the membership degrees $U^{(k+1)}$ , the centroids $v_j^{IFS(k+1)}$
```

Algorithm 5 PIFCM algorithm.

poor segmentation performance. Secondly, the presence of noise in an image makes boundaries around the pixels sensitive and hence affecting the membership degree (20) of a given pixel to cluster. Therefore in this section, we formulate an optimization problem robust to noise, named probabilistic intuitionistic fuzzy c-means with spatial information (PIFCM_S). The inclusion of spatial regularization term in the optimization problem of PIFCM_S makes it robust to handle the problem of noise and uncertainty present between the boundaries in images in the segmentation process. The optimization problem of the PIFCM_S algorithm is defined as:

$$\begin{aligned} \min J_m(U, V^{IFS} : X^{IFS}) = & \sum_{i=1}^{p}\sum_{j=1}^{c} u_{ij}^m \tilde{d}_2^2(X_i^{IFS}, v_j^{IFS}) \\ & + \alpha \sum_{i=1}^{p}\sum_{j=1}^{c} u_{ij}^m \tilde{d}_2^2(\bar{X}_r^{IFS}, v_j^{IFS}) \end{aligned} \tag{22}$$

where, $U = [u_{ij}]_{p \times c}(0 \leq u_{ij} \leq 1)$ represents the fuzzy partition matrix, $X = \{x_i^{IFS}\}_{p \times 1}$ represents the image in terms of IFS, and the i^{th} element $X_i^{IFS} = \langle \mu_X(x_i), \nu_X(x_i), \pi_X(x_i) \rangle$, m is a fuzzy parameter, $V = \{v_j^{IFS}\}_{c \times 1}$ denotes cluster centroid, α is spatial regularization parameter value and should be tuned proportionally to the noise level present in the image, $\bar{X}_r^{IFS} = \langle \bar{\mu}_X(x_r), \bar{\nu}_X(x_r), \bar{\pi}_X(x_r) \rangle$ represents mean value of neighboring pixels around the pixel and $\tilde{d}_2(X_i^{IFS}, v_j^{IFS})$ computes PEDM between image pixel X_i^{IFS} and centroid pixel v_j^{IFS}. The Lagrange method of undetermined multiplier method is used to solve the optimization problem (22). The Lagrangian of optimization problem of PIFCM_S with ζ_i

as Lagrange multiplier is defined as:

$$L(U, V^{IFS},\ X^{IFS} : \zeta_i) = \sum_{i=1}^{p}\sum_{j=1}^{c} u_{ij}^{m}\tilde{d}_2^2(X_i^{IFS}, v_j^{IFS})$$

$$+\alpha\sum_{i=1}^{p}\sum_{j=1}^{c} u_{ij}^{m}\tilde{d}_2^2(\bar{X}_r^{\ IFS}, v_j^{IFS}) - \sum_{i=1}^{p}\zeta_i\left(\sum_{j=1}^{c} u_{ij} - 1\right) \tag{23}$$

Calculating partial derivative of L with respect to $\mu_V(v_j)$, $\nu_V(v_j)$ and $\pi_V(v_j)$ and equate them to zero, we have

$$\underset{\substack{1\le i\le p\\ 1\le j\le c}}{\forall}\ \frac{\partial L}{\partial \mu_V(v_j)} = \frac{\partial L}{\partial \nu_V(v_j)} = \frac{\partial L}{\partial \pi_V(v_j)} = 0 \tag{24}$$

Simplifying (24), $1 \le j \le c$ we obtain

$$\mu_V(v_j) = \frac{\sum\limits_{i=1}^{p} p_{ij}u_{ij}^{m}(\mu_X(x_i) + \alpha\bar{\mu}_X(x_r))}{(1+\alpha)\sum\limits_{i=1}^{p} p_{ij}u_{ij}^{m}} \tag{25a}$$

$$\nu_V(v_j) = \frac{\sum\limits_{i=1}^{p} q_{ij}u_{ij}^{m}(\nu_X(x_i) + \alpha\bar{\nu}_X(x_r))}{(1+\alpha)\sum\limits_{i=1}^{p} q_{ij}u_{ij}^{m}} \tag{25b}$$

$$\pi_V(v_j) = \frac{\sum\limits_{i=1}^{p} \rho_{ij}u_{ij}^{m}(\pi_X(x_i) + \alpha\bar{\pi}_X(x_r))}{(1+\alpha)\sum\limits_{i=1}^{p} \rho_{ij}u_{ij}^{m}} \tag{25c}$$

Similarly, calculate the partial derivative of L with respect to u_{ij} and ζ_i and equating them to zero, we have

$$\underset{\substack{1\le i\le p\\ 1\le j\le c}}{\forall}\ \frac{\partial L}{\partial u_{ij}} = 0 \text{ and } \underset{1\le i\le p}{\forall}\ \frac{\partial L}{\partial \zeta_i} = 0 \tag{26}$$

After simplifying (26), we get

$$u_{ij} = \left\{\sum_{k=1}^{c}\left(\frac{\tilde{d}_2^2(X_i^{IFS}, v_j^{IFS}) + \alpha\tilde{d}_2^2(\bar{X}_r^{IFS}, v_j^{IFS})}{\tilde{d}_2^2(X_i^{IFS}, v_k^{IFS}) + \alpha\tilde{d}_2^2(\bar{X}_r^{IFS}, v_k^{IFS})}\right)^{\frac{1}{m-1}}\right\}^{-1} \tag{27}$$

The final solution is obtained using (25) and (27) with the help of an alternating optimization algorithm which is given in Algorithm 6. The value of spatial regularization parameter $\alpha =$ 0 in (22) reduces to solution of the optimization problem (19)

4 Experimentation setup and results

To check the efficacy of the proposed PIFCM_S algorithm in comparison to other existing counterparts such as FCM [7], IFCM [57], FCM_S [1], FLICM [33], KFCM_S [13], ARKFCM [19], IIFCM [54], KIFCM [38], PIFCM [40], KWFLICM [21], NDFCM [22],

```
Input  : Dataset D with $p \times 1$ dimensions; fuzzy factor $(m)$; spatial regularization
         parameter $(\alpha)$; intuitionistic fuzzy parameter (β); number of centroids (c);
         tolerance level $(\epsilon)$
Output: Intuitionistic fuzzy partition $U$, centroid $v_j^{IFS}$.
1. Calculate $X_i^{IFS} = \langle \mu_X(x_i), \nu_X(x_i), \pi_X(x_i) \rangle \ \forall\ i \in \{1, 2, \ldots, p\}$ as described in
Section 2.1
2. Initialize intuitionistic fuzzy centroid $V^{IFS(k=0)}$ randomly.
3. $k \leftarrow 1$
Repeat
4. Calculate the updated fuzzy partition matrix
   $U^{(k)} = [u_{ij}^{(k)}]_{p \times c}$ using (27).
5. Calculate the Updated weight matrices P, Q and R
   using Algorithms 2 - 4 and normalize them.
6. Update the cluster centroids using ( 25 )
   $v_j^{IFS(k+1)} = \left\langle \mu_V(v_j)^{(k+1)}, \nu_V(v_j)^{(k+1)}, \pi_V(v_j)^{(k+1)} \right\rangle$ .
7. $k \leftarrow k + 1$
   Until
8. $\|v_j^{IFS(k+1)} - v_j^{IFS(k)}\| < \epsilon$
9. Return the membership degrees $U^{(k+1)}$ , the centroids $v_j^{IFS(k+1)}$
```

Algorithm 6 PIFCM_S algorithm.

WRFCM [55], FRFCM [35] and DSFCMN [60], experiments have been conducted on two publicly available brain MRI dataset. The PIFCM_S method performs clustering of the pixels of the image represented in terms of IFS for image segmentation. For this purpose, we have investigated two well-known intuitionistic fuzzy generation functions, namely Sugeno's and Yager's negation functions to convert the MRI images in IFS. Both the variants of proposed method are denoted as PIFCM_S(S) and PIFCM_S(Y) corresponding to Sugeno's negation function and Yager's negation function respectively for representing the image. The segmentation performance of both the variants of proposed PIFCM_S method is compared with the state-of-art methods in terms of average segmentation accuracy (ASA) and Dice score (DS). The mathematical definition of the performance measures indexes are summarized in Table 1. In this table, c is the number of clusters; X_i represents the pixels belonging to the manually segmented MRI image (ground truth) and Y_i represent the pixels belonging to the experimental segmented MRI image corresponding to i^{th} region; $\mod X_i$ represents the cardinality of X_i. The datasets used for experimentation are described in Section 4.1.

Table 1 List of performance measures

Performance measure	Formula
Average Segmentation Accuracy (ASA)	$\sum_{i=1}^{c} \frac{\lvert X_i \cap Y_i \rvert}{\sum_{j=1}^{c} \lvert X_j \rvert}$
Dice Score (DS)	$\frac{2\lvert X_i \cap Y_i \rvert}{\lvert X_i \rvert + \lvert Y_i \rvert}$

4.1 Datasets:

4.1.1 Brain MRI datasets :

Two publicly available real world dataset are also used for experimentation. The description about the brain MRI datasets is given as :

- **Simulated MRI brain volumes**: It is a publicly available dataset from the McConnell Brain Imaging Center of the Montreal Neurological Institute, McGill University [15]. The dataset contains simulated T1-weighted MRI images with different levels of noise ($1\%,\ 3\%,\ 5\%,\ 7\%$ and 9%) and intensity inhomogeneity or intensity non-uniformity (INU) ($0\%,\ 20\%$ and 40%) of resolution $1 \times 1 \times 1\text{mm}^3$ with $181 \times 217 \times 181$ dimension with ground truth.
- **Internet Brain Segmentation Repository (IBSR)**: It is a real MRI brain images that has been acquired from the Internet Brain Segmentation Repository (IBSR)[1] which has the ground truth data along with it. For all the MRI images, the brain extraction tool[2] is utilized for skull striping.

4.1.2 Tool used for experimental results

All the Experimental results are obtained using MATLAB version 9.6 running on a PC having 3.40 GHz frequency and 16 GB of RAM.

4.1.3 Parameter selection:

In this work, we have applied grid search for obtaining the optimal parameter values for all the methods along with the proposed PIFCM_S method based on the optimal value of objective function and the performance measures corresponding to the optimal parameter value is quoted. The proposed PIFCM_S algorithm involves mainly three parameters; fuzzifier factor m, spatial regularization factor α and intuitionistic fuzzy generator parameter β, which have significant impact on the solution of its optimization problem, i.e., cluster centroids and fuzzy partition matrix according to (25) and (27) thereby affecting the cluster performance measures. The optimal values of the parameters in the proposed PIFCM_S and other related methods have been obtained using the grid search method [24]. The parameter value is set based on the maximum average segmentation accuracy obtained. The range of Yager's negation parameter and Sugeno's negation parameter is searched in the interval [0, 2] and [0, 5], respectively, with 0.05 step-size. The optimal value of spatial regularization factor α is chosen in the interval [0, 5] with 0.1 step-size depending on the noise level present in the MRI image. The fuzzifier factor m and tolerance criterion ϵ are set to 2 and 10^{-5}, respectively.

4.2 Results and discussion on BrainWeb datasets

In this section, a detailed discussion and comparison of the performance of the proposed methods, namely, PIFCM_S(S) and PIFCM_S(Y) is presented with other state of art methods in terms of aforementioned performance measure indexes (see Table 1) on BrainWeb

[1] IBSR [online], available: https://www.nitrc.org/projects/ibsr

[2] Brain Extraction Tool (BET) [online], available: http://www.fmrib.ox.ac.uk/fsl/.

simulated MRI datasets. Figure 1 represents the Original image (INU = 40% and noise level = 9%) and ground truth corresponding to WM, GM and CSF. Figure 2 represents the qualitative segmentation results obtained for WM, GM and CSF using the proposed method and the state-of-the-art methods on this image. From Fig. 2, It can be noted that the qualitative segmentation results obtained using the proposed methods, namely, PIFCM_S(S) and PIFCM_S(Y) better in comparison to the state of art methods. Figure 3(a)-(f) depicts the bar chart of variation of average segmentation accuracy with different levels of INU for a given level of noise. Figure 4(a)-(c) shows the line graph of the variation of average segmentation accuracy with different levels of noise for a given level of INU. Table 2 shows the performance in terms of average segmentation accuracy on brainweb simulated MRI datasets for high levels of noise (7 % and 9%) with different levels of INU (0 %, 20 % and 40 %). From Figs. 3(a)-(f), 4 (a)-(c) and Table 2, the observation drawn is discussed as follows:

1. The performance of the proposed method is better than other state of the art methods for a given noise level.
2. For a given level of noise, the performance of the proposed method is steady for different levels of INU over state of the art methods where the performance is debased substantially. Although FCM, FCM_S, IFCM and IIFCM methods perform well on INU (40 %) images with low noise (0 %, 1 %, 3 % and 5 %) compared to the proposed method but lag behind on high level of noises (5 % and 7 %).
3. As the level of the noise increases (see Fig. 3(a)-(f)), the performance of all the methods debased as expected, but it is less in case of our proposed method in comparison to other related methods.
4. Figure 3(f) clearly depicts that the proposed method gives better segmentation accuracy compared to other methods such as ARKFCM, KFCM_S and KIFCM to handle both noise and INU.
5. For a given level of INU, the average segmentation accuracy is always going to be debased as the level of the noise increases. But this debasement of the segmentation performance in the proposed method is less in comparison to other methods. This shows that the proposed method is robust towards noise due to successful exploitation of spatial constraint.
6. For a given level of INU, the performance of all the methods debased as the level of noise increases from 0 % to 9 % (See Fig. 4(b)-(c)). However, the debasement in the performance of the proposed methods is less in comparison to other methods that shows its robustness towards the INU.

Further, to show the effectiveness of the proposed method over the state-of-art methods for tissue segmentation evaluation, the Dice score for GM and WM is summarized in Tables 3 and 4, respectively. The high values of the DS for both GM and WM tissue evidence the correct identification of the regions in an image using the proposed PIFCM_S method in presence of both noise and INU. From Tables 3 and 4, it is clear that the state-of-art methods are unable to provide comparable results in terms of DS for GM and WM corresponding to the proposed PIFCM_S method except for the FLICM and KFCM_S methods. It is also observed from Figs. 3 and 4 that for low levels of noises (0 %, 1 % and 3 %), the proposed PIFCM_S method gives better performance in terms of ASA when the image is represented in IFS using Yager's negation function. Whereas, for higher levels of noises (5 %, 7 % and 9 %), the performance of the proposed method is better when the image is represented in IFS using Sugeno's negation function. This shows the effectiveness of both Yager's negation

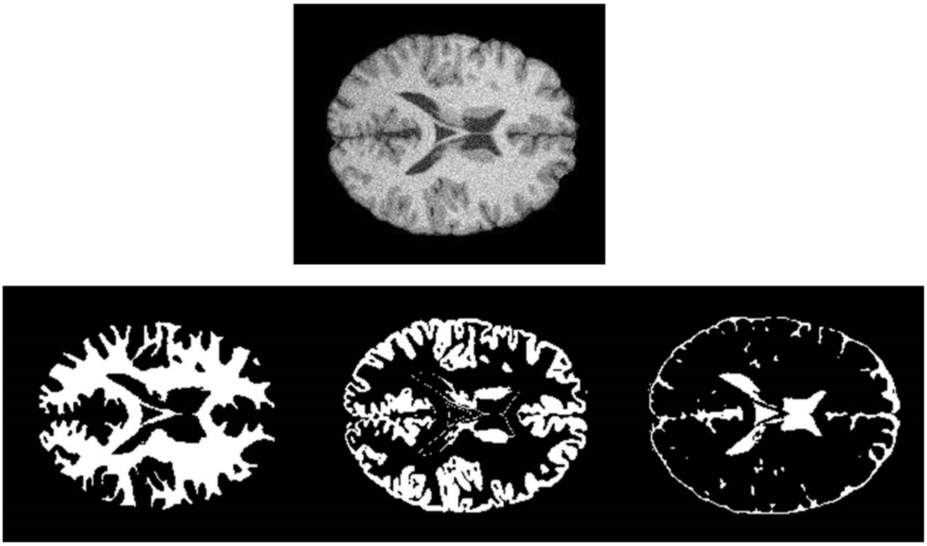
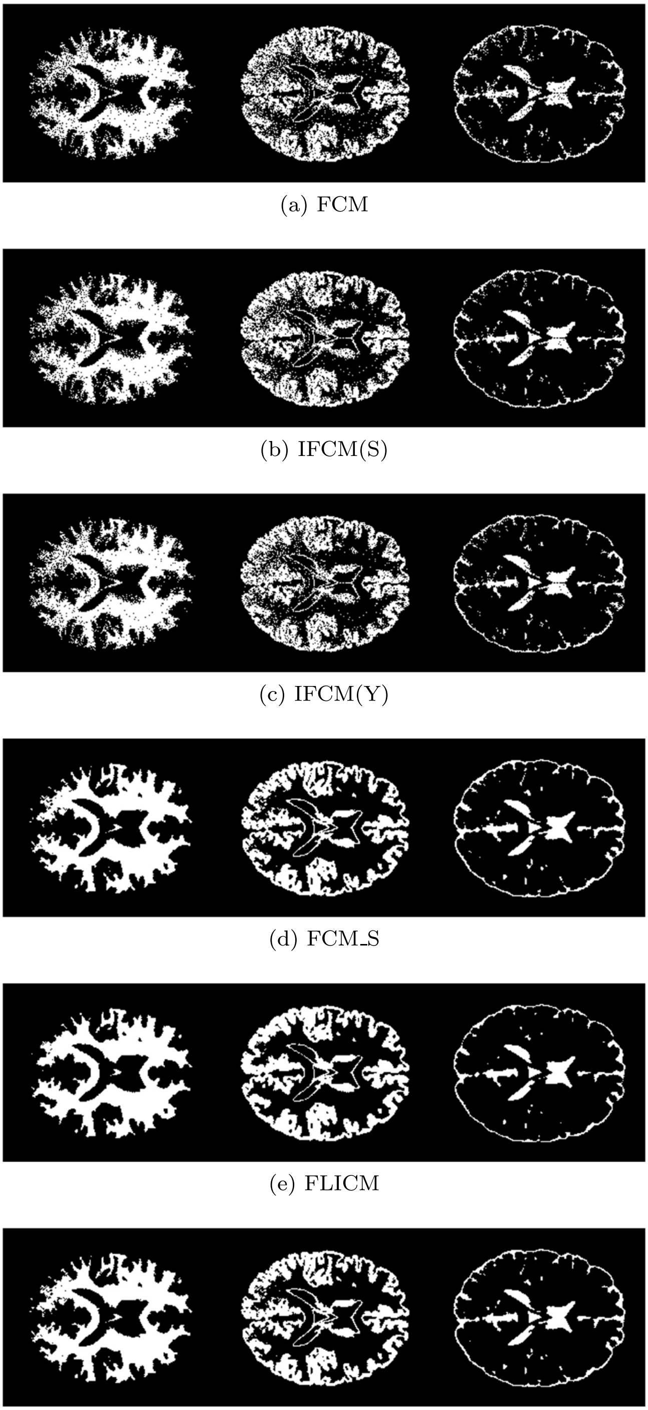


Fig. 1 Original image (INU = 40% and noise level = 9%) and ground truth corresponding to WM, GM and CSF

function and Sugeno's negation function over different levels of noise on Brainweb MRI dataset.

4.3 Results discussion on real brain MRI dataset

The effectiveness of the proposed PIFCM_S method with other state-of-art methods is further checked on real normal brain MR images from IBSR database for which ground truth is available. For this, the 134th axial slice of T1-weighted image is extracted from IBSR dataset for 8 cases 110_3, 111_2, 11_3, 12_3, 15_3, 16_3, 1_24 and 205_3 and corrupted with 10 % Rician noise to test the performance of the segmentation methods in noisy environment. Table 5 shows the performance of the proposed PIFCM_S method along with state-of-art methods in terms of ASA. From Table 5, it can be clearly seen that the proposed method on real brain MRI images corrupted with 10 % Rician noise outperforms the other related methods. Whereas, the performance of the existing methods for high noise images could not provide satisfactory performance. The utilization of the spatial constraints in the proposed PIFCM_S method provides resistance to noise for real brain MRI images in the IFS framework. Table 6 shows the tissue segmentation performance measure in terms of Dice Score (DS) corresponding to GM on these images. It can be noted from these tables that the proposed PIFCM_S method performs well except on the images 11_3 and 15_3 in comparison to other methods in terms of DS. However, the average value of the proposed PIFCM_S method is higher than other state-of-arts methods (see Fig. 5). Figure 5 shows the average value of ASA and DS for GM over 8 cases of real brain MRI images with 10 % Rician noise. It is evident from Fig. 5 that the proposed PIFCM_S method with Sugeno's negation function performs well on average over all 8 cases of real brain MRI image in terms of ASA and DS (GM). It reveals the performance of the proposed PIFCM_S method on these images has significant improvement in terms of the performance measures used while comparing with other state-of-art methods.



(a) FCM

(b) IFCM(S)

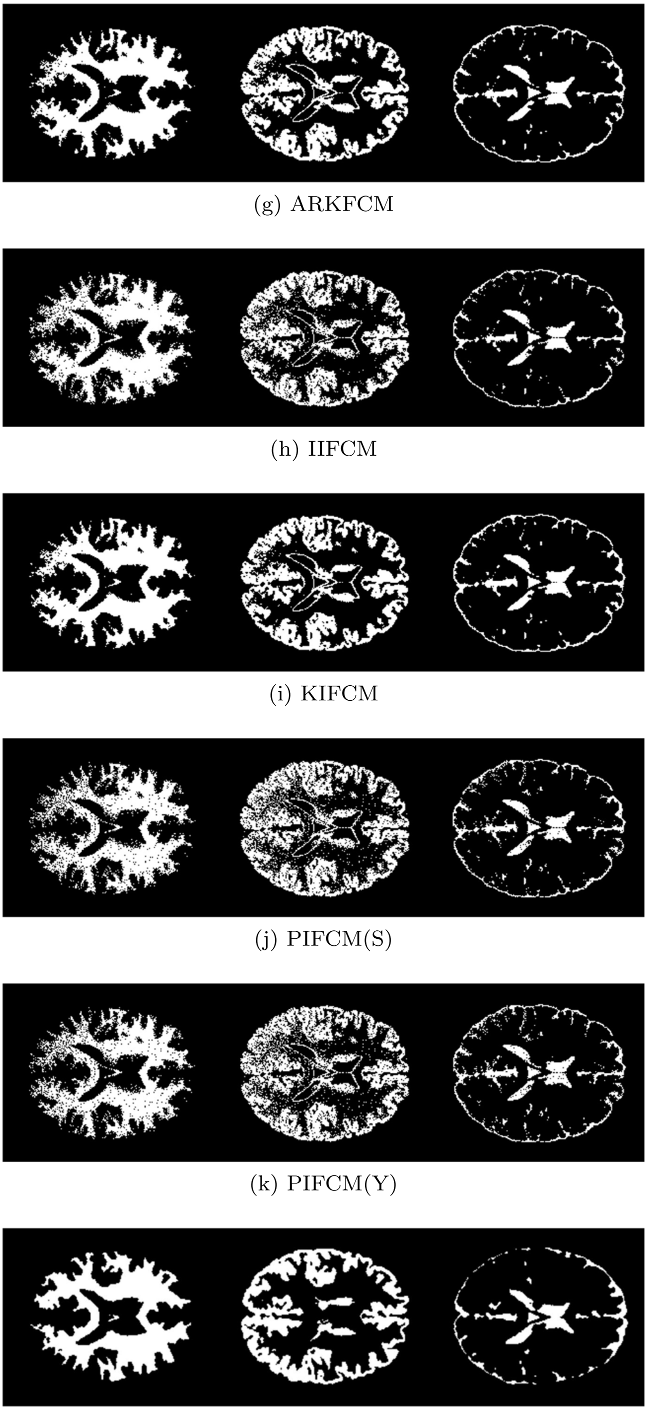
(c) IFCM(Y)

(d) FCM_S

(e) FLICM

(f) KFCM_S

Fig. 2 Qualitative results for WM, GM and CSF for different methods (Cont..)



(g) ARKFCM

(h) IIFCM

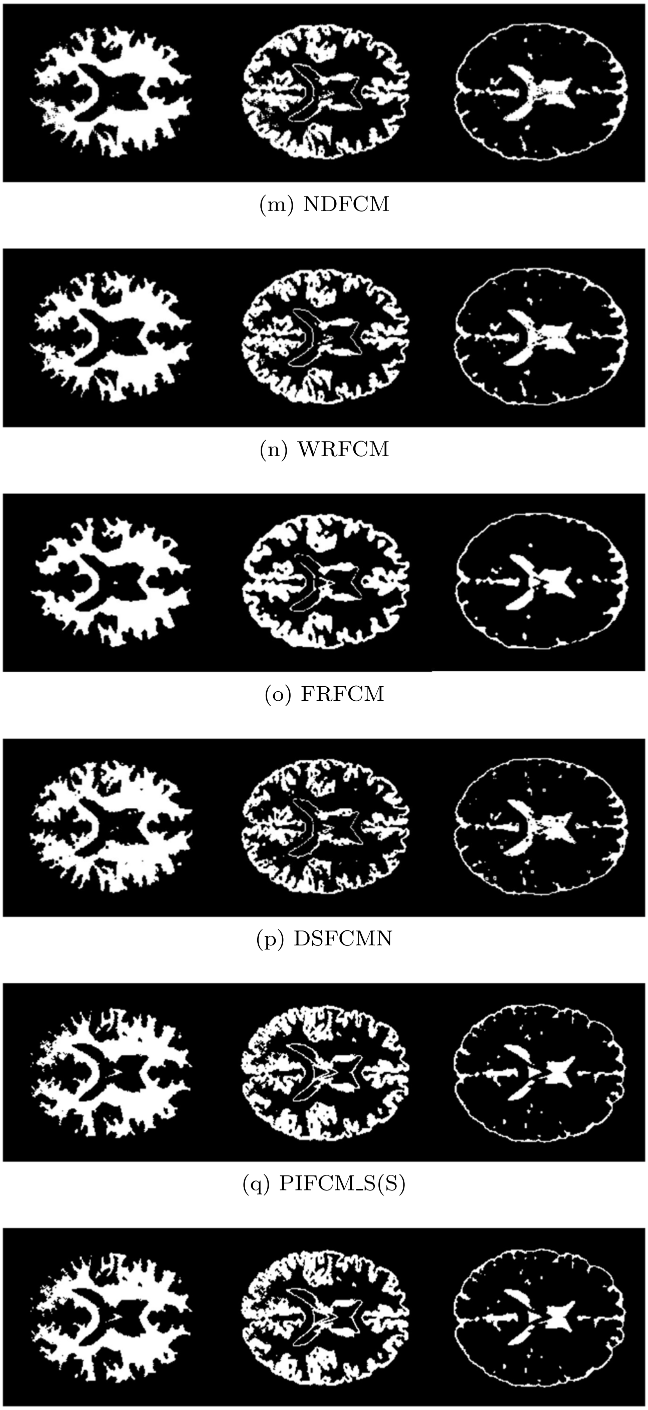
(i) KIFCM

(j) PIFCM(S)

(k) PIFCM(Y)

(l) KWFLICM

Fig. 2 (continued)



(m) NDFCM

(n) WRFCM

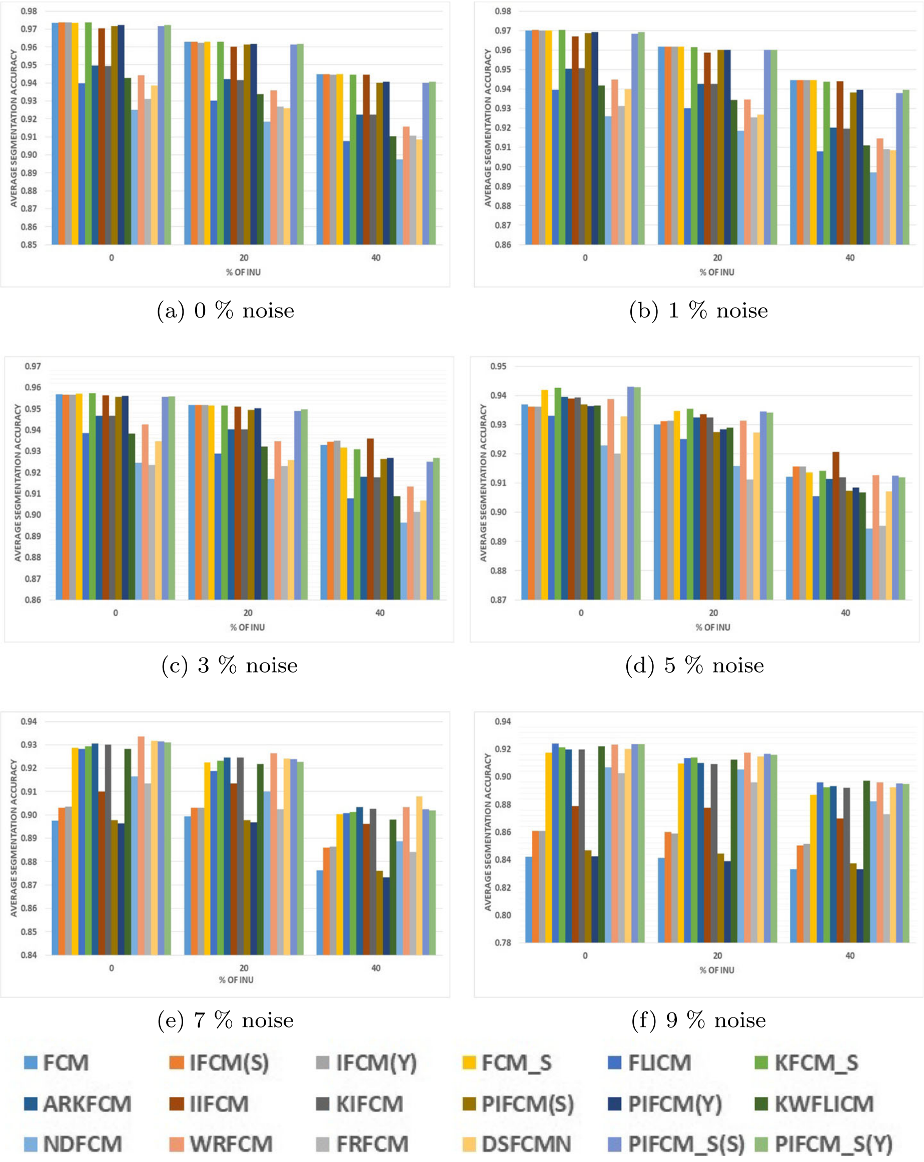
(o) FRFCM

(p) DSFCMN

(q) PIFCM_S(S)

(r) PIFCM_S(Y)

Fig. 2 (continued)



(a) 0 % noise

(b) 1 % noise

(c) 3 % noise

(d) 5 % noise

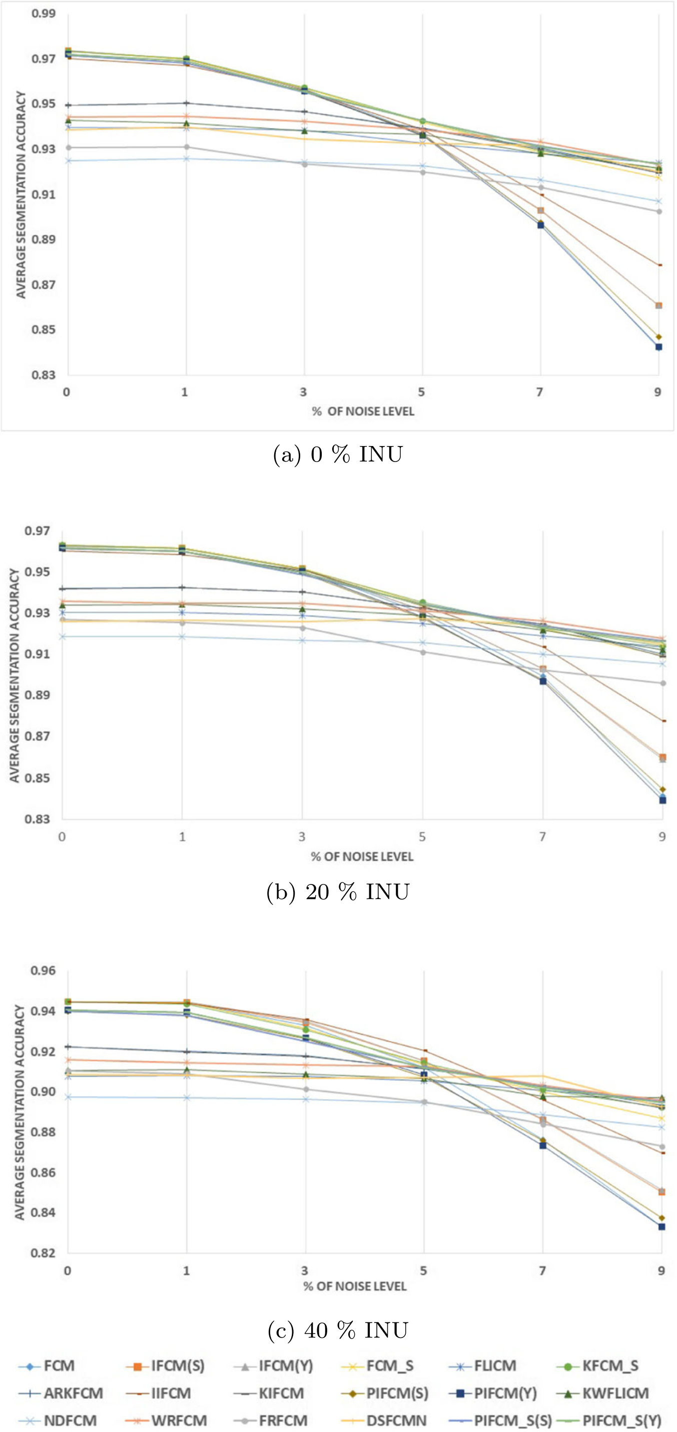
(e) 7 % noise

(f) 9 % noise

Fig. 3 Variation in performance in terms of average segmentation accuracy with INU level of for given level of noise on Brain Web dataset a) 0 % b)1 % c) 3 % d) 5% e) 7% and f) 9%

4.4 Statistical test

Friedman test, a two way non-parametric 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



(a) 0 % INU

(b) 20 % INU

(c) 40 % INU

Fig. 4 Variation in performance in terms of average segmentation accuracy with noise level of for given level of INU on Brain Web dataset a) 0 % b) 20 % c) 40 %

Table 2 Comparison of PIFCM_S with other methods in terms of ASA for Brain Web dataset

Image/Methods	7% Noise			9% Noise		
	0% INU	20% INU	40% INU	0% INU	20% INU	40% INU
FCM	0.8976	0.8994	0.8762	0.8421	0.8414	0.8330
IFCM(S)	0.9031	0.9030	0.8861	0.8608	0.8602	0.8505
IFCM(Y)	0.9035	0.9030	0.8864	0.8609	0.8590	0.8513
FCM_S	0.9288	0.9225	0.9002	0.9175	0.9094	0.8869
FLICM	0.9283	0.9188	0.9007	0.9242	0.9137	0.8959
KFCM_S	0.9294	0.9231	0.9013	0.9214	0.9139	0.8926
ARKFCM	0.9306	0.9244	0.9032	0.9198	0.9101	0.8934
IIFCM	0.9100	0.9135	0.8961	0.8788	0.8777	0.8699
KIFCM	0.9301	0.9245	0.9026	0.9196	0.9090	0.8922
PIFCM(S)	0.8978	0.8977	0.8760	0.8469	0.8444	0.8376
PIFCM(Y)	0.8963	0.8969	0.8732	0.8426	0.8390	0.8331
KWFLICM	0.9283	0.9217	0.8979	0.9220	0.9123	0.8973
NDFCM	0.9165	0.9099	0.8887	0.9070	0.9053	0.8824
WRFCM	0.9335	0.9264	0.9032	0.9232	0.9175	0.8959
FRFCM	0.9134	0.9024	0.8841	0.9027	0.8960	0.8730
DSFCMN	0.9318	0.9240	0.9079	0.9202	0.9148	0.8926
PIFCM_S(S)	0.9314	0.9238	0.9025	0.9238	0.9166	0.8951
PIFCM_S(Y)	0.9311	0.9228	0.9020	0.9238	0.9160	0.8946

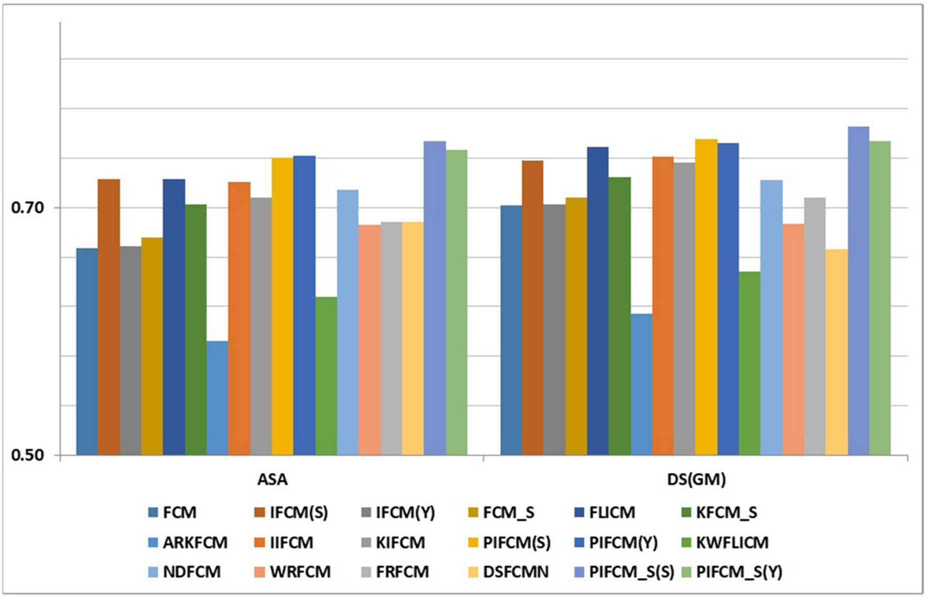


Fig. 5 Average value of ASA and DS (GM) over 8 cases of the IBSR dataset with 10% Rician noise

Table 3 Comparison of PIFCM_S with other methods in terms of DS for GM for Brain Web dataset

Image/Methods	7% Noise			9% Noise		
	0% INU	20% INU	40% INU	0% INU	20% INU	40% INU
FCM	0.8619	0.8669	0.8405	0.7950	0.7958	0.7864
IFCM(S)	0.8696	0.8722	0.8542	0.8193	0.8197	0.8093
IFCM(Y)	0.8696	0.8718	0.8537	0.8183	0.8168	0.8090
FCM_S	0.9017	0.8959	0.8695	0.8881	0.8787	0.8506
FLICM	0.9033	0.8928	0.8728	0.8994	0.8872	0.8658
KFCM_S	0.9034	0.8968	0.8712	0.8941	0.8853	0.8588
ARKFCM	0.9055	0.8992	0.8749	0.8926	0.8814	0.8615
IIFCM	0.8778	0.8850	0.8660	0.8397	0.8397	0.8311
KIFCM	0.9046	0.8990	0.8739	0.8921	0.8797	0.8595
PIFCM(S)	0.8653	0.8679	0.8444	0.8069	0.8053	0.7979
PIFCM(Y)	0.8624	0.8659	0.8396	0.7993	0.7964	0.7903
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
WRFCM	0.9127	0.9043	0.8744	0.8989	0.8934	0.8650
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
PIFCM_S(S)	0.9077	0.8995	0.8750	0.8994	0.8909	0.8653
PIFCM_S(Y)	0.9068	0.8974	0.8734	0.8987	0.8895	0.8637

are different. For a given performance measure M, the H_0 and H_1 can be defined as:

$$\begin{aligned} H_0 \;:\; & \mu_{FCM} = \mu_{IFCM(S)} = \mu_{IFCM(Y)} = \mu_{FCM\_S} \\ & = \mu_{FLICM} = \mu_{KFCM\_S} = \mu_{ARKFCM} = \mu_{IIFCM} \\ & = \mu_{KIFCM} = \mu_{PIFCM(S)} = \mu_{PIFCM(Y)} \\ & = \mu_{KWFLICM} = \mu_{NDFCM} = \mu_{WRFCM} \\ & = \mu_{FRFCM} = \mu_{DSFCMN} = \mu_{PIFCM\_S(S)} = \mu_{PIFCM\_S(Y)} \end{aligned} \tag{28}$$

$$\begin{aligned} H_1 \;:\; & \mu_{FCM} \neq \mu_{IFCM(S)} \neq \mu_{IFCM(Y)} \neq \mu_{FCM\_S} \\ & \neq \mu_{FLICM} \neq \mu_{KFCM\_S} \neq \mu_{ARKFCM} \neq \mu_{IIFCM} \\ & \neq \mu_{KIFCM} \neq \mu_{PIFCM(S)} \neq \mu_{PIFCM(Y)} \\ & \neq \mu_{KWFLICM} \neq \mu_{NDFCM} \neq \mu_{WRFCM} \\ & \neq \mu_{FRFCM} \neq \mu_{DSFCMN} \neq \mu_{PIFCM\_S(S)} \neq \mu_{PIFCM\_S(Y)} \end{aligned} \tag{29}$$

The rank of different segmentation methods, according to the different performance measures is obtained for comparing the methods separately. In 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:

$$R_j = \frac{1}{N}\sum_{i=1}^{N} r_i^j \tag{30}$$

Table 4 Comparison of PIFCM_S with other methods in terms of DS for WM for Brain Web dataset

Image/Methods	7% Noise			9% Noise		
	0% INU	20% INU	40% INU	0% INU	20% INU	40% INU
FCM	0.9300	0.9308	0.9105	0.8860	0.8860	0.8766
IFCM(S)	0.9321	0.9318	0.9132	0.8942	0.8941	0.8832
IFCM(Y)	0.9326	0.9323	0.9144	0.8954	0.8943	0.8850
FCM_S	0.9604	0.9543	0.9343	0.9552	0.9482	0.9292
FLICM	0.9596	0.9504	0.9329	0.9569	0.9470	0.9293
KFCM_S	0.9615	0.9550	0.9353	0.9556	0.9499	0.9302
ARKFCM	0.9605	0.9544	0.9342	0.9512	0.9431	0.9257
IIFCM	0.9390	0.9407	0.9221	0.9111	0.9090	0.9002
KIFCM	0.9606	0.9544	0.9344	0.9513	0.9430	0.9259
PIFCM(S)	0.9258	0.9252	0.9034	0.8799	0.8780	0.8700
PIFCM(Y)	0.9261	0.9264	0.9042	0.8805	0.8780	0.8706
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
WRFCM	0.9575	0.9522	0.9311	0.9511	0.9472	0.9273
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
PIFCM_S(S)	0.9618	0.9540	0.9326	0.9565	0.9494	0.9280
PIFCM_S(Y)	0.9628	0.9541	0.9347	0.9569	0.9500	0.9292

Table 5 Comparison of PIFCM_S with other methods in terms of ASA for IBSR dataset with Rician noise (σ =10)

Methods\Images	110_3	111_2	11_3	12_3	15_3	16_3	1_24	205_3
FCM	0.7293	0.6946	0.7214	0.7352	0.5090	0.5474	0.6864	0.7150
IFCM(S)	0.7403	0.7424	0.7270	0.7522	0.6685	0.6859	0.7460	0.7183
IFCM(Y)	0.7309	0.6955	0.7226	0.7371	0.5090	0.5506	0.6899	0.7145
FCM_S	0.7321	0.7007	0.7306	0.7407	0.5160	0.5737	0.6929	0.7209
FLICM	0.7406	0.7245	0.7695	0.7721	0.5875	0.6826	0.7506	0.7558
KFCM_S	0.7344	0.7219	0.7307	0.7492	0.5820	0.6826	0.6929	0.7269
ARKFCM	0.6139	0.6012	0.7162	0.7413	0.5818	0.6806	0.7652	0.6200
IIFCM	0.7400	0.7470	0.7479	0.7601	0.5915	0.6957	0.7528	0.7302
KIFCM	0.7472	0.7273	0.7406	0.7594	0.5049	0.6949	0.7587	0.7274
PIFCM(S)	0.7642	0.7422	0.7574	0.7605	0.6702	0.7263	0.7508	0.7515
PIFCM(Y)	0.7675	0.7454	0.7594	0.7596	0.6459	0.7407	0.7540	0.7588
KWFLICM	0.5148	0.6790	0.7125	0.7239	0.5698	0.5910	0.6893	0.5448
NDFCM	0.7521	0.7196	0.7389	0.7290	0.6670	0.6998	0.6945	0.7129
WRFCM	0.6919	0.6723	0.6935	0.7106	0.6817	0.6735	0.6714	0.6902
FRFCM	0.6695	0.6762	0.7156	0.7089	0.6742	0.6858	0.6813	0.6924
DSFCMN	0.7397	0.7062	0.6195	0.7119	0.6361	0.6633	0.6799	0.7479
PIFCM_S(S)	0.7709	0.7556	0.7721	0.7759	0.6750	0.7441	0.7670	0.7688
PIFCM_S(Y)	0.7705	0.7499	0.7682	0.7665	0.6446	0.7473	0.7636	0.7622

Table 6 Comparison of PIFCM_S with other methods in terms of DS for GM for IBSR dataset with Rician noise (σ =10)

Methods\Images	110_3	111_2	11_3	12_3	15_3	16_3	1_24	205_3
FCM	0.7667	0.7193	0.7635	0.7535	0.5499	0.6073	0.6880	0.7646
IFCM(S)	0.7623	0.7593	0.7509	0.7665	0.6655	0.6946	0.7432	0.7575
IFCM(Y)	0.7674	0.7200	0.7641	0.7552	0.5491	0.6100	0.6919	0.7637
FCM_S	0.7675	0.7245	0.7720	0.7589	0.5531	0.6244	0.6938	0.7699
FLICM	0.7672	0.7437	0.8028	0.7896	0.6172	0.7168	0.7552	0.7993
KFCM_S	0.7667	0.7387	0.7698	0.7647	0.5936	0.6951	0.6946	0.7741
ARKFCM	0.6638	0.6107	0.7619	0.7567	0.5926	0.6936	0.7556	0.6764
IIFCM	0.7586	0.7589	0.7697	0.7734	0.6293	0.7215	0.7458	0.7696
KIFCM	0.7783	0.7468	0.7857	0.7789	0.5502	0.7126	0.7595	0.7758
PIFCM(S)	0.7929	0.7696	0.7720	0.7821	0.6764	0.7007	0.7567	0.7910
PIFCM(Y)	0.7950	0.7727	0.7704	0.7817	0.6157	0.7234	0.7610	0.7951
KWFLICM	0.5385	0.7062	0.7399	0.7338	0.6017	0.5965	0.7182	0.5514
NDFCM	0.7694	0.7442	0.7705	0.7050	0.6642	0.7050	0.7199	0.7000
WRFCM	0.7006	0.6883	0.7194	0.7114	0.6603	0.6589	0.6911	0.6656
FRFCM	0.6909	0.7079	0.7511	0.7302	0.6867	0.6953	0.7134	0.6852
DSFCMN	0.7479	0.6935	0.6135	0.6844	0.5485	0.6367	0.6791	0.7288
PIFCM_S(S)	0.7946	0.7787	0.7822	0.7964	0.6644	0.7273	0.7742	0.8061
PIFCM_S(Y)	0.7972	0.7759	0.7813	0.7878	0.6048	0.7127	0.7708	0.7997

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 7 shows the average Friedman ranking of different segmentation methods corresponding to ASA for 9 BrainWeb brain images and 8 synthetic images used for experiment [17, 20]. 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 PIFCM_S(S) performs better in terms of *ASA*. The statistical hypothesis test proposed by Iman and Davenportis is used. The statistic F_{ID} is defined by Iman and Davenport [30] is given as:

$$F_{ID} = \frac{(N-1)\chi_F^2}{N(k-1) - \chi_F^2} \tag{31}$$

which is distributed according to F-distribution with $k - 1$ and $(k - 1)(N - 1)$ degrees of freedom, where χ_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 $k = 18$ and $N = 17$. The p-value obtained by Iman and Davenport statistic is 0.0 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 obtained are not suitable for comparison with the control method, i.e. the one that emerges with the lowest rank. So adjusted p-values [17] 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 PIFCM_S(S) (lowest rank for *ASA*). For this, a set of post-hoc procedures are defined and adjusted

Table 7 Average Friedman Rankings of the algorithms

Algorithm	Ranking	Algorithm	Ranking
PIFCM_S(S)	2.50	DSFCMN	10.09
PIFCM_S(Y)	3.94	PIFCM(S)	10.21
KIFCM	7.35	IFCM(S)	10.44
KFCM_S	7.50	PIFCM(Y)	10.47
FLICM	8.03	NDFCM	11.29
IIFCM	8.18	KWFLICM	11.97
WRFCM	8.74	IFCM(Y)	12.74
FCM_S	9.50	FRFCM	14.06
ARKFCM	9.59	FCM	14.41

p-values are computed. The most widely used post-hoc method [17] to obtain adjusted *p*-values is Holm procedure. Table 8 shows the various values of adjusted *p*-values obtained. Table 8 indicate that the performance of proposed PIFCM_S method with Sugeno's negation function and Yager's negation function in terms of ASA performance measures have no significant difference.

5 Conclusion

In this research work, we have presented a intuitionistic fuzzy set theoretic clustering for image segmentation problem that uses probabilistic Euclidean distance measure with a spatial regularization term (PIFCM_S). For this, we have utilized the mean filter image in the spatial regularization term in the segmentation process to dampen the effect of noise. The optimization problem of the proposed approach has the advantage of probabilistic Euclidean distance measure and regularization term to handle the noise in IFS framework. The image representation in terms of IFS increases the representational capability and hence improves segmentation performance. For this, two well-known intuitionistic fuzzy negation functions, namely Yager's negation function and Sugeno's negation function have been utilized to convert the gray image in terms of IFS. The experiments are carried out on two publicly available brain MRI dataset for checking the efficacy of the proposed method. Moreover,

Table 8 Adjusted p-values (Friedman) corresponding to performance measure ASA

Algorithm	Unadjusted p value	p_{Holm} value	Algorithm	Unadjusted p value	p_{Holm} value
FCM	7.76E-11	1.32E-09	ARKFCM	1.08E-04	8.67E-04
FRFCM	2.75E-10	4.39E-09	FCM_S	1.32E-04	9.23E-04
IFCM(Y)	2.27E-08	3.41E-07	WRFCM	6.61E-04	0.004
KWFLICM	2.32E-07	3.24E-06	IIFCM	0.002	0.010
NDFCM	1.57E-06	2.04E-05	FLICM	0.003	0.010
PIFCM(Y)	1.34E-05	1.61E-04	KFCM_S	0.006	0.019
IFCM(S)	1.45E-05	1.61E-04	KIFCM	0.008	0.019
PIFCM(S)	2.57E-05	2.57E-04	PIFCM_S(Y)	0.431	0.431
DSFCMN	3.41E-05	3.07E-04			

the comparison of the performance of the proposed PIFCM_S method with other state-of-art methods is carried out on the datasets. The results obtained on these two publicly available datasets show significant improvement in the segmentation performance of the proposed PIFCM_S method in comparison to other related methods in terms of average segmentation accuracy and Dice score. It is clearly depicted from the results that Sugeno's negation function gives better performance for higher level of noise whereas Yager's negation function gives better performance for lower level of noise. Further, a statistical test has been performed to check the significant difference in the performance of the proposed PIFCM_S method with the state-of-art methods. The statistical test shows that the performance of the proposed PIFCM _S method is superior over other related methods. The limitation of the proposed PIFCM_S method is the manual tuning of the intuitionistic negation parameter and spatial regularization parameter, which is important to obtain the accurate segmentation. In the future direction, we may investigate an adaptive way to choose the optimal value of these parameters based on the image itself.

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.

Declarations

Ethics 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.

Conflict of Interests The authors declare that they have no conflict of interest.

References

1. 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 Imaging 21(3):193–199
2. Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ (2017) Deep learning for brain mri segmentation: state of the art and future directions. J Digital Imaging 30(4):449–459
3. Atanassov KT (1986) Intuitionistic fuzzy sets. Fuzzy Sets and Systems 20(1):87–96. https://doi.org/10.1016/S0165-0114(86)80034-3, http://www.sciencedirect.com/science/article/pii/S0165011486800343
4. Balafar M, Ramli AR, Saripan MI, Mahmud R, Mashohor S, Balafar M (2008) New multi-scale medical image segmentation based on fuzzy c-mean (fcm). In: 2008 IEEE Conference on innovative technologies in intelligent systems and industrial applications. IEEE, pp 66–70
5. Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Kluwer Academic Publishers
6. Bezdek JC (2013) Pattern recognition with fuzzy objective function algorithms. Springer Science & Business Media
7. Bezdek JC, Douglas Harris J (1978) Fuzzy partitions and relations; an axiomatic basis for clustering. Fuzzy Sets Syst 1(2):111–127
8. Bezdek JC, Hall L, Clarke L (1993) Review of mr image segmentation techniques using pattern recognition. Med Phys 20(4):1033–1048
9. Brandt ME, Bohant TP, Kramer LA, Fletcher JM (1994) Estimation of csf, white and gray matter volumes in hydrocephalic children using fuzzy clustering of mr images. Comput Med Imaging Graph 18(1):25–34

10. Cabezas M, Oliver A, Lladó X, Freixenet J, Cuadra MB (2011) A review of atlas-based segmentation for magnetic resonance brain images. Computer Methods and Programs in Biomedicine 104(3):e158–e177
11. Cai W, Chen S, Zhang D (2007) Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. Pattern Recogn 40(3):825–838
12. Chaira T (2011) A novel intuitionistic fuzzy c means clustering algorithm and its application to medical images. Appl Soft Comput 11(2):1711–1717
13. Chen S, Zhang D (2004) Robust image segmentation using fcm with spatial constraints based on new kernel-induced distance measure. IEEE Trans Syst Man Cybern Part B (Cybernetics) 34(4):1907–1916
14. Chintalapudi KK, Kam M (1998) A noise-resistant fuzzy c means algorithm for clustering. In: 1998 IEEE International conference on fuzzy systems proceedings. IEEE world congress on computational intelligence (cat. no. 98CH36228), vol 2. IEEE, pp 1458–1463
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
16. Cocosco CA, Zijdenbos AP, Evans AC (2003) A fully automatic and robust brain mri tissue classification method. Med Image Anal 7(4):513–527
17. Derrac J, García S, Malian 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
18. Dubey YK, Mushrif MM, Mitra K (2016) Segmentation of brain mr images using rough set based intuitionistic fuzzy clustering. Biocybern Biomed Eng 36(2):413–426
19. Elazab A, Wang C, Jia F, Wu J, Li G, Hu Q (2015) Segmentation of brain tissues from magnetic resonance images using adaptively regularized kernel-based fuzzy-means clustering. Computational and Mathematical Methods in Medicine, pp 2015
20. 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
21. 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
22. 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
23. Hall LO, Bensaid AM, Clarke LP, Velthuizen RP, Silbiger MS, Bezdek JC (1992) A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. IEEE Trans Neural Netw 3(5):672–682
24. Hsu CW, Lin CJ (2002) A comparison of methods for multiclass support vector machines. IEEE Trans Neural Netw 13(2):415–425
25. 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
26. Hylton N (2006) Mr imaging for assessment of breast cancer response to neoadjuvant chemotherapy. Magn Reson Imaging Clin 14(3):383–389
27. Iakovidis DK, Pelekis N, Kotsifakos E, Kopanakis I (2008) Intuitionistic fuzzy clustering with applications in computer vision. In: International conference on advanced concepts for intelligent vision systems. Springer, pp 764–774
28. Iancu I (2014) Intuitionistic fuzzy similarity measures based on frank t-norms family. Pattern Recogn Lett 42:128–136
29. Iglesias JE, Sabuncu MR (2015) Multi-atlas segmentation of biomedical images: a survey. Med Image Anal 24(1):205–219
30. Iman RL, Davenport JM (1980) Approximations of the critical region of the fbietkan statistic. Commun Stat Theory Methods 9(6):571–595
31. Ji Z, Sun Q, Xia Y, Chen Q, Xia D, Feng D (2012) Generalized rough fuzzy c-means algorithm for brain mr image segmentation. Comput Methods Programs Biomed 108(2):644–655
32. Kaya IE, Pehlivanlı AÇ, Sekizkardeş EG, Ibrikci T (2017) Pca based clustering for brain tumor segmentation of t1w mri images. Comput Methods Programs Biomed 140:19–28
33. Krinidis S, Chatzis V (2010) A robust fuzzy local information c-means clustering algorithm. IEEE Trans Image Process 19(5):1328–1337
34. Kumar D, Agrawal RK, Kirar JS (2019) Intuitionistic fuzzy clustering method with spatial information for mri image segmentation. In: 2019 IEEE International conference on fuzzy systems (FUZZ-IEEE). IEEE, pp 1–7
35. 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

36. Liew AC, Yan H, Law NF (2005) Image segmentation based on adaptive cluster prototype estimation. IEEE Trans Fuzzy Syst 13(4):444–453
37. Liew AWC, Yan H (2003) An adaptive spatial fuzzy clustering algorithm for 3-d mr image segmentation. IEEE Trans Med Imaging 22(9):1063–1075
38. Lin KP (2014) A novel evolutionary kernel intuitionistic fuzzy-means clustering algorithm. IEEE Trans Fuzzy Syst 22(5):1074–1087
39. Lloyd S (1982) Least squares quantization in pcm. IEEE Trans Inform Theory 28(2):129–137
40. Lohani QD, Solanki R, Muhuri PK (2018) Novel adaptive clustering algorithms based on a probabilistic similarity measure over atanassov intuitionistic fuzzy set. IEEE Trans Fuzzy Syst 26(6):3715–3729
41. Ma L, Staunton RC (2007) A modified fuzzy c-means image segmentation algorithm for use with uneven illumination patterns. Pattern Recogn 40(11):3005–3011
42. Moeskops P, Viergever MA, Mendrik AM, De Vries LS, Benders MJ, Išgum I. (2016) Automatic segmentation of mr brain images with a convolutional neural network. IEEE Trans Med Imaging 35(5):1252–1261
43. Murofushi T, Sugeno M et al (2000) Fuzzy measures and fuzzy integrals. Fuzzy measures and integrals: theory and applications, pp 3–41
44. Pelekis N, Iakovidis DK, Kotsifakos EE, Kopanakis I (2008) Fuzzy clustering of intuitionistic fuzzy data. Int J Business Intell Data Mining 3(1):45–65
45. Pham DL (2001) Spatial models for fuzzy clustering. Comput Vis Image Under 84(2):285–297
46. Pham DL, Xu C, Prince JL (2000) Current methods in medical image segmentation. Annual Rev Biomed Eng 2(1):315–337
47. Selvaraj H, Selvi ST, Selvathi D, Gewali L (2007) Brain mri slices classification using least squares support vector machine. Int J Intell Comput Med Sci Image Process 1(1):21–33
48. Shen S, Sandham W, Granat M, Sterr A (2005) Mri fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization. IEEE Trans Inform Technol Biomed 9(3):459–467
49. Singh C, Bala A (2019) A local zernike moment-based unbiased nonlocal means fuzzy c-means algorithm for segmentation of brain magnetic resonance images. Expert Syst Appl 118:625–639
50. Szilagyi L, Benyo Z, Szilágyi SM, Adam H (2003) Mr brain image segmentation using an enhanced fuzzy c-means algorithm. In: Proceedings of the 25th annual international conference of the IEEE engineering in medicine and biology society (IEEE Cat. No. 03CH37439), vol 1. IEEE, pp 724–726
51. Szmidt E, Kacprzyk J (2000) Distances between intuitionistic fuzzy sets. Fuzzy Sets Syst 114(3): 505–518
52. Tolias YA, Panas SM (1998) Image segmentation by a fuzzy clustering algorithm using adaptive spatially constrained membership functions. IEEE Trans Syst Man Cybern Part A: Systems and Humans 28(3):359–369
53. Varshney AK, Lohani QD, Muhuri PK (2020) Improved probabilistic intuitionistic fuzzy c-means clustering algorithm: Improved pifcm. In: 2020 IEEE International conference on fuzzy systems (FUZZ-IEEE). IEEE, pp 1–6
54. 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
55. 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
56. Wang Z, Boesch R (2007) Color-and texture-based image segmentation for improved forest delineation. IEEE Trans Geosci Remote Sens 45(10):3055–3062
57. Xu Z, Wu J (2010) Intuitionistic fuzzy c-means clustering algorithms. J Syst Eng Electron 21(4):580–590
58. YAGER RR (1979) On the measure of fuzziness and negation part i: Membership in the unit interval. Int J Gen Syst 5(4):221–229. 10.1080/03081077908547452
59. Yager RR (1980) On the measure of fuzziness and negation. ii. lattices. Inf Control 44(3):236–260
60. 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
61. Zhao F, Jiao L, Liu H (2013) Kernel generalized fuzzy c-means clustering with spatial information for image segmentation. Digit Signal Process 23(1):184–199
62. Zhou H, Schaefer G, Shi C (2008) A mean shift based fuzzy c-means algorithm for image segmentation. In: 2008 30Th annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 3091–3094
63. Zhu L, Chung FL, Wang S (2009) Generalized fuzzy c-means clustering algorithm with improved fuzzy partitions. IEEE Trans Syst Man Cybern Part B (Cybernetics) 39(3):578–591

Received: 20 December 2022 | Revised: 5 February 2023 | Accepted: 1 March 2023

DOI: 10.1002/cnr2.1807

Cancer Reports WILEY

REVIEW

Racial disparities in cancer care, an eyeopener for developing better global cancer management strategies

Bharmjeet | Asmita Das

Department of Biotechnology, Delhi Technological University, Delhi, 110042, India

Correspondence
Asmita Das, Department of Biotechnology, Delhi Technological University, Main Bawana Road, Delhi-110042, India.
Email: asmitadas1710@dce.ac.in

Abstract

Background: In the last few decades, advancements in cancer research, both in the field of cancer diagnostics as well as treatment of the disease have been extensive and multidimensional. Increased availability of health care resources and growing awareness has resulted in the reduction of consumption of carcinogens such as tobacco; adopting various prophylactic measures; cancer testing on regular basis and improved targeted therapies have greatly reduced cancer mortality among populations, globally. However, this notable reduction in cancer mortality is discriminate and reflective of disparities between various ethnic populations and economic classes. Several factors contribute to this systemic inequity, at the level of diagnosis, cancer prognosis, therapeutics, and even point-of-care facilities.

Recent Findings: In this review, we have highlighted cancer health disparities among different populations around the globe. It encompasses social determinants such as status in society, poverty, education, diagnostic approaches including biomarkers and molecular testing, treatment as well as palliative care. Cancer treatment is an active area of constant progress and newer targeted treatments like immunotherapy, personalized treatment, and combinatorial therapies are emerging but these also show biases in their implementation in various sections of society. The involvement of populations in clinical trials and trial management is also a hotbed for racial discrimination. The immense progress in cancer management and its worldwide application needs a careful evaluation by identifying the biases in racial discrimination in healthcare facilities.

Conclusion: Our review gives a comprehensive evaluation of this global racial discrimination in cancer care and would be helpful in designing better strategies for cancer management and decreasing mortality.

KEYWORDS

cancer care, cancer health disparities, cancer management, cancer therapy, racial disparities

1 | INTRODUCTION

Despite coherent efforts that have led to a significant reduction in cancer mortality, it remains to be the second major cause of death, following cardiovascular diseases. With 215 deaths from cancer per 100000 individuals, the mortality rate peaked in 1991.[1] At the beginning of the present year, the American Cancer Society estimated a total of 1918030 collective cancer cases with 609360 deaths in

2022[2] and in 2020 there were about 10 million cancer deaths and an anticipated 19.3 million additional cancer incidences worldwide.[3] Cancer prevalence and cancer mortality are fortunately dropping in the world due to efficient healthcare facilities, better monitoring, early detection, and better cancer management. However, some populations continue to exhibit a greater risk of predominance and mortality concerning specific types of malignancies. Human populations all over the world are impacted by cancer but certain types of cancer are predominant in certain geographical locations.[4] Various factors have been attributed to the impact of this skewness, that include genetic,[5] socioeconomic,[6,7] and environmental factors.[8] The National Cancer Institute (NCI) specifies cancer health disparities as discrepancies in disease metrics, namely the occurrence rates, death rates, complications, survival rates, budgetary stress, and living standards [Courtesy: Cancer Disparities—NCI]. There appears to be a huge disparity in screening and early detection of cancer and the choice of treatment that is predominant among population subgroups. Disparities are apparent in the fact that although overall results show increasing awareness, better screening facilities, and significantly improved cancer mitigation, certain subgroups are not seeing the same gains as other groups. Such observations require a better understanding of the factors responsible for such differential mortality rates and improved cancer management and call for designing strategies for better implementation of the same.

These disparities are the outcome of complex and interconnected factors, making it challenging to separate them and analyze each factor's independent relative impact. Major cancer health disparities and associated mortality that have been noticed in different sections of the population are related to geographical locations, socioeconomic status, and genetics.[5,9] There is a large regional variation in both cancer cases, kind, and disease prognosis.

Breast cancer accounts for 25% of all women screened and leads to 16.6% of deaths due to cancer.[3] Incidence of breast cancer is significantly higher in developed countries like North America and Oceania (Figure 1). The proportion of risk factors for breast cancer has been known to be impacted by significant alterations in diet, lifestyle, sociocultural, and architectural environments brought on by developing countries and an increase in the number of women in the industries. Lung cancer represents 11.4% of total cancer cases.[3] The incidence of lung cancer as seen in Figure 1 (the primary data to synthesize the following secondary data was obtained from GLOBOCAN, 2020) is reflective of greater exposure to pollutants and is an unfortunate outcome of industrialization, hence Africa and Latin America exhibit relatively lower incidences. The prevalence of colorectal cancer is about 10% of the total incidences of cancers.[3] Incidence rates of colorectal cancer in North America and Oceania are higher than in others due to the predominance of junk food in the diet and a sedentary lifestyle. Heavy alcohol use, tobacco consumption, and intake of red or processed meat are other contributing risk factors. Prostate cancer, being one of the most diagnosed cancers in men accounts for variable frequencies in incidence worldwide. Latin America, North America, Europe, and Oceania show greater incidence predominantly due to regular monitoring and marker-based screening. The greatest incidence rates for prostate cancer among black males are found in the Caribbean and the United States.[11]

A higher mortality rate due to breast cancer in the African population, despite low incidence (Figure 2) exemplifies the fact that due to a lack of regular screening early detection of breast cancer does not take place, resulting in higher mortality rates. Lung cancer is a very aggressive cancer and hence leads to higher mortality in all populations with a higher incidence of the disease. The incidence and consequently, mortality rates are highly impacted by the state of industrialization and exposure to associated pollutants. Mortality rates due to colorectal cancers are more or less similar in all regions; the underlying reason for this could be urbanization, dependence on processed foods, a sedentary lifestyle, and lack of physical activities. Prostate cancer is curable if detected early, and hence, the lack of early detection of cancer in the African population due to a lack of

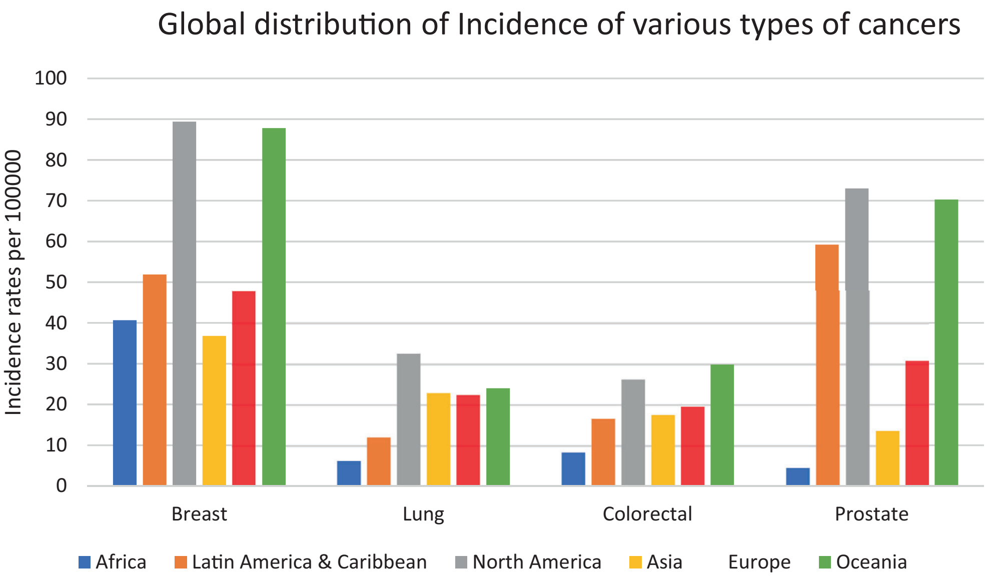


FIGURE 1 Incidence rates of common types of cancers in different ethnicities per 100 000. Blue, orange, gray, yellow, red, and green bars represent incidence rates per 100 000 for African, Latin American & Caribbean, North American, Asian, European, and Oceanic populations respectively for Breast, Lung, Colorectal, and Prostate cancers. *Source*: Data adapted from GLOBOCAN, 2020.[10]

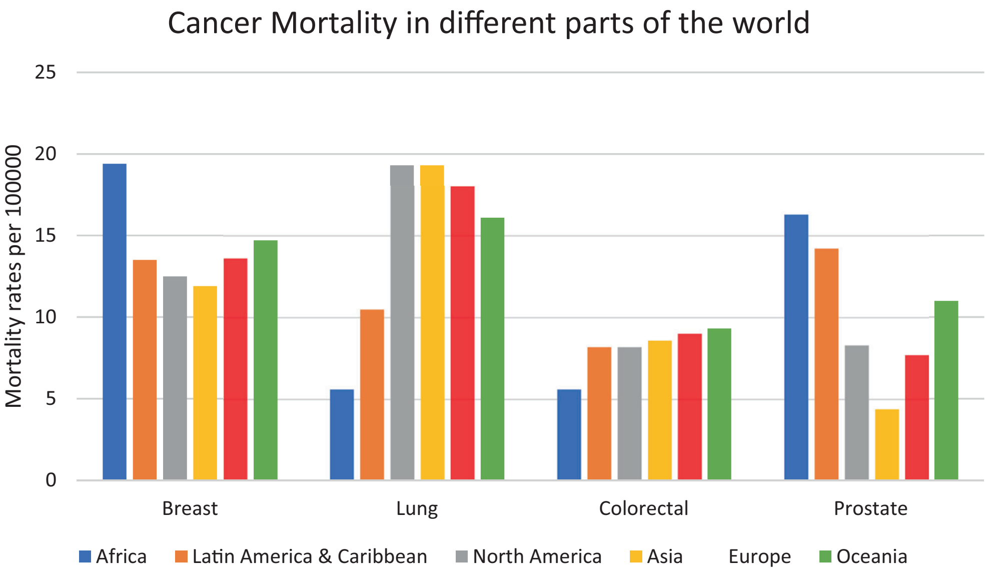


FIGURE 2 Mortality rates of common types of cancers in different ethnicities per 100 000. Blue, orange, gray, yellow, red, and green bars represent incidence rates per 100 000 for African, Latin American & Caribbean, North American, Asian, European, and Oceanic populations respectively for Breast, Lung, Colorectal, and Prostate cancers. *Source*: Data adapted from GLOBOCAN, 2020.[10]

effective screening facilities results in higher mortality rates as compared to North America, where the mortality rate is low despite a higher incidence, as seen in the graph.

Factors like insufficient knowledge, poverty, and health insurance appear to be equally important when compared to biological factors in accessing early diagnosis and appropriate treatment.[12] Additionally, societal injustices like the lingering effects of racial discrimination still have an impact on how patients and doctors connect.[13] Furthermore, cultural traits may influence how people behave in terms of their health management including regular health monitoring, prophylactic treatment, and whether they trust conventional medicine over alternative types of treatment.[14]

Reducing cancer deaths and increasing survival among the underprivileged constitute the eradication of inequities.[12] Our understanding of the mechanism and parameters contributing to the reduction in this disparity are inadequate due to the dearth of data, with regards to cancer mitigation and medical & palliative support. To overcome this shortcoming, large cohort studies are needed to be conducted and meta-analysis of the disparities in a fatality in various ethnic populations needs to be investigated more comprehensively. However, such widespread analyses are very limited due to prohibitive costs and lack of sensitivity. Our review aims to portray a comprehensive analysis of multiple parameters that are crucial in influencing disparities in cancer mortalities in different racial populations with an intent to suggest better mitigation strategies for cancer management across different sections of society.

2 | DISPARITIES DUE TO SOCIAL DETERMINANTS

Incidence, mortality, and risk factors for cancer vary not just by race and ethnicity but also by socioeconomic levels.[15,16] Indigence, culture, and societal injustice are socioeconomic factors that influence the discrepancy in deaths due to cancer.[17] A significant societal factor causing health disparities is poverty.[18] Tobacco use, inadequate diet, idleness, and being overweight are additional cancer hazard factors connected to socioeconomic disparities. Tobacco corporations frequently use poor and minority groups as sales targets. These groups frequently lack access to appropriate nutrition and fresh foods, as well as few possibilities for appropriate recreational body exercises.[19]

Research indicates that racial differences in the incidence of breast cancer seems less pronounced while there is a significant difference in the mortality rates in different ethnic groups. Social and economical aspects impact the choice of treatment and cancer management so significantly that the cancer outcome shows the immense disparity in ethnic populations.[20] Irrespective of ethnicity, poverty is linked to worse breast carcinoma results for all; nevertheless, because more black Americans compared to white populations are poor and are, therefore, very prone to exhibiting higher death rates.[21] Low income and unavailability of insurance coverage deter women from regular breast cancer screenings, resulting in higher chances of detection at a later stage resulting in a greater risk of mortality. Similarly, African and Asian women hardly visit a healthcare practitioner regularly, mammography frequencies are lower and the likelihood of early-stage detection is low.[22] Additionally, prohibitive costs often get subpar and unsuitable treatment increases the risk of death in these patients.[15] Recent advances in monoclonal antibody-based therapies specifically administered to breast cancer patients based on their marker profile have resulted in an immense increment in cancer regression in patients unresponsive to conventional treatments.[23] However, such treatments are expensive and often require advanced medical facilities which are unavailable to a great majority of women, the world over. A majority of the world population lack or have insufficient health insurance and relies on governmental interventions. Globally, women in various countries are living in locations with poor infrastructure, which makes it difficult for them to reach basic service facilities and doctors for diagnosis, treatment, or even follow-ups.[24]

Genetic profiling for the determination of propensity to certain forms of cancer and prophylactic vaccination for cancer is followed in less than 1% of the global population and is also disparate in different ethnic groups. This is often due to a lack of awareness and also due to social deterrents.

Cardiovascular disease, diabetes, hypertension, obesity, and respiratory illness are more common comorbidities among low-income women, which restricts their treatment choices.[25] Black females are more inclined than white females to consume a diet rich in fat, deficient, in vegetables and fruits, and are less likely to engage in routine exercise, therefore, more prone to be overweight.[26] The discrepancy in breast carcinoma rates among females is, therefore, affected by nutritional and lifestyle factors that are indirectly linked to socioeconomic constraints. Further ignorance of disease, lower levels of education, and religious and cultural taboos are additional factors that often lead to late-stage diagnosis and inappropriate treatments, leading to deaths.[27] Contrary to white women, black women are more likely to depend on supernatural and spiritual intervention, instead of getting the proper medical care, which can be harmful to their survival.[28] In general, societal injustice, poverty, and other variables play a direct and indirect role in the gap in breast carcinoma rates among females. Similar socioeconomic discrepancies are also observed in various developing nations including India.[29] Interestingly in India, there is a steep rise in incidences of breast cancer in urban women, mainly due to stress, lifestyle choices, late pregnancies, and late menopause. Among rural women, breast cancer incidences are however, lesser as compared to their urban counterparts, although there is a higher incidence of cervical cancer among them.[30,31] In rural India, the constraints leading to cancer mortality are often a lack of early diagnostics, advanced facilities and health care options. Hence, it is obvious that prudent cancer management is not only achievable by addressing economical and infrastructural constraints but also requires a holistic understanding of the factors in specific ethnic populations in addition to the uplifting of care facilities.

The unavailability of exposure to high-quality healthcare and therapeutic trials is considered to be the main cause of racial discrepancies in lung carcinoma survival.[32] It is significant to note that social determinants of wellness may contribute to differences in lung carcinoma therapy. These include, (1) both social and economic considerations, such as having health insurance or having the capacity to spend money for treatment, which affects the uninsured and disadvantaged populations, which comprises of many impoverished populations, in terms of access to effective adjuvant therapies.[33,34] (2) Lack of healthcare awareness, and literacy levels have an impact on the patient's choice of treatment and decisions of adherence and follow-ups for the treatment which are often prolonged. Poor healthcare awareness will probably have an impact on how well lung carcinoma sufferers comprehend their condition and can handle their respective treatment plans. (3) Improper patient treatment decisions are often the consequence of mistrust of the medical profession, which is a result of their past interactions with the medical system. Negative surgical views, fatalism, and skepticism have been put out as possible explanations for why certain patients are less able to adhere to and get prescribed therapy.[35] Mistrust is often fuelled by a dearth of knowledge about advancements in ethical principles controlling healthcare and the substandard treatment delivered in unregulated healthcare facilities. (4) Therapeutic inequities may be related to localities or neighborhoods having insufficient practical availability or usage of therapeutic facilities. People living in remote versus metropolitan locations, or living in a community with a high or low socioeconomic status are all connected to the lack of sufficient diagnostic and therapeutic facilities.

Colorectal cancer (CRC) is the third most typical cancer in the United States, irrespective of gender.[36] The chances of men getting CRC are more than women (4.3% vs. 4%). Age and hereditary risk factors are just two of the several variables that have been found to influence CRC development risk.[37–39] Being an aging-related condition, the probability of CRC increases with a person's age; according to recommendations, those with medium probability should begin screening tests at 50 years of age.[37] However, there are complicated correlations between race/ethnicity, socioeconomic status (SES), and CRC.[39,40] Poor diet and a sedentary lifestyle are two modifiable factors linked to CRC risk that is also linked to SES. Lifestyle choices may have an impact on the microbiota and biological behavior of colonic stem cells as well as the regional colonic environment.[41] A balanced diet, hormone replacement therapy, and aspirin prescription or NSAIDs may all lower the risk of CRC; exposure to these elements may also be correlated with SES and accessibility to healthcare. SES-related elements like income, education level, and medical insurance have an impact on who has access to resources and services for healthcare.[39]

It is well-accepted that social and economic variables influence prostate cancer occurrence. Prostate cancer risk frequently has an opposing relationship with social and economic position. Poor SES is linked to a reduced likelihood of surviving or quality of life. Based on socioeconomic background, race, level of education, and unemployment, prostate cancer survival vary dramatically. The adverse relationship between social assistance and the advanced stage of prostate cancer detection may be explained by several causes.[42–45] Men may be persuaded to get screened for prostate malignancy by their partner, other family members, or friends in their network. Married men are better at prostate cancer management due to early screening and better therapy than unmarried men.[46]

3 | DISPARITIES IN DIAGNOSIS

While Asian women often get breast cancer between the ages of 40 and 50, Non-Hispanic White women typically develop it between the ages of 60 and 70.[47] It is estimated that genetic factors account for 5%–10% of breast carcinoma cases.[48] The majority of autosomal dominant hereditary breast malignancies are resulting from alterations in BRCA genes (1 and 2), which are present at the 17th and 13th chromosomes, respectively. Human genes called BRCA1 and BRCA2 translate into tumor suppressors, which contribute to DNA repair as well as aid in preserving the integrity of the genomic information.

Whenever they are modified, it results in DNA damage and mutation and cells are more prone to further genetic changes that might result in the establishment of cancer. Depending on one's race and ethnicity, these alterations might occur more, or less frequently. For instance, Ashkenazi Jewish females (8.3%) had the greatest incidence of BRCA1 mutations. Hispanic females (3.5%), non-Hispanic white females (2.2%), Black females (1.3%), and Asian females (0.5%) are next.[49] Asian women are unlikely to undergo regular breast cancer screening as advised by WHO[49] and this may be the cause of the historically lower incidence of Breast cancer among Asian women, compared to their western counterparts. 55%–65% of BRCA1 mutation-containing females and 45% of BRCA2 mutation-containing females have a probability of developing breast cancer after the age of 70 years. Additionally, before 70 years of age, ovarian cancer may manifest in 39% of females with detrimental BRCA1 alteration and 11%–17% of females with BRCA2 alteration.[50] Despite the fact that deleterious BRCA 1 and BRCA 2 mutations are known to result in breast carcinoma in greater than 50% of the households with recurrent cases, mutations in other genes have also been associated with increased risks of the disease.[51,52] ATM, BRIP1, CHEK2, CDH1, MLH1, MLH2, MRE11A, NBN, PTEN, PALB2, RAD50, RAD51C, SEC23B, STK11, and TP53 genes all have rare mutations. By the time they are 70 years old, 33% of females who have a dangerous PALB2 gene alteration will have breast cancer. Those who have a hereditary background of breast malignancy and the dangerous PALB2 alteration are at even greater risk, 58%.[53] Several Asian racial groups are more likely to develop HER2-positive breast carcinoma.[54] Compared to more prevalent hormone-receptor-positive kinds of breast cancer, this biological subtype is more aggressive and has a worse prognosis.[55] Thus, genetic screening may give sufficient insight into the propensity of certain cancers and may serve as a basis for regular monitoring or even prophylactic surgical or vaccine-mediated interventions. However, such interventions are very rarely followed due to a lack of knowledge as well as social deterrents and taboos. Substantial variations were observed in 5-year survival rates of different ethnic populations in breast cancer research that included 777 Hispanic patients, 1016 Black patients, and 4885 White patients. Patients of Hispanic descent had survival rates of 70% ± 2%, Black patients of 65% ± 2%, and White patients of 75% ± 1%.[56] These variations are reflective of the stage of diagnosis as the principal factor. The percentage mortality rate for US patients for different stages of breast cancer is shown in Figure 3A and indicates that early detection irrespective of racial bias can lead to complete recovery in most cases.

According to estimates, there were 654620 individuals in the United States who have a background of pulmonary carcinoma, and another 236740 incidents have been discovered in 2022. The

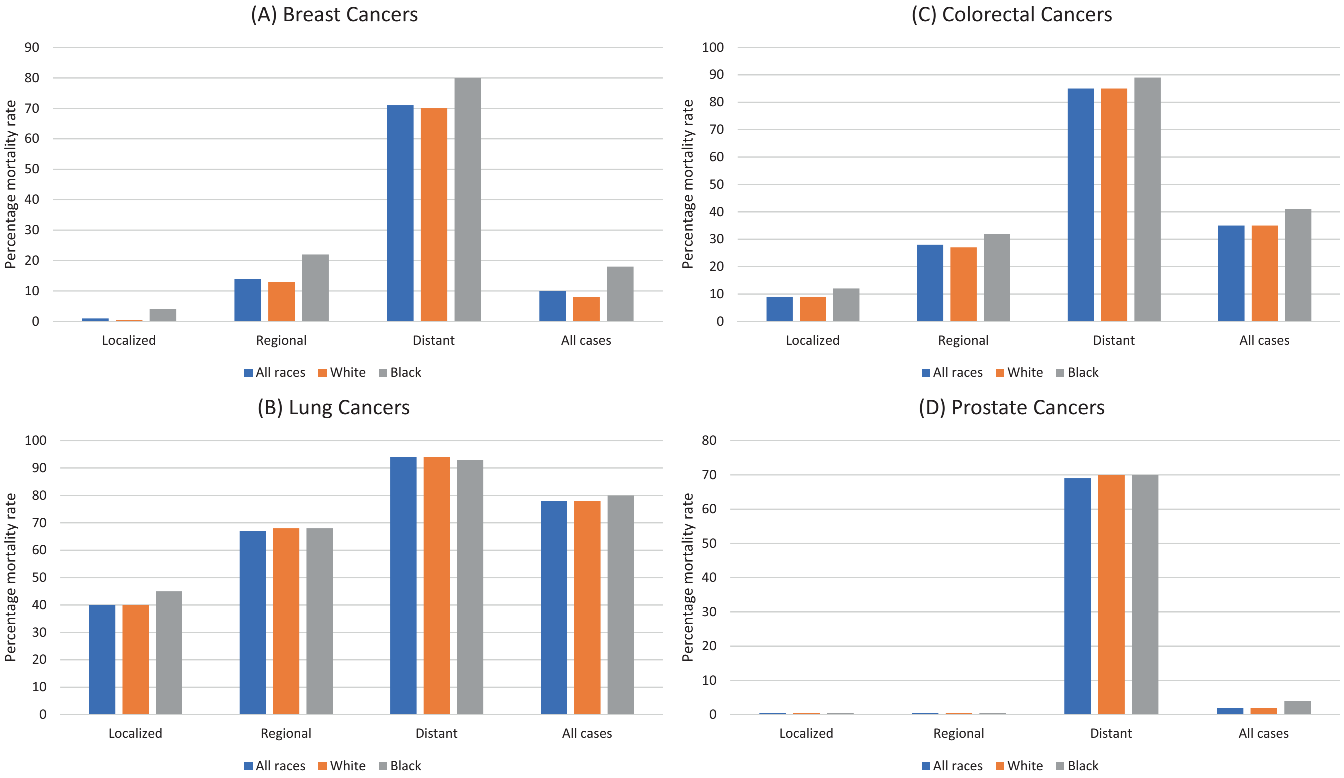


FIGURE 3 Percentage mortality rate for US patients with respect to different stages of cancers. (A) Breast cancers, (B) Lung cancers, (C) Colorectal cancers, and (D) Prostate cancers. Cancer is classified as localized, regional, or distant depending on the site of the disease. Localized refers to a disease that is limited to the site of origin, regional refers to cancer that has spread to an adjacent area, and distant refers to the post-malignancy stage of cancer. Mortality rates are shown in blue for all races, orange for white patients, and gray for black patients. *Source*: Graphs adapted from a study by Siegel et al.[2]

percentage mortality rate for US patients with respect to different stages of lung cancer is shown in Figure 3B. For therapeutic reasons, small cell lung cancer (SCLC; 14% of incidents) and non-small cell lung cancer (NSCLC; 82% of incidents) are two different types of lung cancer, with around 3% of cases having undetermined histology. The emergence of targeted cancer medicines has irrevocably altered the therapeutic scenario for NSCLC, making personalized treatment for lung carcinoma (i.e., NGS) the treatment of choice. In subpopulations of NSCLC sufferers who are eligible for this therapy, the clinical implementation of specific kinase medicines has increased the survival rate significantly.[32] ALK rearrangement, EGFR mutation, and PDL-1 testing are all recommended as part of molecular testing for metastatic NSCLC by NCCN in light of the significance of directed treatment for the overall control of lung cancer. As part of a comprehensive molecular profile, screening for mutations in BRAF, KRAS, RET, NTRK1/2/3, METex14 skipping, and ROS1 should also be done. It is crucial to take into account how race may affect biomarker screening and molecular analysis in lung cancer because these techniques are increasingly vital for improving cancer outcomes for individuals with NSCLC. Although targeted treatments were previously acknowledged as improvements in the management of NSCLC, it was highlighted that the distribution of their use across racial and socioeconomic strata was uneven.[57] One early research revealed that individuals with low incomes and those who lived in highly impoverished places were less probable to have EGFR screening. African Americans underwent lower rates of erlotinib and EGFR screening in univariate analysis than Whites, even after excluding the effects of socioeconomic, clinical, and demographic variables.[57] Other research has looked at the correlation between the probability of ordering an EGFR test and institutional and geographical features of the treating hospital as well as differences in socioeconomic determinants. It was shown that if the region had a wealthier or more educated population, hospitals were extra inclined to seek EGFR screening for individuals having advanced NSCLC.[58] Lynch et al. emphasized that there have been long-standing issues with getting anti-EGFR treatments and EGFR screening in regional hospitals, raising concerns that this might exacerbate the imbalance in cancer inequalities.[58] Although there were no racial differences in the frequencies of EGFR alterations and ALK rearrangements, individuals from most impoverished nations remained with a lower probability of ever having had any type of biomarker examined. The frequency of thorough genetic testing with NGS was even lower in all populations around the world. Compared to white patients, black patients had a lower likelihood of having had NGS analysis (39.8% vs. 50.1%, p 0.0001).[59]

More than 1.4 million males and females were anticipated to have received a colorectal carcinoma diagnosis as of January 1, 2022 and 151030 more patients are anticipated to get the diagnosis this year. KRAS mutations are present in around 45% of colorectal malignancy (CRC) patients.[60] About 12% of CRCs have a BRAF alteration (V600E), which is connected to a worse prognosis.[61] The percentage mortality rate for US patients with respect to different stages of breast cancer is shown in Figure 3C and is indicative of the fact that late detection results in a very high probability of fatal consequences, thus necessitating the emphasis on biomarker screening and early detection.

One of the most inherited cancers is prostate cancer which can be easily diagnosed at early stages by marker-based screening.[62] Incidents of prostate cancer strike one in nine American men throughout the course of their lifetimes. However, this ratio is one in seven for Black males, whose mortality rate is 1.7 times higher than their white counterparts.[9] The mortality rate for prostate cancer concerning different stages of cancers in different racial cross sections of patients is shown in Figure 3D. Notably, in prostate cancer, there is a significantly low mortality rate at the regional and localized stages as compared to other types of cancer. Prostate cancer is generally curable if detected early when the cancer is at a localized or regional stage. Black patients, however, often come for treatment only in the advanced stages of the disease and have high PSA values.[63] Black men had lower PSA screening rates than White men.[64,65] A similar scenario is seen in African men where lower incidence (Figure 1) is only reflective of poor screening and unfortunately results in a high mortality rate (Figure 2) as a result of late diagnosis. Due to advances in screening technologies, more than 50% reduction in prostate cancer occurrence has been seen since 1992 with an increase of more than 2% in overall survival rates.[66]

4 | DISPARITIES IN TREATMENT

Due to population expansion as well as improvements in early identification and treatment, there are more cancer survivors than ever before. As of Jan 1, 2022, over 4000000 females in the United States were projected to have a background of metastatic breast carcinoma, and an additional 287850 females will receive a new diagnosis. According to a study, three-fourths of the 150000 approx. breast carcinoma survivors who have the metastatic illness, are survivors who were detected early and initially confirmed at cancer stage I, II, or III.[67] More than 2.7 million females that are two-thirds of breast carcinoma survivors are 65 years of age or older, while just 6% are under 50. While one-third (34%) of females with stage I and stage II carcinoma receive mastectomy, frequently without chemotherapy or radiation, the remaining half of these women choose breast-conserving surgery (BCS) plus adjuvant radiotherapy. In contrast, 65% of females with stage III carcinoma, elect a mastectomy as their therapy of choice and in addition typically get chemotherapy. For stage I and stage II illness, Black females are less probable than White females to undergo BCS (60% vs. 64%, respectively). Black females are more probable to have just chemotherapy and/or irradiation for stage III illness (9% vs. 6%) and are less certain to undergo excision (57% vs. 66%). Sixty percent of female individuals with metastatic illness (stage IV) get just irradiation or chemotherapy. Adjuvant hormonal treatment is administered to at least 50% of females with invasive breast carcinoma who have tumors that express hormone receptors and who do not undergo carcinoma-targeted surgery, chemotherapy, or irradiation. Some BCS-eligible females choose to have surgery because they are reluctant to receive irradiation therapy, dread a recurrence, or have a medical condition that makes it impossible for them to receive irradiation.[68–70] Trends in the treatment of breast cancer are shown in Figure 4A,B.

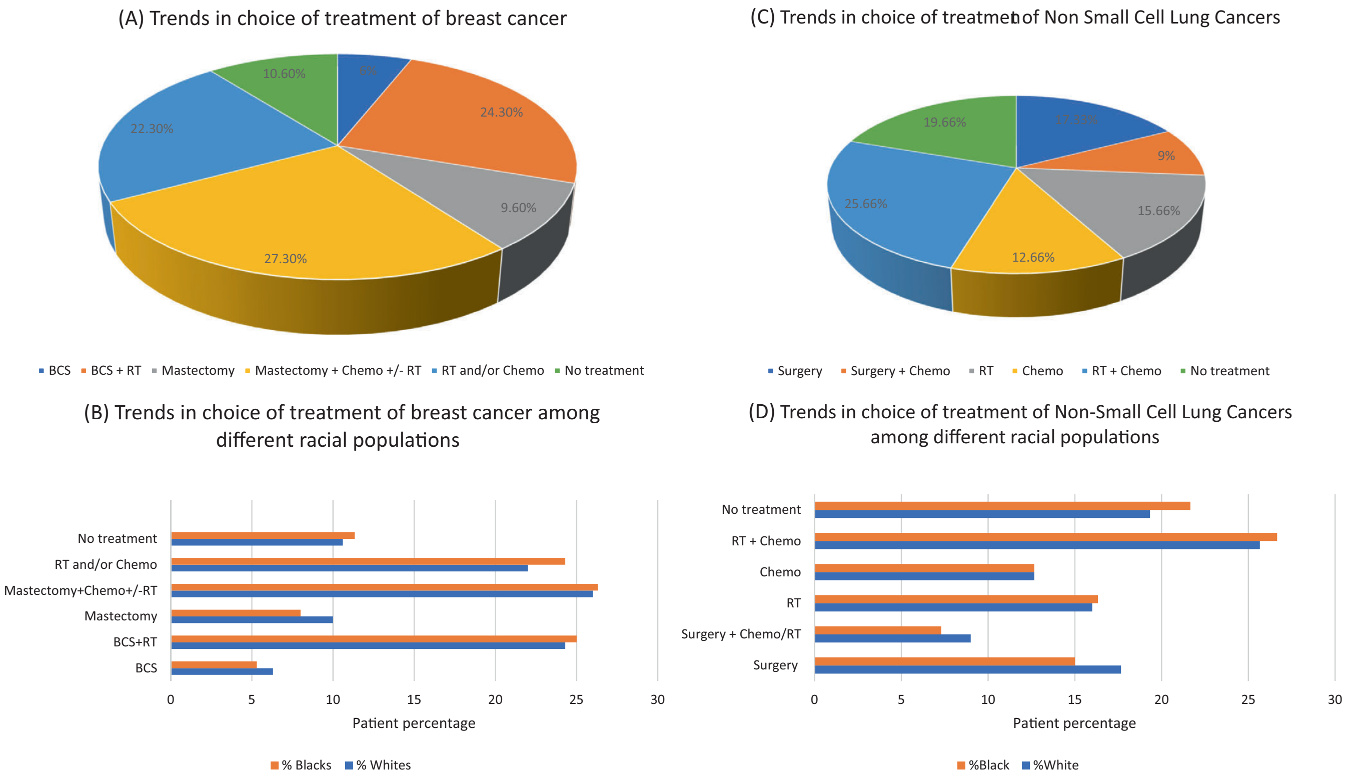


FIGURE 4 Trends in the choice of treatment for various cancers is a reflection of racial biases. (A) Trends in choice of treatment of breast cancer, (B) Different racial population show differences in their choice of treatment for breast cancer among US patients, (C) Trends in choice of treatment of non-small cell lung cancers, and (D) Different racial population show differences in their choice of treatment for non-small cell lung cancers among US patients. Orange shows the population percentage of Black patients for a particular type of treatment and blue shows the population percentage of White patients. *Source*: Data were obtained from a study by Miller et al.[72]

The accessibility of conveyance and/or the proximity to the treatment site might be structural barriers to undergoing irradiation therapy.[71] Black females are less inclined toward diagnosis at stage I carcinoma than White females that is 53% versus 68% of cases which leads to lower survival rates of black females for all stages, with the highest discrepancy for advanced malignancy that is 65% versus 77% for stage III and 19% versus 30% for stage IV.

Among NSCLC patients having stage I or II NSCLC, over 55% have surgery that involves wedge resection, sleeve resection, lobectomy, or pneumonectomy. Wedge resection involves the elimination of some part from a lung lobe; sleeve resection involves the elimination of the tumor plus a section of the damaged air track. In comparison, approximately one-fifth of patients with NSCLC stage III are able to receive surgical intervention; the majority (61%) receives chemotherapy and/or radiation therapy. Blacks are substantially lesser inclined than Whites to undergo surgery—16% versus 22% for stage III and 49% versus 55% for stages I and II. In comparison to whites, the (10%), frequency of receiving therapy in blacks (15%) is low in stages I and II illness. There is mixed evidence regarding whether, Black patients who receive platinum-based chemotherapy have a worse treatment outcome or more severe toxicity, which may influence survival, coupled with lower post-operative death rates. Trends in the treatment of lung cancer are shown in Figure 4C,D which compares the choice of treatment in different ethnic groups. Although it has been seen that surgery remains to be the choice of treatment, it is not preferred by most African Americans.

The majority of patients surviving colorectal carcinoma, involving both genders belong to the age group of 65 and above. About 67% of individuals with colorectal carcinoma (stage III) get chemotherapy involving adjuvant to discourage the chances of recurrence, compared to the bulk of patients at stage I & II colorectal carcinoma (84%), who undergo partial surgical removal of colon, avoiding chemotherapy. Proctectomy and related procedures are the most prevalent therapy for individuals suffering from rectal cancer stage I (61%) and almost half additionally get neoadjuvant irradiation or chemotherapy. Surgical plus neoadjuvant chemotherapy and irradiation treatment are often used to treat rectal tumors in stages II and III. Individuals having stage IV colon carcinoma (49%) and rectal carcinoma (29%), respectively, typically have surgery along with radiation and/or chemotherapy. Rectal cancer treatment differences between races are far higher than for colon carcinoma, which is probably due, at least in part, to the more complicated nature of care management. For both initial-stage colon and rectal malignancies, Black individuals are lesser inclined than White individuals to have surgery with the gap being substantially bigger for rectal carcinoma than it was previously noted for colon cancer.[73,74] Proctectomy or proctocolectomy is significantly less common for Black individuals with stage I rectal carcinoma than for White individuals (41% against 66%). While 7% of Blacks are unable to get any treatment, this figure is 3% for whites. Fifty-seven percent of Stage II/III black patients receive neoadjuvant chemo-radiotherapy before proctectomy or proctocolectomy as contrasted with 60% of White patients. The unavailability of skilled practitioners also contributes to treatment disparities. For example, there is fewer than one pathologist for every 500000 people in sub-Saharan Africa.[75] There are not enough trained cancer surgeons, less than two surgeons are available for every 100000 people. These ratios are significantly lower than the American average of 35 surgeons per 100000 people and one pathologist for every 15000 people.[76]

In the United States, there were more than 3.5 million males who have had prostate cancer previously, and in 2022, there might be 268490 new cases identified. Eighty-five percent of male prostate cancer survivors are above 65 years of age, while only 1% (12630) are under the age of 50. Active monitoring of low-risk illness climbed from 15% in 2010 to 42% in 2015[77] according to the National Comprehensive Cancer Network 2010 publication. This recommends toning down the excessive treatment,[78] whereas radical prostatectomy decreased from 47% to 31%. Numerous studies indicate a rise in proactive monitoring among elderly males aged 75 years and above, making them prone to early detection and 100% cancer regression.[79] However, although there is not much racial discrepancy as per genetic predisposition, regular screening remains to be the only determinant between 100% cancer recovery versus fatal outcome upon late diagnosis.[80–82] Hence the greatest solution to the racial discrepancy in its mortality lies in greater awareness and screening outreach measures that need to be implemented on a global scale.[83]

5 | DISPARITIES IN TARGETED TREATMENTS AND IMMUNOTHERAPY

Cancer health disparities have continued to increase despite advancements in treatment strategies. Cost and availability are the major factors that widen this gap.[84] Immunotherapy is currently regarded as a regular part of the first and foremost therapy for metastatic tumors which lack targetable mutations.[85] Mortality due to cancer has been reduced significantly due to the elevated use of immune checkpoint blockers. Disparities in targeted treatments like immunotherapy are seen even at the stage of recruitment for their clinical trials. Despite the fact that black populations have more cases of lung carcinoma, only 4.5% of blacks participated in screening trials[86] and 2% in the durvalumab trial for stage III NSCLC.[87] A retrospective cohort study using NCDB 2004–2012, also found a huge difference in their study sample among blacks who have opted for immunotherapy for melanoma, 97.7% less than whites.[88] A recent study shows the disparity in pembrolizumab trials for breast cancer patients, where approximately 12 white females have participated as opposed to only 1 black female patient for the immunotherapy trial.[89] In metastatic HCC, immunotherapy is preferred over chemotherapy for survival in general. Early accessibility to immunotherapy is characterized by major differences among Hispanics and Blacks as compared to Whites.[90]

PDL1 expression and tumor genomic profiles in breast, lung, and colorectal cancers have not been shown to differ in different ethnic groups, although there are notable variations in the immune cell population in the tumor microenvironment.[91] For example, compared to non-Hispanic white patients, breast tumors from black patients had a compelling immune cell prevalence and enhanced expression of inhibitory receptors like PD1, CTLA4, and LAG3.[92] Similar trends were also seen in prostate tumors for increased expression of proinflammatory genes among this population.[93] All these findings suggest better immunotherapeutic options for black patients but the reality is contrasting, due to patient-level factors (SES, behavior toward treatment and ethnicity, etc.), provider-level factors (cost of immunotherapy, knowledge, beliefs, and attitude toward patient, etc.) and system level factors (reimbursement and infrastructure quality, etc.).

Regulatory bodies play a role in specialized therapies like immunotherapy. It has been found that the percentage of patients having immunotherapy before and after approval by the FDA is increased to 12.4% in NSCLC.[94] Immune checkpoint inhibitor use has increased exponentially over the past 10 years, significantly decreasing carcinogenicity-based deaths. However, overall cancer registration among populations indicates the advantage of targeted treatments in non-Hispanic Whites, as compared to other minority subgroups. Contrary to popular belief, Asian lung cancer patients seemed to survive better as compared to other ethnic populations.[95]

6 | DISPARITIES IN CLINICAL TRIAL INVOLVEMENT

Clinical trials with sound design have improved the diagnosis and treatment of cancer. Cancer treatment is showing new innovations every day due to extensive improvements in our scientific knowledge. However, the implementation of novel treatment strategies requires the informed consent of patients, participating in clinical trials. Statistics reveal that less than 5% of cancer patients are confident enough to do so.[96–99] A meta-analysis revealed a somewhat better level of participation of 8% of patients in the case of industry-sponsored projects,[100] possibly due to perks given by industries but at the same time the recruitment of patients for trials by the industry is largely from academic centers whereas the proportion of patients in investigator-initiated trials is from community centers.[84] Geographical accessibility to a clinical study may affect its enrollment. There is evidence that unequal geographical accessibility to health care is correlated to adverse consequences and inferior quality of life as well as inadequate treatment compliance.[101] Syed et al.[102] have further demonstrated that minorities and individuals with lower incomes are disproportionately impacted by these differences. This hinders equitable representation in therapeutic studies as well. Only 37% of people with cancer in Pennsylvania who participated in a nationwide poll said they would commute to take part in a clinical trial.[103] Similar results were shown by Lara et al. in a prospective analysis of cancer patients being conducted at the University of California Davis Cancer Center, where the second most frequent explanation given for not participating in a trial was the patient's remoteness from the study center.[104] Recent research showed that even in the United States, clinical trials are not equally accessible.[105] According to a study by Galsky et al., 38.4%, 45.6%, 50.2%, and 52.2%, respectively, of patients with NSCLC, breast, prostate, and colorectal cancers need to travel more than an hour to reach a trial location.[106]

Phase 3 prostate cancer studies in the United States had a significant underrepresentation of Black males between 1987 and 2016; of the 72 clinical trials examined, 83.4% of the males who enrolled were White, compared to 6.7% who were Black.[107] In lower and middle-income nations throughout the world, the difference in access to trials for cancers is more pronounced. Only 1951 trials were available in lower and middle-income nations, compared to approximately 4700 trials in high-income nations for lung, breast, and cervical malignancies.[108] For instance, there are not many cancer clinical studies available in India, despite the significant patient burden. Only 350 interventional studies were filed between 2007 and 2017 according to a recent CTRI-listed audit of trials.[109] According to Carneiro et al., there are between 0.14 and 10.7 interventional trials per 100000 people in Europe.[110] African Americans make up about 5% of cancer clinical study enrollment. The execution of and accessibility to preventative clinical trials are likely to be impacted by the socioeconomic disparity and multiculturalism of the country.[111] Major issues responsible for this disparity included inadequate patient rights protection and compensation for the harm resulting from clinical trials, poor compliance with informed consent protocol, inadequate scientific and ethical review processes, subpar regulatory procedures for new drugs, and, most importantly, lack of post-trial cohort population's access to prohibitively expensive cancer treatments that had been demonstrated to be effective in low- and middle-income country settings. According to Agarwal et al.[112] and Joseph et al.,[113] there are a number of obstacles in conducting these researches, including the workforce's mobility, socioeconomic difficulties (such as gender inequality, casteism, and sickness stigma), and an absence of availability of primary healthcare facilities in low- and middle-income countries.

For NSCLC-focused treatment studies, genomic analysis is frequently a requirement for inclusion; as a result, disparities in comprehensive molecular profiling/NGS analysis may be significant in discrepancies in trial enrollment among racial groups. Recent research has focused on whether racial inequities exist in the use of biomarker screening and if inclusion in clinical trials is correlated with thorough genetic testing. Importantly, individuals were considerably more likely to take part in a clinical study if their tumors had undergone NGS analysis,[59] since such marker profiling gave greater promise of favorable outcomes, as evidenced due to targeted therapies.

7 | DISPARITIES IN PALLIATIVE CARE

Palliative care simply means "active overall care for patients whose conditions don't improve with treatment."[114] The effectiveness of pain treatment and the use of hospice care are major tenets of discrepancies based on socioeconomic factors and the availability of medical facilities. Nearly 58% of the world has palliative care facilities,[115] but they are not similar in all regions. USA, Australia, Europe, and Canada have modern facilities whereas South American and African regions are devoid of similar services.[115] While there are some similarities in palliative cancer care worldwide, significant differences exist in the prevalence, knowledge, and accessibility of palliative care facilities. There is often a lack of integrated cancer-specific treatments into care and cultural considerations that necessitate a customized approach to treatment.[116] A study found that obtaining any palliative care was considerably less likely at hospitals that served impoverished communities.[117] Non-Hispanic Blacks continue to be under-represented among hospice patients despite improvements in the accessibility of hospice care.[118] Despite a 14% increase in a hospice facility, an analysis of 204175 hospitalizations with late-stage cancer found that non-Hispanic Blacks were significantly less probable than their White counterparts to avail hospice care in terminally ill patients.[119] When palliative and hospice care among 133 non-Hispanic Black and White patients at a cancer treatment center were compared, non-Hispanic Blacks were found to have considerably lower levels of state-of-the-art facilities than Whites. A comprehensive study evaluating the effectiveness of cancer pain therapy studies conducted in North America, Europe, and Africa was published by Odonkor and colleagues.[120] Only 3 of the 18 studies were conducted in Africa (Egypt), and the investigators highlighted the uneven distribution of trials worldwide.[120] Only 41.4% of the respondents in a survey of 15 Middle Eastern nations said their organization had a palliative care facility.[121] The first step toward reducing inequities in the usage of appropriate treatment for cancer patients is increased knowledge and uniform accessibility of hospice & palliative care.[122] Reluctance to utilize hospice care more frequently is mostly due to: (1) Prohibitive costs; (2) Cultural or personal values at variance with modern hospice concept; (3) Ignorance of hospice care; (4) Absence of trust in the medical care; and (5) Reluctance of engaging financial burden of palliative care, especially in terminal conditions.[123] Both, gaps in critical disease treatment and inequities related to palliative care must be understood to eliminate the disparity. Analysis of 187 individuals who underwent hospitalized palliative care at a hospital revealed that location of birth and racial group were strongly linked with disposition.[124]

Studies have also revealed that minority groups, such as African Americans, Asian Americans, and Hispanics/Latinos seek hospice care less frequently.[125] Cultural variations in the impact of the disease or its mitigation and palliative care among different sections of patients and their families are also often variable in the western world as compared to close-knit societal frameworks as seen in South Asia, Southeast Asia, and the far East. Buddhism's predominant faith in "natural fate" urges sufferers to face pain as they await death. Since Buddhism predominates in China and Southeast Asia, there are reluctances to palliative hospice care.[126] In many countries, talking about cancer or palliative care is fraught with cultural taboos and fears of the disease.[127] Some ethnic groups continue to believe that cancer is infectious, especially in some regions of Africa[128] which has made it difficult to manage palliative treatment and has led to the isolation of patients due to social stigma.[129] This societal outlook toward palliative care also must be acknowledged and weighed against scientific rationale with compassionate patient management and the spread of awareness. State-of-the-art palliative treatment techniques need to be more uniformly distributed to ease pain and suffering, notwithstanding racial discrimination globally.[130] To evaluate and eradicate racial/ethnic inequities in hospice and palliative care, investigational strategies are required and the financial burden needs to be effectively managed.

8 | CONCLUSION

Growing industrialization is often associated with drastic changes in diet, lifestyle, and socio-economic conditions of people resulting in skewed incidences of certain types of cancers in a disparate manner around the world. Cancer care and novel treatment strategies have indeed resulted in a significant reduction in mortality rates of certain cancers and have also made preventive interventions possible. Despite significant advancements in our knowledge of certain biomarkers and their regular screening having an immense impact on cancer outcomes and mortality rates, inequalities exist in global populations sheerly due to late diagnosis. The prohibitive cost of treatment and unequal distribution of state-of-the-art hospice facilities and trained medical staff are significant causes of disparities in cancer treatment and global mortality rates. More studies are recommended to further streamline meticulous data collection of the medical system, genetic, and sociocultural environment in order to better identify and comprehend the pertinent levels of arbitration needed to reduce and finally eradicate cancer-related health discrepancies. However, just an increased understanding of its reasons, may not suffice alone to eradicate cancer health inequalities. To better understand the etiology of cancer and develop effective therapies oriented toward specific ethnicities, effort needs to be undertaken for acquiring genome analysis data sets and combinatorial therapeutics available in a broader scope of ethnic populations. Biomarker evaluation and prophylactic measures for high-risk groups compounded with regular screening are crucial for the ability to detect cancer at an early stage. As has been clearly shown by our study, despite the availability of state-of-the-art therapeutic options, late detection of cancer can in most cases adversely impact the cancer outcome. Thus early detection of most cancers can significantly lower mortality rates, irrespective of ethnicity necessitating more aggressive regular screening initiatives all over the world. In the case of certain cancers like prostate cancer, it may even result in 100% regression across all ethnic populations, provided the early diagnosis is effectuated. Prophylactic vaccination in certain cancers is not widely accepted due to social taboos which can be mitigated only by better education and awareness. It is crucial to broaden ongoing cultural and linguistic programs directed toward cancer awareness and broaden our outreach for better cancer management. The institutional elements and regulations that enable behavioral changes, such as tobacco control, should also be supported. Most crucially, government-driven schemes for improvements that support health equity, ubiquitous insurance policies, and availability of standard treatment for all must be ensured if the aforementioned disparities are to be erased.

AUTHOR CONTRIBUTIONS

Bharmjeet: Data curation (supporting); investigation (supporting); writing – original draft (supporting). **Asmita Das:** Conceptualization (lead); supervision (lead).

ACKNOWLEDGMENTS

The work has been supported by the Department of Biotechnology, Delhi Technological University.

CONFLICT OF INTEREST

The authors have stated explicitly that there are no conflicts of interest in connection with this article.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ETHICS STATEMENT

The present work has been done in compliance with the ethical guidelines of Delhi Technological University.

ORCID

Asmita Das https://orcid.org/0000-0001-9846-1005

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68(1):7-30. doi:10.3322/caac.21442
2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin*. 2022;72(1):7-33. doi:10.3322/caac.21708
3. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249. doi:10.3322/caac.21660
4. Goodwin B, Rowe A, Crawford-Williams F, et al. Geographical disparities in screening and cancer-related health behaviour. *Int J Environ Res Public Health*. 2020;17(4):1246. doi:10.3390/ijerph17041246
5. Fejerman L, John EM, Huntsman S, et al. Genetic ancestry and risk of breast cancer among U.S. Latinas. *Cancer Res*. 2008;68(23):9723-9728. doi:10.1158/0008-5472.CAN-08-2039
6. Silber JH, Rosenbaum PR, Ross RN et al. Disparities in breast cancer survival by socioeconomic status despite medicare and medicaid insurance. *Milbank Q*. 2018;96(4):706-754. doi:10.1111/1468-0009.12355
7. Freeman HP. Cancer in the socioeconomically disadvantaged. *CA Cancer J Clin*. 1989;39(5):266-288. doi:10.3322/canjclin.39.5.266
8. Zavala VA, Bracci PM, Carethers JM, et al. Cancer health disparities in racial/ethnic minorities in the United States. *Br J Cancer*. 2021;124(2):315-332. doi:10.1038/s41416-020-01038-6
9. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70(1):7-30. doi:10.3322/caac.21590
10. "Global Cancer Observatory.". https://gco.iarc.fr/. Accessed Jan 30, 2023
11. Rebbeck TR, Devesa SS, Chang BL, et al. Global patterns of prostate cancer incidence, aggressiveness, and mortality in men of african descent. *Prostate Cancer*. 2013;2013:1-12. doi:10.1155/2013/560857
12. Bhatia S. Disparities in cancer outcomes: lessons learned from children with cancer. *Pediatr Blood Cancer*. 2011;56(6):994-1002. doi:10.1002/pbc.23078

13. *Unequal Treatment*. National Academies Press; 2003. doi:10.17226/12875
14. Freeman HP. Commentary on the meaning of race in science and society. *Cancer Epidemiol Biomark Prev*. 2003;12(3):232s-236s.
15. Bigby J, Holmes MD. Disparities across the breast cancer continuum. *Cancer Causes Control*. 2005;16(1):35-44. doi:10.1007/s10552-004-1263-1
16. Newman LA, Martin IK. Disparities in breast cancer. *Curr Probl Cancer*. 2007;31(3):134-156. doi:10.1016/j.currproblcancer.2007.01.003
17. Freeman HP, Chu KC. Determinants of cancer disparities: barriers to cancer screening, diagnosis, and treatment. *Surg Oncol Clin N Am*. 2005;14(4):655-669. doi:10.1016/j.soc.2005.06.002
18. Freeman HP. Poverty, culture, and social injustice: determinants of cancer disparities. *CA Cancer J Clin*. 2004;54(2):72-77. doi:10.3322/canjclin.54.2.72
19. Bernstein L, Teal CR, Joslyn S, Wilson J. Ethnicity-related variation in breast cancer risk factors. *Cancer*. 2003;97(1 Suppl):222-229. doi:10.1002/cncr.11014
20. Yedjou CG, Sims JM, Miele L, Sims JM, Miele L et al. Health and racial disparity in breast cancer. *Adv Exp Med Biol*. 2019;1152:31-49. doi:10.1007/978-3-030-20301-6_3
21. Gerend MA, Pai M. Social determinants of black-white disparities in breast cancer mortality: a review. *Cancer Epidemiol Biomark Prev*. 2008;17(11):2913-2923. doi:10.1158/1055-9965.EPI-07-0633
22. O'Malley AS, Forrest CB, Mandelblatt J. Adherence of low-income women to cancer screening recommendations. *J Gen Intern Med*. 2002;17(2):144-154. doi:10.1046/j.1525-1497.2002.10431.x
23. Vagia E, Mahalingam D, Cristofanilli M. The landscape of targeted therapies in TNBC. *Cancers (Basel)*. 2020;12(4):916. doi:10.3390/cancers12040916
24. Lacey L, Whitfield J, Dewhite W, et al. Referral adherence in an inner city breast and cervical cancer screening program. *Cancer*. 1993;72(3):950-955. doi:10.1002/1097-0142(19930801)72:33.0.CO;2-S
25. Tammemagi CM. Comorbidity and survival disparities among black and white patients with breast cancer. *JAMA*. 2005;294(14):1765. doi:10.1001/jama.294.14.1765
26. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999-2004. *JAMA*. 2006;295(13):1549. doi:10.1001/jama.295.13.1549
27. Johnson KS, Elbert-Avila KI, Tulsky JA. The influence of spiritual beliefs and practices on the treatment preferences of African Americans: a review of the literature. *J Am Geriatr Soc*. 2005;53(4):711-719. doi:10.1111/j.1532-5415.2005.53224.x
28. Lannin DR, Mathews HF, Mitchell J, Swanson MS. Impacting cultural attitudes in African-American women to decrease breast cancer mortality. *Am J Surg*. 2002;184(5):418-423. doi:10.1016/S0002-9610(02)01009-7
29. Negi J, Nambiar D. Intersectional social-economic inequalities in breast cancer screening in India: analysis of the National Family Health Survey. *BMC Womens Health*. 2021;21(1):324. doi:10.1186/s12905-021-01464-5
30. Blake KD, Moss JL, Gaysynsky A, Srinivasan S, Croyle RT. Making the case for investment in rural cancer control: an analysis of rural cancer incidence, mortality, and funding trends. *Cancer Epidemiol Biomark Prev*. 2017;26(7):992-997. doi:10.1158/1055-9965.EPI-17-0092
31. Moss JL, Liu B, Feuer EJ. Urban/rural differences in breast and cervical cancer incidence: the mediating roles of socioeconomic status and provider density. *Womens Health Issues*. 2017;27(6):683-691. doi:10.1016/j.whi.2017.09.008
32. Lin JJ, Cardarella S, Lydon CA, et al. Five-year survival in EGFR-mutant metastatic lung adenocarcinoma treated with EGFR-TKIs. *J Thorac Oncol*. 2016;11(4):556-565. doi:10.1016/j.jtho.2015.12.103
33. Verma V, Haque W, Cushman TR, et al. Racial and insurance-related disparities in delivery of immunotherapy-type compounds in the United States. *J Immunother*. 2019;42(2):55-64. doi:10.1097/CJI.0000000000000253
34. Farrow NE, An SJ, Speicher PJ, et al. Disparities in guideline-concordant treatment for node-positive, non–small cell lung cancer following surgery. *J Thorac Cardiovasc Surg*. 2020;160(1):261-271.e1. doi:10.1016/j.jtcvs.2019.10.102
35. Lin JJ, Mhango G, Wall MM, et al. Cultural factors associated with racial disparities in lung cancer care. *Ann Am Thorac Soc*. 2014;11(4):489-495. doi:10.1513/AnnalsATS.201402-055OC
36. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA Cancer J Clin*. 2021;71(1):7-33. doi:10.3322/caac.21654
37. Rex DK, Boland RC, Dominitz JA, et al. Colorectal cancer screening: recommendations for physicians and patients from the U.S. multi-society task force on colorectal cancer. *Am J Gastroenterol*. 2017;112(7):1016-1030. doi:10.1038/ajg.2017.174
38. Carethers JM. Lynch syndrome and Lynch syndrome mimics: the growing complex landscape of hereditary colon cancer. *World J Gastroenterol*. 2015;21(31):9253-9261. doi:10.3748/wjg.v21.i31.9253
39. Carethers JM. Screening for colorectal cancer in African Americans: determinants and rationale for an earlier age to commence screening. *Dig Dis Sci*. 2015;60(3):711-721. doi:10.1007/s10620-014-3443-5
40. Carethers JM. Clinical and genetic factors to inform reducing colorectal cancer disparitites in African Americans. *Front Oncol*. 2018;8.
41. Carethers JM, Doubeni CA. Causes of socioeconomic disparities in colorectal cancer and intervention framework and strategies. *Gastroenterology*. 2020;158(2):354-367. doi:10.1053/j.gastro.2019.10.029
42. Li ML, Lin J, Hou JG, et al. Environmental and psycho-social factors related to prostate cancer risk in the Chinese population: a case-control study. *Biomed Environ Sci*. 2014;27(9):707-717. doi:10.3967/bes2014.089
43. Lynch SM, Mitra N, Ross M, et al. A neighborhood-wide association study (NWAS): example of prostate cancer aggressiveness. *PLoS One*. 2017;12(3):e0174548. doi:10.1371/journal.pone.0174548
44. Aizer AA, Chen MH, McCarthy EP, et al. Marital status and survival in patients with cancer. *J Clin Oncol*. 2013;31(31):3869-3876. doi:10.1200/JCO.2013.49.6489
45. Bergelt C, Prescott E, Grønbæk M, Koch U, Johansen C. Social ties and risk for cancer—a prospective cohort study. *Acta Oncol (Madr)*. 2009;48(7):1010-1018. doi:10.1080/02841860903036230
46. Coughlin SS. A review of social determinants of prostate cancer risk, stage, and survival. *Prostate Int*. 2020;8(2):49-54. doi:10.1016/j.prnil.2019.08.001
47. Leong SPL, Shen ZZ, Liu TJ, et al. Is breast cancer the same disease in Asian and Western countries? *World J Surg*. 2010;34(10):2308-2324. doi:10.1007/s00268-010-0683-1
48. Bogdanova N, Helbig S, Dörk T. Hereditary breast cancer: ever more pieces to the polygenic puzzle. *Hered Cancer Clin Pract*. 2013;11(1):12. doi:10.1186/1897-4287-11-12
49. John EM, Miron A, Gong G, et al. Prevalence of pathogenic BRCA1 mutation carriers in 5 US racial/ethnic groups. *JAMA*. 2007;298(24):2869-2876. doi:10.1001/jama.298.24.2869
50. Chen S, Parmigiani G. Meta-analysis of *BRCA1* and *BRCA2* penetrance. *J Clin Oncol*. 2007;25(11):1329-1333. doi:10.1200/JCO.2006.09.1066
51. Campeau PM, Foulkes WD, Tischkowitz MD. Hereditary breast cancer: new genetic developments, new therapeutic avenues. *Hum Genet*. 2008;124(1):31-42. doi:10.1007/s00439-008-0529-1
52. Walsh T, Casadei S, Coats KH, et al. Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at High risk of breast cancer. *JAMA*. 2006;295(12):1379-1388. doi:10.1001/jama.295.12.1379
53. Antoniou AC, Casadei S, Heikkinen T, et al. Breast-cancer risk in families with mutations in *PALB2*. *N Engl J Med*. 2014;371(6):497-506. doi:10.1056/NEJMoa1400382
54. Gomez SL, Yao S, Kushi LH, Kurian AW. Is breast cancer in Asian and Asian American women a different disease? *JNCI: J Natl Cancer Inst*. 2019;111(12):1243-1244. doi:10.1093/jnci/djz091

55. Telli ML, Chang ET, Kurian AW, et al. Asian ethnicity and breast cancer subtypes: a study from the California cancer registry. *Breast Cancer Res Treat*. 2011;127(2):471-478. doi:10.1007/s10549-010-1173-8
56. PDQ Cancer Genetics Editorial Board, *Genetics of Breast and Gynecologic Cancers* (PDQ®): Health Professional Version. 2002.
57. Palazzo LL, Sheehan DF, Tramontano AC, Kong CY. Disparities and trends in genetic testing and erlotinib treatment among metastatic non–small cell lung cancer patients. *Cancer Epidemiol Biomarkers Prev*. 2019;28(5):926-934. doi:10.1158/1055-9965.EPI-18-0917
58. Lynch JA, Khoury MJ, Borzecki A, et al. Utilization of epidermal growth factor receptor (EGFR) testing in the United States: a case study of T3 translational research. *Genet Med*. 2013;15(8):630-638. doi:10.1038/gim.2013.5
59. Bruno DS, Hess LM, Li X, Su EW, Zhu YE, Patel M. Racial disparities in biomarker testing and clinical trial enrollment in non-small cell lung cancer (NSCLC). *J Clin Oncol*. 2021;39(15_suppl):9005. doi:10.1200/JCO.2021.39.15_suppl.9005
60. Bos JL. The ras gene family and human carcinogenesis. *Mutat Res/Rev Genet Toxicol*. 1988;195(3):255-271. doi:10.1016/0165-1110(88)90004-8
61. Cantwell-Dorris ER, O'Leary JJ, Sheils OM. BRAFV600E: implications for carcinogenesis and molecular therapy. *Mol Cancer Ther*. 2011;10(3):385-394. doi:10.1158/1535-7163.MCT-10-0799
62. Rebbeck TR. Prostate cancer disparities by race and ethnicity: from nucleotide to neighborhood. *Cold Spring Harb Perspect Med*. 2018;8(9):a030387. doi:10.1101/cshperspect.a030387
63. Jemal A, Culp MB, Ma J, Islami F, Fedewa SA. Prostate cancer incidence 5 years after US preventive services task force recommendations against screening. *JNCI: J Natl Cancer Inst*. 2021;113(1):64-71. doi:10.1093/jnci/djaa068
64. Kensler KH, Pernar CH, Mahal BA, et al. Racial and ethnic variation in PSA testing and prostate cancer incidence following the 2012 USPSTF recommendation. *JNCI: J Natl Cancer Inst*. 2021;113(6):719-726. doi:10.1093/jnci/djaa171
65. Oliver JS, Allen RS, Eichorst MK, et al. A pilot study of prostate cancer knowledge among African American men and their health care advocates: implications for screening decisions. *Cancer Causes Control*. 2018;29(7):699-706. doi:10.1007/s10552-018-1041-0
66. Siegel DA, O'Neil ME, Richards TB, Dowling NF, Weir HK. Prostate cancer incidence and survival, by stage and race/ethnicity—United States, 2001–2017. *MMWR Morb Mortal Wkly Rep*. 2020;69(41):1473-1480. doi:10.15585/mmwr.mm6941a1
67. Mariotto AB, Etzioni R, Hurlbert M, Penberthy L, Mayer M. Estimation of the number of women living with metastatic breast cancer in the United States. *Cancer Epidemiol Biomarkers Prev*. 2017;26(6):809-815. doi:10.1158/1055-9965.EPI-16-0889
68. Montagna G, Morrow M. Contralateral prophylactic mastectomy in breast cancer: what to discuss with patients. *Expert Rev Anticancer Ther*. 2020;20(3):159-166. doi:10.1080/14737140.2020.1732213
69. Albornoz CR, Matros E, Lee CN, et al. Bilateral mastectomy versus breast-conserving surgery for early-stage breast cancer. *Plast Reconstr Surg*. 2015;135(6):1518-1526. doi:10.1097/PRS.0000000000001276
70. Kummerow KL, Du L, Penson DF, Shyr Y, Hooks MA. Nationwide trends in mastectomy for early-stage breast cancer. *JAMA Surg*. 2015;150(1):9-16. doi:10.1001/jamasurg.2014.2895
71. Lautner M, Lin H, Shen Y, et al. Disparities in the use of breast-conserving therapy among patients with early-stage breast cancer. *JAMA Surg*. 2015;150(8):778-786. doi:10.1001/jamasurg.2015.1102
72. Miller KD, Nogueira L, Devasia T, et al. Cancer treatment and survivorship statistics, 2022. *CA Cancer J Clin*. 2022;72(5):409-436. doi:10.3322/caac.21731
73. Hao S, Snyder RA, Irish W, Parikh AA. Association of race and health insurance in treatment disparities of colon cancer: a retrospective analysis utilizing a national population database in the United States. *PLoS Med*. 2021;18(10):e1003842. doi:10.1371/journal.pmed.1003842
74. Murphy CC, Harlan LC, Warren JL, Geiger AM. Race and insurance differences in the receipt of adjuvant chemotherapy among patients with stage III colon cancer. *J Clin Oncol*. 2015;33(23):2530-2536. doi:10.1200/JCO.2015.61.3026
75. Adesina A, Chumba D, Nelson AM, et al. Improvement of pathology in sub-Saharan Africa. *Lancet Oncol*. 2013;14(4):e152-e157. doi:10.1016/S1470-2045(12)70598-3
76. Meara JG, Leather AJM, Hagander L, et al. Global surgery 2030: evidence and solutions for achieving health, welfare, and economic development. *Int J Obstet Anesth*. 2016;25:75-78. doi:10.1016/j.ijoa.2015.09.006
77. Mahal BA, Butler S, Franco I, et al. Use of active surveillance or watchful waiting for low-risk prostate cancer and management trends across risk groups in the United States, 2010-2015. *JAMA*. 2019;321(7):704-706. doi:10.1001/jama.2018.19941
78. Mohler J, Bahnson RR, Boston B, et al. Prostate cancer. *J Natl Compr Canc Netw*. 2010;8(2):162-200. doi:10.6004/jnccn.2010.0012
79. Cooperberg MR, Carroll PR. Trends in management for patients with localized prostate cancer, 1990-2013. *JAMA*. 2015;314(1):80-82. doi:10.1001/jama.2015.6036
80. Nocera L, Wenzel M, Collà Ruvolo C, et al. The effect of race/ethnicity on active treatment rates among septuagenarian or older low risk prostate cancer patients. *Urol Oncol: Semin Orig Investig*. 2021;39(11):785.e11-785.e17. doi:10.1016/j.urolonc.2021.04.004
81. Liu Y, Hall IJ, Filson C, Howard DH. Trends in the use of active surveillance and treatments in Medicare beneficiaries diagnosed with localized prostate cancer. *Urol Oncol: Semin Orig Investig*. 2021;39(7):432.e1-432.e10. doi:10.1016/j.urolonc.2020.11.024
82. Washington SL, Jeong CW, Loner Gan PE et al. Regional variation in active surveillance for low-risk prostate cancer in the US. *JAMA Netw Open*. 2020;3(12):e2031349. doi:10.1001/jamanetworkopen.2020.31349
83. Presley CJ, Raldow AC, Cramer LD, et al. A new approach to understanding racial disparities in prostate cancer treatment. *J Geriatr Oncol*. 2013;4(1):1-8. doi:10.1016/j.jgo.2012.07.005
84. Osarogiagbon RU, Sineshaw HM, Unger JM, Acuña-Villaorduña A, Goel S. Immune-based cancer treatment: addressing disparities in access and outcomes. *Am Soc Clin Oncol Educ Book*. 2021;41:66-78. doi:10.1200/EDBK_323523
85. Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al. Pembrolizumab plus chemotherapy in metastatic non–small-cell lung cancer. *N Engl J Med*. 2018;378(22):2078-2092. doi:10.1056/NEJMoa1801005
86. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5):395-409. doi:10.1056/NEJMoa1102873
87. Antonia SJ, Villegas A, Daniel D, et al. Durvalumab after chemoradiotherapy in stage III non–small-cell lung cancer. *N Engl J Med*. 2017;377(20):1919-1929. doi:10.1056/NEJMoa1709937
88. Disparities of immunotherapy utilization in patients with stage III cutaneous melanoma: a National Perspective. *Anticancer Res*. 2018;38(5):2897-2901. doi:10.21873/anticanres.12536
89. Grette KV, White AL, Awad EK et al. Not immune to inequity: minority under-representation in immunotherapy trials for breast and gynecologic cancers. *Int J Gynecol Cancer*. 2021;31(11):1403-1407. doi:10.1136/ijgc-2021-002557
90. Ahn JC, Lauzon M, Luu M, et al. Racial and ethnic disparities in early treatment with immunotherapy for advanced HCC in the United States. *Hepatology*. 2022;76(6):1649-1659. doi:10.1002/hep.32527
91. Mitchell KA, Zingone A, Toulabi L, Boeckelman J, Ryan BM. Comparative transcriptome profiling reveals coding and noncoding RNA differences in NSCLC from African Americans and European Americans. *Clin Cancer Res*. 2017;23(23):7412-7425. doi:10.1158/1078-0432.CCR-17-0527
92. Yao S, Cheng TYD, Elkhanany A, et al. Breast tumor microenvironment in black women: a distinct signature of CD8+ T-cell exhaustion. *JNCI: J Natl Cancer Inst*. 2021;113(8):1036-1043. doi:10.1093/jnci/djaa215

93. Awasthi S, Berglund A, Abraham-Miranda J, et al. Comparative genomics reveals distinct immune-oncologic pathways in African American men with prostate cancer. *Clin Cancer Res*. 2021;27(1):320-329. doi:10.1158/1078-0432.CCR-20-2925
94. Ermer T, Canavan ME, Maduka RC, et al. Association between Food and Drug Administration approval and disparities in immunotherapy use among patients with cancer in the US. *JAMA Netw Open*. 2022;5(6):e2219535. doi:10.1001/jamanetworkopen.2022.19535
95. Qian J, Nie W, Lu J, et al. Racial differences in characteristics and prognoses between Asian and white patients with nonsmall cell lung cancer receiving atezolizumab: An ancillary analysis of the POPLAR and OAK studies. *Int J Cancer*. 2020;146(11):3124-3133. doi:10.1002/ijc.32717
96. Wong AR, Sun V, George K, et al. Barriers to participation in therapeutic clinical trials as perceived by community oncologists. *JCO Oncol Pract*. 2020;16(9):e849-e858. doi:10.1200/JOP.19.00662
97. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials. *JAMA*. 2004;291(22):2720. doi:10.1001/jama.291.22.2720
98. Sateren WB, Trimble EL, Abrams J, et al. How sociodemographics, presence of oncology specialists, and hospital cancer programs affect accrual to cancer treatment trials. *J Clin Oncol*. 2002;20(8):2109-2117. doi:10.1200/JCO.2002.08.056
99. Tejeda HA, Green SB, Trimble EL, et al. Representation of African-Americans, Hispanics, and whites in National Cancer Institute cancer treatment trials. *JNCI J Natl Cancer Inst*. 1996;88(12):812-816. doi:10.1093/jnci/88.12.812
100. Unger JM, Vaidya R, Hershman DL, Minasian LM, Fleury ME. Systematic review and meta-analysis of the magnitude of structural, clinical, and physician and patient barriers to cancer clinical trial participation. *JNCI: J Natl Cancer Inst*. 2019;111(3):245-255. doi:10.1093/jnci/djy221
101. Ambroggi M, Biasini C, del Giovane C, Fornari F, Cavanna L. Distance as a barrier to cancer diagnosis and treatment: review of the literature. *Oncologist*. 2015;20(12):1378-1385. doi:10.1634/theoncologist.2015-0110
102. Syed ST, Gerber BS, Sharp LK. Traveling towards disease: transportation barriers to health care access. *J Community Health*. 2013;38(5):976-993. doi:10.1007/s10900-013-9681-1
103. Meropol NJ, Buzaglo JS, Millard J, et al. Barriers to clinical trial participation as perceived by oncologists and patients. *J Natl Compr Canc Netw*. 2007;5(8):753-762. doi:10.6004/jnccn.2007.0067
104. Lara PN, Higdon R, Lim N et al. Prospective evaluation of cancer clinical trial accrual patterns: identifying potential barriers to enrollment. *J Clin Oncol*. 2001;19(6):1728-1733. doi:10.1200/JCO.2001.19.6.1728
105. Feyman Y, Provenzano F, David FS. Disparities in clinical trial access across US urban areas. *JAMA Netw Open*. 2020;3(2):e200172. doi:10.1001/jamanetworkopen.2020.0172
106. Galsky MD, Stensland KD, McBride RB, et al. Geographic accessibility to clinical trials for advanced cancer in the United States. *JAMA Intern Med*. 2015;175(2):293-295. doi:10.1001/jamainternmed.2014.6300
107. Rencsok EM, Bazzi LA, McKay RR, et al. Diversity of enrollment in prostate cancer clinical trials: current status and future directions. *Cancer Epidemiol Biomarkers Prev*. 2020;29(7):1374-1380. doi:10.1158/1055-9965.EPI-19-1616
108. Ramaswami R, Paulino E, Barrichello A, et al. Disparities in breast, lung, and cervical cancer trials worldwide. *J Glob Oncol*. 2018;4:1-11. doi:10.1200/JGO.17.00226
109. Roy AM, Mathew A. Audit of cancer clinical trials in India. *J Glob Oncol*. 2019;5:1. doi:10.1200/JGO.19.00156
110. Carneiro A, Amaral TMS, Brandao M, et al. LBA66_PR disparities in access to oncology clinical trials in Europe in the period 2009-2019. *Ann Oncol*. 2020;31:S1196. doi:10.1016/j.annonc.2020.08.2301
111. Thulaseedharan JV, Frie KG, Sankaranarayanan R. Challenges of health promotion and education strategies to prevent cervical cancer in India: a systematic review. *J Educ Health Promot*. 2019;8:216. doi:10.4103/jehp.jehp_156_19
112. Agrawal T, Fathima F, Hegde SK, Joshi R, Srinivasan N, Misquith D. Challenges in conducting community-based trials of primary prevention of cardiovascular diseases in resource-constrained rural settings. *WHO South East Asia J Public Health*. 2015;4(1):98-103. doi:10.4103/2224-3151.206628
113. Joseph LM, Lekha TR, Boban D, Jose P, Jeemon P. Perceived facilitators and barriers of enrolment, participation and adherence to a family based structured lifestyle modification interventions in Kerala, India: a qualitative study. *Wellcome Open Res*. 2019;4:131. doi:10.12688/wellcomeopenres.15415.2
114. Gluyas H. Patient-centred care: improving healthcare outcomes. *Nurs Stand*. 2015;30(4):50-59. doi:10.7748/NS.30.4.50.E10186
115. Lynch T, Connor S, Clark D. Mapping levels of palliative care development: a global update. *J Pain Symptom Manage*. 2013;45(6):1094-1106. doi:10.1016/j.jpainsymman.2012.05.011
116. Brant JM, Silbermann M. Global perspectives on palliative care for cancer patients: not all countries are the same. *Curr Oncol Rep*. 2021;23(5):60. doi:10.1007/s11912-021-01044-8
117. Cole AP, Nguyen DD, Meirkhanov A, et al. Association of care at minority-serving vs non–minority-serving hospitals with use of palliative care among racial/ethnic minorities with metastatic cancer in the United States. *JAMA Netw Open*. 2019;2(2):e187633. doi:10.1001/jamanetworkopen.2018.7633
118. Rhodes RL, Ukoha NCE, Williams KA, et al. Understanding underuse of advance care planning among a cohort of African American patients with advanced cancer: formative research that examines gaps in intent to discuss options for care. *Am J Hosp Palliat Med*. 2019;36(12):1057-1062. doi:10.1177/1049909119843276
119. Lee K, Gani F, Canner JK, Johnston FM. Racial disparities in utilization of palliative care among patients admitted with advanced solid organ malignancies. *Am J Hosp Palliat Med*. 2021;38(6):539-546. doi:10.1177/1049909120922779
120. Odonkor CA, Kim G, Erdek M. Global cancer pain management: a systematic review comparing trials in Africa, Europe and North America. *Pain Manag*. 2017;7(4):299-310. doi:10.2217/pmt-2016-0047
121. Silbermann M, Fink RM, Min SJ, et al. Evaluating palliative care needs in middle eastern countries. *J Palliat Med*. 2015;18(1):18-25. doi:10.1089/jpm.2014.0194
122. Matsuyama RK, Balliet W, Ingram K, Lyckholm LJ, Wilson-Genderson M, Smith TJ. Will patients want hospice or palliative care if they do not know what it is? *J Hosp Palliat Nurs*. 2011;13(1):41-46. doi:10.1097/NJH.0b013e3182020520
123. Washington KT, Bickel-Swenson D, Stephens N. Barriers to hospice use among African Americans: a systematic review. *Health Soc Work*. 2008;33(4):267-274. doi:10.1093/hsw/33.4.267
124. Isenberg SR, Bonares M, Kurahashi AM, Algu K, Mahtani R. Race and birth country are associated with discharge location from hospital: a retrospective cohort study of demographic differences for patients receiving inpatient palliative care. *EClinicalMedicine*. 2022;45:101303. doi:10.1016/j.eclinm.2022.101303
125. Escobedo LE, Cervantes L, Havranek E. Barriers in healthcare for Latinx patients with limited English proficiency—a narrative review. *J Gen Intern Med*. 2023;2023:1-8. doi:10.1007/S11606-022-07995-3
126. Masel EK, Schur S, Watzke HH. Life is uncertain. Death is certain. Buddhism and palliative care. *J Pain Symptom Manage*. 2012;44(2):307-312. doi:10.1016/j.jpainsymman.2012.02.018
127. Abazari P, Taleghani F, Hematti S, Ehsani M. Exploring perceptions and preferences of patients, families, physicians, and nurses regarding cancer disclosure: a descriptive qualitative study. *Support Care Cancer*. 2016;24(11):4651-4659. doi:10.1007/s00520-016-3308-x
128. Wigginton B, Farmer K, Kapambwe S, Fitzgerald L, Reeves MM, Lawler SP. Death, contagion and shame: the potential of cancer survivors' advocacy in Zambia. *Health Care Women Int*. 2018;39(5):507-521. doi:10.1080/07399332.2018.1424854

129. Abdullah R, Guo P, Harding R. Preferences and experiences of Muslim patients and their families in Muslim-majority countries for end-of-life care: a systematic review and thematic analysis. *J Pain Symptom Manage*. 2020;60(6):1223-1238.e4. doi:10.1016/j.jpainsymman.2020.06.032

130. Crawley LM, Marshall PA, Lo B, Koenig BA. Strategies for culturally effective end-of-life care. *Ann Intern Med*. 2002;136(9):673. doi:10.7326/0003-4819-136-9-200205070-00010

How to cite this article: Bharmjeet, Das A. Racial disparities in cancer care, an eyeopener for developing better global cancer management strategies. *Cancer Reports*. 2023;e1807. doi:10.1002/cnr2.1807

Resume Classification using Elite Bag-of-Words Approach

Muskan Sharma
Dept. of Information Technology
Delhi Technological University
Delhi, India
muskan_2k19it082@dtu.ac.in

Gargi Choudhary
Dept. of Information Technology
Delhi Technological University
Delhi, India
gargi_2k19it048@dtu.ac.in

Seba Susan
Dept. of Information Technology
Delhi Technological University
Delhi, India
seba_406@yahoo.in

2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT) | 978-1-6654-7467-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSSIT55814.2023.10061036

Abstract—**As technology is advancing day by day, new trends are booming up, like automation, where traditional libraries are being automated to digital libraries. Therefore, instead of manually screening the resumes of all candidates, algorithms and models are being employed to screen resumes in job and career portals. This complete process of mapping resumes to their corresponding job profiles could be efficiently accomplished by making use of various machine learning and Natural Language Processing (NLP) tools. This article utilizes a recently introduced text vectorization technique called Elite bag-of-words for the vectorization of resumes. To implement this method, words in each class are ranked based on their occurring frequency, and then applied maximum entropy partitioning (MEP) to derive the top-ranked significant keywords in each class. These keywords, defined as the Elite keywords, were extracted from each class, and concatenated without redundancy, for predicting the resume type. This research study presents an experimental comparison of the proposed method with existing bag-of-words approaches. This paper implements four vectorization techniques and it is proved that the Elite bag-of-words approach outperforms the other methods for resume classification.**

Keywords – Bag-of-words, Elite keywords, Term frequency, Resume classification.

I. INTRODUCTION

In an increasingly competitive world, the strife to get selected and secure a job is day by day becoming even more difficult and complicated. With the onset of the surge in the number of job profiles versus the number of candidates in the current market scenarios, the sole motive or intention of every other organisation or corporate house is to find the best and most appropriate candidate with the required skill set as per the job profile description. The companies or organizations receive a humongous number of resumes on career portals. Categorizing them on the basis of their job title/posting as per their skills set is quite a tedious task if performed manually; hence automated resume screening is the need of the hour [1]. Since the problem of classification of resumes is a subset of the document classification problem, just like text or sentiment analysis, the tokenization methods of document analysis followed by machine learning can be used for resume classification as well [2].

The efficiency of document categorization depends on how well the machine learning algorithms are able to learn from the given data. Since these algorithms can be applied only on numerical representations, therefore, as a prerequisite, first the resume data having text, paragraphs, sentences should be transformed into feature vectors that can be further applied as input to various machine learning and deep learning models [3].

Now this problem can be solved by making use of popular text vectorization techniques [4] that transform text into numerical data representation called the bag-of-words (BoW) feature representation where the feature columns are the keywords that comprise the input vocabulary. Examples of BoW are the Term Frequency (TF) and Term Frequency - Inverse Document Frequency (TF-IDF) that have been amply used in literature for representing text in various domains and applications [5, 6, 7]. Another alternative to BoW is the use of word embeddings like Word2Vec [8]. Bag-of-words models like TF, TF-IDF in conjunction with machine learning classifiers have been used before for resume classification [9, 10]. Word embeddings have also been used along with convolutional neural networks [11, 12] and recurrent neural networks [13] for resume classification.

This paper aims to determine whether the Elite keywords which is a recently introduced bag-of words model for document representation [14] can serve out to be helpful in the categorization of curriculum vitae as per varied job roles. Elite keywords are defined by the authors in [14] to be the most significant keywords in each class, distinctive in terms of their frequencies of occurrences. The iterative Maximum Entropy Partitioning (MEP) algorithm is used to determine the threshold of the number of significant keywords in a class. The Elite keywords are then concatenated across classes after eliminating redundancy.

The organization of the rest of this paper is as follows. Section II discusses some preliminaries on text pre-processing and vectorization, section III presents the methodology followed, section IV discusses the results, and section V concludes the paper.

II. PRELIMINARIES ON TEXT PRE-PROCESSING AND TEXT VECTORIZATION

Before applying the suitable machine learning models on the proposed dataset comprising of multiple resumes, a series of text pre-processing steps needs to be carried out [15]. The first step is to pre-process the resume data in order to remove insignificant words or noises so that the machine learning techniques could work more efficiently. Below are the steps used to perform text pre-processing.

1) Elimination of irrelevant punctuation marks - The presence of these delimiters can contribute a lot towards adding noise to our dataset which further might affect the accuracy.

2) Deletion of stop-words - Since stop-words do not carry much significance or add meaning to the classification task at hand, so they can be ignored and direct all the focus on the

words that are actually important in our predictions, and vectorize them and proceed further. The stop-words are language specific, that means there exists different sets of stop-words for – English, German, Spanish etc. In case of English, some of the stop-words include- “the”, “a”, “upon”, “above”, “is”, “of”, “below”.

3) The resume cleaning also included removal of URL’s, hashtags, mentions, extra whitespaces, numbers and non-English characters. This is followed by conversion of all uppercase letters into lowercase so that there is no ambiguity in case of two or more occurrences of the same word in both uppercase and lowercase.

4) Stemming and Lemmatization- Stemming refers to the phenomenon of reducing a given word to its stem or root word. The root of the word in this case generally is not a meaningful word. It is implemented by the class Porter Stemmer in the module: nltk.stem.porter present in the NLTK library [15]. Whereas, lemmatization refers to the phenomenon which is quite similar to the process of stemming. There exists only a slight difference; that in case of lemmatization, it reduces the given word to its corresponding root word which comes out to be a meaningful word. For the implementation purpose, an object is created of the class WordNet Lemmatizer inside the nltk.stem module present in the NLTK library. It comes with an added advantage over the previous method of stemming, that the final word representation after reductions is understandable and meaningful. For the resume text, first lemmatization has been performed and then stemming process has been carried out on the lemmatized words.

5) Feature extraction by text vectorization – This is the constitutional model which forms the basis of Natural Language Processing, that underlines the importance of transforming the tokens or words in a document into a numeric representation called the feature vector. In this work, we explore the bag-of-words model for text representation in which the words form feature columns and the rows represent resume samples. Bag-of-words (BoW) model has been successfully used before for the representation of large document classes like 20-Newsgroups [16], named entity recognition [17], sentiment analysis [18], and topic modelling from social media posts [19]. We therefore found the bag-of-words model an apt choice for the vectorization and representation of resume documents in our current study. Some popular bag-of-words models that we used in our experiments are described next.

a) *One-hot encoding*

The one-hot encoding is the most simplistic and conventionally used BoW model. The crux of this model lies in the fact that after we are done with all the pre-processing, we create a dictionary of all the keywords present in the corpus and based on their occurrences within a text document they are mapped with binary output i.e.- either 0 or 1, where 1 denotes that particular keyword is present in the document, and on the other hand 0 denotes the absence of the keyword. Most of the times while applying one-hot encoding, we are left with a sparse matrix having elements- 0 and 1.

b) *Term frequency (TF)*

TF stands for Term Frequency which is the count or frequency of keywords in a document. The underlying principle on which the TF works, is based on the fact that it takes into consideration the number of times a particular keyword k is present in the text document- in our case the document is the resume r.

TF can be represented mathematically as

$$TF\left(k,\ r\right)=count\left(k\right)\middle| r \qquad (1)$$

where, k stands for keyword and r represents the document.

c) *Term frequency – inverse document frequency (TF-IDF)*

The acronym TF-IDF stands for Term Frequency - Inverse Document Frequency, wherein we create a word frequency map or dictionary where each word is mapped to its corresponding frequency, and multiply this frequency by a weight that represents how rare this keyword is across all documents. TF-IDF is a modified version of the original Term Frequency (TF) wherein, in addition to the basic functionalities of the TF an added benefit is there - it aims to focus more on those frequently occurring keywords that do not occur commonly in all documents.

The ability of the TF-IDF to distinguish and emphasize on unique keywords is an added advantage over TF. The usual process for calculating it is divided into two parts. We individually create the Term Frequency (TF) matrix and the Inverse Document Frequency (IDF) matrix. And we then multiply the two matrices. Mathematically, TF-IDF can be represented as

$$TF-IDF\left(k,r\right)=TF\left(k,\ r\right)\times IDF\left(k\right)$$

(2)

Here $IDF(k)$ is the logarithm of the inverse fraction of documents that contain the keyword k. TF-IDF is one of the most popular and reliable BoW models used in NLP, noted for its advantages over TF and one-hot encoding.

III. Resume Classification using Elite Bag-of-Words

One disadvantage of TF and other BoW approaches is that there is no scheme of separating out redundant keywords that may affect the performance of the resume classification. Removing redundant and non-informative keywords or feature columns is the need of the hour. Usually feature selection schemes are additionally used for selecting important keywords [20]. However, feature selection by itself is an unstable method and the set of selected features depends heavily on the training samples [21].

In this paper we explore the use of Elite keywords [14], a recently proposed bag-of-words approach, for extracting significant keywords separately from each resume class. After shortlisting the significant keywords, they are concatenated across classes after removing the redundant keywords. This ensures that class-specific keywords are included in the feature columns. Since resume text is expected to contain keywords specific to a class that may not be as important for the other classes, therefore, the procedure of Elite keyword extraction from each class, and concatenation, will ensure that only significant keywords are selected. The method also returns stable results since the maximum entropy partitioning (MEP) method is used to separate the significant keywords from the non-significant keywords in each resume class.

The methodology followed in this paper is divided-majorly into five steps as shown in the process flow in Fig. 1: a) Data preparation- which involves collecting data from online resources b) Applying data pre-processing techniques c) Converting the text inside the documents into feature vectors d) Training the machine learning models e) Feeding resumes for testing purposes and predicting the resume category using the trained model.

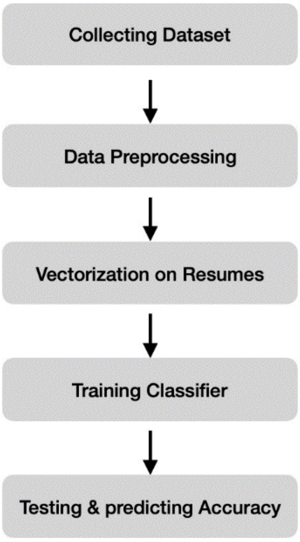


Fig. 1. Process flow

In order to adopt an optimal approach for the purpose of categorization of resumes, we first need to find a minimized subset of most relevant keywords for every category in the dataset.

The Elite keywords extraction was introduced in a recent work [14] as an automated technique for finding subsets of significant keywords from each class using maximum entropy partitioning (MEP). These significant keywords provide ample knowledge about each class and are based on relative term frequencies within each class. The detailed procedure for finding the Elite keywords for resume classification is described next.

1) Creating vectorizers

The first step is similar to that of BoW, in which we need to create a TF matrix for all the words occurring in the resumes of a specific class. This step would give us *C* matrices, each corresponding to the term frequencies of keywords in each category, where *C* is the number of resume classes.

2) Calculating cumulative Frequencies

From the TF matrices obtain the cumulative count of each word for an entire category; this can be achieved by simply adding all the frequencies corresponding to that word for a specific class.

3) Calculating Relative Frequency

For each class, first sum up all the cumulative frequencies and then divide each cumulative frequency with that sum in order to obtain a relative frequency value for each keyword.

4) Applying Maximum Entropy Partitioning

After obtaining the relative term frequencies in each class, we need to sort the frequencies in descending order and then apply the iterative MEP algorithm to partition the sorted relative frequency values into two parts - the upper part is defined as the Elite keywords and the lower part is to be discarded. MEP algorithm involves the computation of the Shannon entropy for the upper and lower parts, and summing up the two entropies. The sum of the probabilities in each of the upper and lower parts should sum up to one prior to the computation of entropy. The partition which gives the maximum sum of entropies is the optimal partition, since at this point, the probability distributions in the two sections approximate uniform distributions.

Mathematically, Shannon entropy is calculated by the equation shown below:

$$E = -\sum_{s} s \log s \tag{3}$$

where, *E* denotes the entropy and *s* stands for the probability values. MEP will be returning the optimal index at which we need to partition the sorted relative probabilities in order to get the subset of the most significant keywords. Therefore, to achieve the MEP index we need to run a loop from 2 to the length of the relative probabilities array, that would partition the array into two groups, and then we calculate the sum of the entropies of both the groups using the formula in (3). For the optimal partition we need to compare entropies at all the indices and hence return the index *i* at which the sum of the two entropies is maximum.

5) Concatenation of obtained Elite keywords

The Elite keywords obtained by maximum entropy partitioning of each class are concatenated across classes after removing the redundant keywords. This final array that we obtained after concatenation is the set of Elite keywords that will be used as tokens to calculate the TF matrix for the training set of the resume dataset.

6) Creating a TF matrix

The TF matrix is obtained by calculating the frequency of the concatenated Elite keywords in each resume document. Then we feed the matrix as input to the classifier and use the trained model for prediction.

IV. Results and Discussions

All experiments are performed on the Kaggle resume dataset available online[1] that has 24 categories such as Accountant, Teacher etc. as shown in Table I. There are 1738 training samples and 744 testing samples in this dataset. The dataset consists of multiple attributes like ID, Resume_html, Resume_str. Resume_html which contains html tags corresponding to each resume, and ID which associates a unique numeric value to each resume, were dropped off during the pre-processing state because both of these hardly carried any significance to our experiments.

We compare the performance of the Elite bag-of-words for resume classification with different bag-of-words approaches found in literature, using one of the most effective classifiers used for text classification – random forest (with Grid search for hyperparameter optimization). Python 3.7 version software is used. All the BoW codes just took a few seconds to execute on a 2.6 GHz Intel PC. We

[1] https://www.kaggle.com/datasets/snehaanbhawal/resume-dataset

have made our Python code for extracting Elite keywords available online[2] for facilitating future research. We have shown the number of Elite keywords extracted from the 24 resume classes in Table I, along with the total number of keywords extracted from each class. It is observed that the application of MEP drastically reduces the number of features as verified from the Elite keywords' column in Table I.

TABLE I. RESUME CATEGORIES AND NUMBER OF KEYWORDS

Resume Categories	Elite keywords	Total Keywords
Accountant	2272	5084
Advocate	2573	5530
Agriculture	1994	4662
Arts	2198	5227
Apparel	2141	5384
Automobile	1750	3227
Aviation	2735	6222
Banking	2481	5396
BPO	1359	2481
Business-Development	2312	5424
Chef	2566	5674
Construction	2395	5514
Consultant	2687	6202
Designer	2305	5770
Digital-Media	1986	5150
Engineering	2692	6310
Finance	2189	5051
Fitness	2344	5575
Healthcare	2676	5948
HR	2027	4538
Information-Technology	2844	6275
Public-Relation	2505	6239
Sales	2226	5090
Teacher	2011	4636

For the visual representation of keywords, we have used the Word cloud in Python which primarily helps us in visualizing the text where the size of a specific word denotes the frequency or significance of the word in the resume. Fig. 2 shows the word cloud obtained using the vanilla BoW model for the "Accountant" class. The Elite keywords derived for the class "Accountant" are shown in the word cloud in Fig. 3. It is observed from the comparison of Fig. 2 and Fig. 3 that that class-specific significant words like *financial* and *company* are given more prominence in the list of Elite keywords for the class "Accountant". On the other hand, common words like *work*, *instruction*, *provide* etc. found in Fig. 2 have been removed in the Elite keyword subset as observed from Fig. 3.

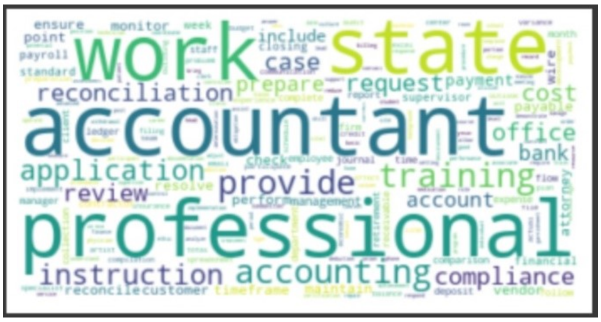


Fig. 2. Keywords detected in the class "Accountant"

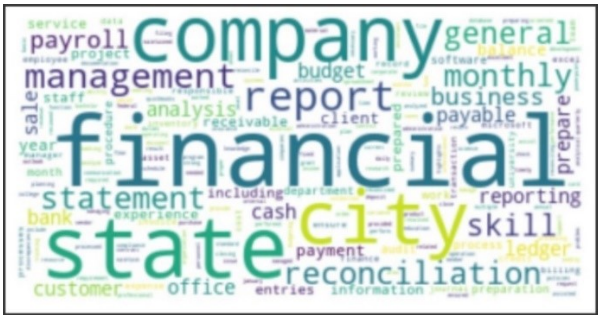


Fig. 3. Elite keywords detected in the class "Accountant"

We train all features on the random forest classifier. Further detailed performances of the BoW models can be analyzed from Table II which shows the results of the random forest classifier on the different bag-of-words representations like One-hot encoding, TF, TF-IDF and Elite keywords. Therefore, upon comparison the best performing model is observed to be the Elite keywords which gave the highest test accuracy of 62.60%.

TABLE II. PERFORMANCE COMPARISON OF DIFFERENT BAG-OF-WORDS APPROACHES FOR RESUME CLASSIFICATION

Method	Test accuracy
One-hot encoding	54.55%
TF-IDF	55.36%
TF	58.98%
Elite keywords	62.60%

The second-best performing model was TF with an accuracy of 58.98%, followed by TF-IDF with an accuracy score of 55.36%. One-hot encoding performed worst giving an accuracy score of 54.55%.

V. CONCLUSION

There is an increasing demand for automated resume screening from job and career portals. Natural language processing is regarded as the most popular means for understanding the content of resume, in order to classify the resume to different job profiles. We explore the most popular NLP tool in our work called the bag-of-words representation in which the text is transformed into a feature vector where the keywords constitute the feature columns. Removing redundant and non-informative keywords or feature columns is the need of the hour. Usually feature selection schemes are used for selecting important

[2] https://github.com/Muskankalonia/Resume-Classification-Using-Elite-Bag-of-Words-Approach

keywords. However, feature selection is an unstable method and varies depending on the training samples. In this paper, we explore Elite keywords, a recently proposed bag-of-words approach, for shortlisting the significant keywords separately for each class. After shortlisting the keywords, they are concatenated across classes after removing the redundant keywords. This ensures that class-specific keywords are included in the feature columns. Since resume document contains keywords specific to a class that may not be as important for the other classes, therefore, the procedure of Elite keyword extraction from each class individually followed by concatenation will ensure that class-specific significant keywords are selected. The method is also stable since the maximum entropy partitioning method is used to automatically separate the significant keywords from the non-significant keywords in each class. We train the features on the random forest classifier using Grid search for hyperparameter optimization. After carefully observing the outcomes of each model, we noted that the Elite keyword subset was by far the most reliable and accurate bag-of-words model for classifying different categories of resumes in the benchmark dataset. Graphical methods for representing the significant keywords in resumes will be explored in our future work. Semantic representations such as the fuzzy bag-of-words approach will also be explored in future. Resume matching and retrieval based on input job profiles is also less explored research that will be the subject of our future study.

REFERENCES

[1] Zaroor, Abeer, Mohammed Maree, and Muath Sabha. "JRC: a job post and resume classification system for online recruitment." In *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 780-787. IEEE, 2017.

[2] Roy, Pradeep Kumar, Sarabjeet Singh Chowdhary, and Rocky Bhatia. "A Machine Learning approach for automation of Resume Recommendation system." *Procedia Computer Science* 167 (2020): 2318-2327.

[3] Swami, Pratibha, and Vibha Pratap. "Resume Classifier and Summarizer." In *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, vol. 1, pp. 220-224. IEEE, 2022.

[4] Wang, Yanzhe. "Basic Methodologies Used in NLP Area." In *2020 IEEE 3rd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*, pp. 505-511. IEEE, 2020.

[5] Hamisu, Muhammad, and Ali Mansour. "Detecting advance fee fraud using nlp bag of word model." In *2020 IEEE 2nd International Conference on Cyberspac (CYBER NIGERIA)*, pp. 94-97. IEEE, 2021.

[6] Junior, Antonio P. Castro, Gabriel A. Wainer, and Wesley P. Calixto. "Weighting construction by bag-of-words with similarity-learning and supervised training for classification models in court text documents." *Applied Soft Computing* (2022): 108987.

[7] Helaskar, Mukund N., and Sheetal S. Sonawane. "Text Classification Using Word Embeddings." In *2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, pp. 1-4. IEEE, 2019.

[8] Sivakumar, Soubraylu, Lakshmi Sarvani Videla, T. Rajesh Kumar, J. Nagaraj, Shilpa Itnal, and D. Haritha. "Review on Word2Vec Word Embedding Neural Net." In *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 282-290. IEEE, 2020.

[9] Ali, Irfan, Nimra Mughal, Zahid Hussain Khand, Javed Ahmed, and Ghulam Mujtaba. "Resume classification system using natural language processing and machine learning techniques." *Mehran University Research Journal Of Engineering & Technology* 41, no. 1 (2022): 65-79.

[10] Duan, Liting, Xiaolin Gui, Mingan Wei, and You Wu. "A Resume Recommendation Algorithm Based on K-means++ and Part-of-speech TF-IDF." In *Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing*, pp. 1-5. 2019.

[11] Mridha, M. F., Rabeya Basri, Muhammad Mostafa Monowar, and Md Abdul Hamid. "A Machine Learning Approach for Screening Individual's Job Profile Using Convolutional Neural Network." In *2021 International Conference on Science & Contemporary Technologies (ICSCT)*, pp. 1-6. IEEE, 2021.

[12] Nasser, Shabna, C. Sreejith, and M. Irshad. "Convolutional neural network with word embedding based approach for resume classification." In *2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR)*, pp. 1-6. IEEE, 2018.

[13] Xu, Qiqiang, Ji Zhang, Youwen Zhu, Bohan Li, Donghai Guan, and Xin Wang. "A block-level RNN model for resume block classification." In *2020 IEEE International Conference on Big Data (Big Data)*, pp. 5855-5857. IEEE, 2020.

[14] Susan, Seba, and Juli Keshari. "Finding significant keywords for document databases by two-phase Maximum Entropy Partitioning." Pattern Recognition Letters 125 (2019): 195-205.

[15] Loper, Edward, and Steven Bird. "NLTK: the Natural Language Toolkit." In *Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics-Volume 1*, pp. 63-70. 2002.

[16] Raj, Anshula, and Seba Susan. "Clustering Analysis for Newsgroup Classification." In *Data Engineering and Intelligent Computing*, pp. 271-279. Springer, Singapore, 2022.

[17] Antony, J. Betina, and G. S. Mahalakshmi. "Named entity recognition for Tamil biomedical documents." In *2014 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2014]*, pp. 1571-1577. IEEE, 2014.

[18] Haddi, Emma, Xiaohui Liu, and Yong Shi. "The role of text pre-processing in sentiment analysis." *Procedia computer science* 17 (2013): 26-32.

[19] Sharma, Anubhav, Seba Susan, Anmol Bansal, and Arjun Choudhry. "Dynamic Topic Modeling of Covid-19 Vaccine-Related Tweets." In *2022 the 5th International Conference on Data Storage and Data Engineering*, pp. 79-84. 2022.

[20] Lee, Lam Hong, Dino Isa, Wou Onn Choo, and Wen Yeen Chue. "High Relevance Keyword Extraction facility for Bayesian text classification on different domains of varying characteristic." *Expert Systems with Applications* 39, no. 1 (2012): 1147-1155.

[21] Singh, Yashpal, and Seba Susan. "SMOTE-LASSO-DeepNet Framework for Cancer Subtyping from Gene Expression Data." In *2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1-6. IEEE, 2022.

[22] Zhao, Rui, and Kezhi Mao. "Fuzzy bag-of-words model for document representation." *IEEE transactions on fuzzy systems* 26, no. 2 (2017): 794-804.

[23] Li, Changmao, Elaine Fisher, Rebecca Thomas, Steve Pittard, Vicki Hertzberg, and Jinho D. Choi. "Competence-Level Prediction and Resume & Job Description Matching Using Context-Aware Transformer Models." In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 8456-8466. 2020.

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-14867-3

Reversible data hiding with high visual quality using pairwise PVO and PEE

Neeraj Kumar[1] · Rajeev Kumar[2] · Aruna Malik[3] · Samayveer Singh[3] · Ki-Hyun Jung[4]

Received: 18 August 2021 / Revised: 16 June 2022 / Accepted: 6 February 2023

Abstract

Pixel-value ordering (PVO) and prediction-error expansion (PEE) are the two most popular strategies of reversible data hiding (RDH) as PVO provides high-fidelity stego-images with decent embedding capacity (EC) and PEE provides high EC with limited distortion. Further, pairwise embedding scheme introduced by Ou et al. again boosts the EC and reduces distortion of both the strategies. However, there has been a dearth of RDH schemes which can optimally utilize both the pairwise PVO and pairwise PEE strategies to provide a least trade-off between EC and visual quality. In this paper, we propound an adaptive RDH (ARDH) scheme which optimally selects the embedding strategy based on image block category. The proposed scheme reads the image in the block-wise manner using a sliding window of 4 × 4 size to get the image block of same size, then divides the block into inner and outer sub-block. The outer sub-block is considered as a reference block for the inner sub-block to determine statistical properties of the inner sub-block using standard deviation. An enhanced pairwise PEE is adopted for

✉ Rajeev Kumar
rajeevkumar@dtu.ac.in

Neeraj Kumar
neeraj.mohiwal@gmail.com

Aruna Malik
malika@nitj.ac.in

Samayveer Singh
samays@nitj.ac.in

Ki-Hyun Jung
kingjung@anu.ac.kr

[1] Department of Electronics and Communication, Jamia Millia Islamia University, New Delhi, India

[2] Department of Computer Science & Engineering, Delhi Technological University, New Delhi, India

[3] Department of Computer Science & Engineering, NIT Jalandhar, Jalandhar, India

[4] Department of Software Convergence, Andong National University, Andong (Gyeongbuk), Republic of Korea

Published online: 28 February 2023

embedding when the standard deviation of outer sub-block's pixels is smaller than a first user-defined threshold. In case the standard deviation is greater than the first threshold but lower than a second user-defined threshold, then pairwise PVO is adopted. Otherwise, the sub-block is skipped without embedding the secret data. As a result, the ARDH scheme utilizes both the PEE and PVO strategies in optimum manner, which in turn provides higher EC and image quality than the most of the existing RDH schemes as validated by experimental results.

Keywords Adaptive RDH · Reversible data hiding · Pairwise PVO · Prediction error expansion · Pixel distribution · Standard deviation

1 Introduction

Internet of Things (IoT) is one of the trending technologies in current technological era. It connects millions of physical devices through the Internet network with high speed and reliable data transfer to implement automatization. Various types of sensors attached [6, 33] to the IoT devices facilitate in the automatization as they collect and share environmental information in digital form such as text, image, audio, video and so on. IoT network deployed in the medical and a defence field is sensitive to the data integrity and protection, and hence digital content is pre-processed before sharing onto the network for the security reasons. Pre-processing is performed using one of the two well-known techniques i.e., encoding and information hiding [20]. Encoding technique converts digital content into mysterious form using the popular public/private key based encoding methods. Mysterious content would be unrecognizable for the intruders and also cannot be decoded without having the corresponding public/private key. However, it may raise a doubt of suspicious activity to outsiders as the contents (though encoded) are visible to everyone in some circumstances, whereas information/data hiding technique reduces such risk by embedding confidential messages covertly in some trivial media known as cover/host media such as text, image, audio, video etc. At some times, reversible data hiding (RDH) technique is adopted when both the cover media and embedding contents are needed in their original form at receiving end. The RDH technique guarantees the reversibility of both the cover media and embedding content at the receiving end.

In early phase, RDH techniques have been proposed based on the lossless compression [5, 16, 17, 27] (which usually creates room inside the cover media by compressing its trivial elements for embedding the secret message), but these techniques provide very limited EC. Thereafter, RDH techniques are advanced into three major directions, which are difference expansion (DE) [1, 38], histogram modification via shifting and expansion (HS) [11, 25, 28, 41], and prediction-error expansion (PEE) [10, 24, 37, 42]. The difference expansion based RDH scheme exploits inter-pixel differences to embed the data. In histogram modification-based approach, histogram of the cover image is expanded into positive and negative directions to the embed data into high frequency bins of the histogram. Histogram modification based RDH schemes usually provide decent image quality with the limited EC. In 2007, Thodi et al. [37] presented a novel high EC with least distortion RDH technique using prediction error expansion (PEE). The PEE strategy makes use of a predictor to predict a reference pixel based on surrounding pixels and then modifies the reference pixel to embed the bits of secret message based on the identified error between the reference pixel and the predicted value. To modify the pixel value, a prediction error histogram (PEH) is plotted and its peak bins are utilized to embed secret bits by expanding them and shifting the others to ensure the reversibility. The PEE seeks data embedding in all pixels and makes only small change in magnitude (+1/−1) per pixel. Thus, it improves both EC and marked-image quality. Thereafter, a number of researches have been carried out over the PEE [2–4, 10, 24, 29, 30, 34, 37, 39, 42] to further improve its performance. When the PEE was matured enough in one-dimensional (1D) scenario, Ou et al. [29] unfolded the two-dimensional (2D) scenario of PEE which is known as pairwise PEE. The pairwise PEE made a slight but impactful change in the 2D mapping of PEH. The 2D PEH mapping is generated with the assumption that a pair of prediction errors have the same correlation as that of adjacent pixels. Different from the conventional PEE, the pairwise PEE expands prediction error pairs such as (0, 0) to three possible pairs (0, 0), (0, 1) and (1, 0) only to embed $\log_2 3$ bits, while discarding expansion from (0, 0) to (1, 1) to minimize distortion but at the cost of EC. However, it additionally exploits prediction error pair (1,1) to embed one-bit data by expanding (1,1) to either (1,1) or (2,2). Thus, the pairwise PEE boosts overall EC and also enhances marked-image quality. The pairwise embedding in PEE scheme is further researched in [2–4, 30] and its review is presented in section 2B of this paper.

In addition to the pairwise PEE, a noteworthy development of PEE was introduced by Li et al. [26]. Li et al. proposed pixel value ordering (PVO) technique which provides high-fidelity stego-images with decent EC. The PVO technique divides the image into uniform sized blocks and then sorts the pixels of each block. Next, the PVO generates prediction errors by taking pixel's intensity difference between the first and second pixels located at extreme ends of the block. Thus, the prediction errors histogram (PEH) generated from the difference of sorted pixels is usually sharper than the one produced by convention PEE which allows embedding in one of the peak bins while shifting the other bins to ensure reversibility. Thus, it provides high-fidelity stego-image, however, it has limited EC as only one of the peak bins is utilized for embedding the secret data. To overcome this limitation, Peng et al. introduced a value-added extension to PVO which is popularly known as Improved PVO (I-PVO) [31]. I-PVO computes the prediction errors in a different way which considers relative locations of pixels inside the original block, so that two peak bins i.e., '0' and '1' can be expanded to increase the EC and at the same time, number of shifted pixels can be reduced to increase the quality of stego-image. After this, a lot of research in the domain of PVO based RDH schemes have been conducted [7–9, 11, 15, 18, 19, 21–23, 32, 40, 41, 43–45] to further enhance the performance. Among them, the work of Kumar et al. [18] has been very fascinating as Kumar et al.'s scheme significantly increased the EC with marginally declining in the marked image quality. Kumar et al.'s scheme makes use of both pairwise PEE and pairwise PVO strategies to embed the secret data in a block, however, the scheme has not been truly effective in generating smooth blocks by exploiting spatial location efficiently. Additionally, the scheme sometimes (though in the worst case) makes modification of ±2 in a pixel which drastically impacts the stego-image quality. The work of [18] is further researched by Kaur et al. [15] to further enhance its embedding performance in 2021. The detailed discussion regarding the working of I-PVO [31] and Kumar et al. [18] is provided in section 2C and 2D, respectively. To overcome these limitations of existing schemes, the proposed work introduces an adaptive RDH scheme using pairwise PVO and PEE. The proposed scheme promotes a different block generation and categorisation method. It makes the use of a sliding window during image scanning for block generation and statistical properties of the block for their categorization. Additionally, the proposed RDH scheme suggests adaptive selection of embedding strategies i.e., pairwise PVO and PEE based on block type/category. The main motivation and contributions of the proposed RDH scheme can be summarized as follows:

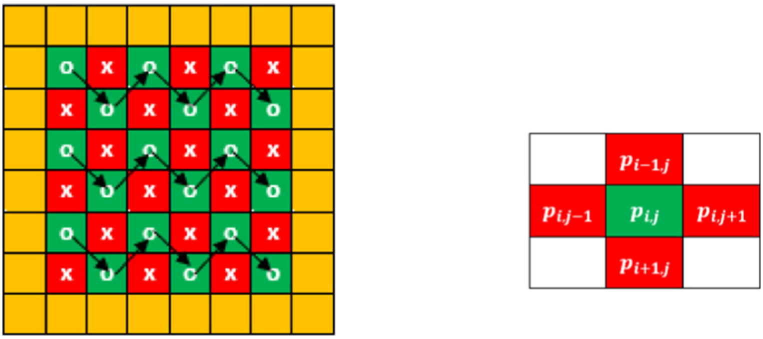
- From the literature review, it has been noted that there is a dearth of RDH schemes which can optimally utilize both the pairwise PVO and pairwise PEE strategies built on their merits. The proposed RDH scheme addresses this gap by introducing a new adaptive RDH scheme.
- The existing RDH methods have not been truly effective in generating smooth blocks by efficiently exploiting spatial correlation. To address this concern, the proposed RDH scheme movably partitions the host image into blocks by scanning the image using the sliding window in a raster scan fashion. This partition allows exploiting of spatial correlation and also allows the extraction of independent rhombus context for embeddable pixels to optimally embed the secret data.
- Before embedding the secret data, the block of the image further partitioned into outer sub-block and inner sub-block, where the pixels of outer sub-block are the peripheral pixels of inner sub-block which are utilized to determine the category of inner-sub-block which can be smooth, moderately complex or highly complex.
- The proposed scheme then makes adaptive selection of embedding strategies i.e., pairwise PEE and pairwise PVO, based on the category of inner sub-block i.e., smooth, moderately complex and highly complex.
- Consequently, the proposed scheme has higher PSNR than all the existing related RDH schemes. More specifically, the average PSNR of the adaptive RDH scheme is 60.39 dB and 56.99 dB on 10 K and 20 K bits in EC.

2 Related works

In this section, some of the existing and related RDH schemes such as PEE based Sachnev et al.'s scheme [34] & Ou et al.'s pairwise scheme [29] along with PVO based Peng et al.'s scheme [31] and Kumar et al.'s scheme [18] are briefly reviewed. Both the PEE and PVO schemes are reviewed as PVO provides high-fidelity images with decent EC and PEE provides high EC with limited distortion. The understanding of both the strategies is important to comprehend the proposed adaptive RDH scheme which makes use of both PVO and PEE strategies. So, Sachnev et al.'s PEE scheme [34] is briefly reviewed followed by Ou et al.'s pairwise scheme [29] in the next sub-section.

2.1 Sachnev et al.'s PEE scheme [34]

In 2009, Sachnev et al. [34] discussed a PEE based RDH scheme using sorting and rhombus predictor. The scheme introduces rhombus predictor which takes all of the closest neighbouring pixels into account to predict the value of reference pixel. For embedding the secret data, the host image is transformed into a chessboard pattern as shown in Fig. 1a and the secret data is embedded in two passes. In the first pass, the "dot" sign pixels (shown in green background of Fig. 1a) are processed followed by the second pass in which "cross" signed pixels (shown in red background of Fig. 1a) are processed. Next, the pixels are sorted based on their local variance before embedding the secret data so that the least complex pixel is used firstly for embedding the secret data. For this, first of all value of reference pixel say ($p_{i,j}$) is



(a) Chessboard transformation of the host image (b) Rhombus context of pixel $p_{i,j}$

Fig. 1 Image scanning and rhombus representation (**a**) Chessboard transformation of the host image (**b**) Rhombus context of pixel $p_{i,\,j}$

predicted using rhombus context which includes pixels $p_{i,\,j\,-\,1}$, $p_{i\,-\,1,\,j}$, $p_{i,\,j\,+\,1}$, and $p_{i\,+\,1,\,j}$ as shown in Fig. 1b. To predict the pixel value, Eq. (1) is given as follows:

$$\widehat{p}_{i,j}=\left\lfloor\frac{1}{4}\left(p_{i,j-1}+p_{i-1,j}+p_{i,j+1}+p_{i+1,j}\right)\right\rfloor \tag{1}$$

where $\widehat{p}_{i,j}$ represents the predicted value of the pixel $p_{i,\,j}$. Next, prediction error ($E_{i,\,j}$) is calculated using Eq. (2) given as follows:

$$E_{i,j}=p_{i,j}-\widehat{p}_{i,j} \tag{2}$$

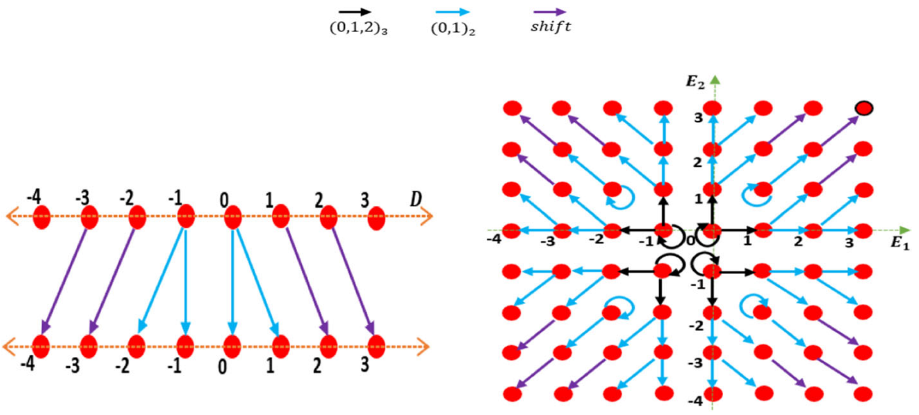
Thus, the prediction-error sequence is derived by processing each "dot" pixel. The sequence is then used to generate the prediction error histogram (PEH) by counting the frequencies of prediction-errors. Generally, the PEH follows a Laplacian-like distribution which peaks at 0 or close to 0. The more the PEH distributes sharply, the less distortion is for embedding the same amount of secret data bits. To embed the secret data, the prediction error is modified by either expanding or shifting the bins as per Eq. (3) which is given below.

$$E'_{i,j}=\begin{cases}2E_{i,j}+b, & if\ E_{i,j}\in(-Thr,Thr),\\ E_{i,j}+Thr, & if\ E_{i,j}\in[Thr,\infty),\\ E_{i,j}-Thr, & if\ E_{i,j}\in[-\infty,-Thr).\end{cases} \tag{3}$$

where Thr is a user-defined threshold to adjust the EC and $b\in\{0,\ 1\}$ is a secret data bit. The complete representation of prediction error modification is shown in Fig. 2a, where the embeddable errors, '0' and '-1' are expanded by the secret data bit value ($b\in\{0,\ 1\}$) and the remaining errors (beyond 0 and $-$ 1) are shifted to ensure reversibility. Finally, the pixel ($p_{i,\,j}$) is modified to embed the secret data using following Eq. (4).

$$p_{i,j}=\widehat{p}_{i,j}+E'_{i,j} \tag{4}$$

This modification is applied on every dot signed pixel to complete the first pass embedding. In the next pass, the same rocess is repeated on the plus signed pixels using the updated image.



(a) Prediction error modification [16]

(b) 2D mapping of prediction errors of pairwise PEE technique [18]

Fig. 2 Prediction error modification/mapping for conventional PEE and pairwise PEE (**a**) Prediction error modification [34] (**b**) 2D mapping of prediction errors of pairwise PEE technique [29]

Thus, the secret data can be embedded into the host image and marked image can be obtained. In the next sub-section, the review of pairwise PEE scheme [18] is provided.

2.2 Ou et al.'s pairwise PEE scheme [29]

In 2013, Ou et al. [29] introduced a novel RDH scheme using pairwise PEE. The pairwise PEE scheme claims and validates that adjacent prediction errors have correlation on the similar lines as that of nearby pixels. By exploiting the correlation, the pair of pixels are used simultaneously to embed the secret data using two-dimensional prediction error histogram modification. Similar to [34], pairwise PEE transforms first the host image into a chessboard pattern as shown in Fig. 1a and does the embedding in two passes. For pairwise embedding, rhombus mean of each pixel is determined to predict its value firstly. The predicted value ($\widehat{p}_{i,j}$) of pixel say ($p_{i,j}$) and the predicted value ($\widehat{p}_{i+1,j+1}$) of pixel say ($p_{i+1,j+1}$) are determined using their surrounding pixels such as $p_{i,j-1}$, $p_{i-1,j}$, $p_{i,j+1}$, and $p_{i+1,j}$ and $p_{i+1,j}$, $p_{i,j+1}$, $p_{i+1,j+2}$, and $p_{i+2,j+1}$ using Eqs. (1) and (5), respectively.

$$\widehat{p}_{i+1,j+1} = \left[\frac{1}{4}\left(p_{i+1,j} + p_{i,j+1} + p_{i+1,j+2} + p_{i+2,j+1}\right)\right] \tag{5}$$

The calculated rhombus means $\widehat{p}_{i,j}$ and $\widehat{p}_{i+1,j+1}$ for the pair of pixels i.e., $p_{i,j}$ and $p_{i+1,j+1}$ are used to determine prediction errors ($E_{i,j}$) and ($E_{i+1,j+1}$), respectively, as follows using Eqs. (2) and (6):

$$E_{i+1,j+1} = p_{i+1,j+1} - \widehat{p}_{i+1,j+1} \tag{6}$$

Then, the pixels $p_{i,j}$ and $p_{i+1,j+1}$ are simultaneously modified to embed the secret data as per the 2D mapping of prediction errors shown in Fig. 2b. Similar to the conventional one-dimensional (1D) mapping, the value of $p_{i,j}$ and $p_{i+1,j+1}$ is also updated maximum by ±1 in 2D PEH mapping. However, it is clear from the Fig. 2 that the pairwise embedding expands prediction error pairs such as (0, 0) to only three possible pairs (0, 0), (0, 1) and (1, 0) to embed $\log_2 3$ bits, while discarding expansion from (0, 0) to (1, 1) to limit the changes to 1 only which minimizes distortion while shrinking EC a bit. However, to compensate the shrinking EC, it additionally exploits prediction error (1,1) to embed one-bit data by expanding (1,1) to either (1,1) or (2,2) based on the bit of secret data which in turn increases the EC.

2.3 I-PVO method [31]

In 2014, Peng et al. [31] introduced the improved pixel value ordering (I-PVO) method which provides high-fidelity marked images with good EC. The I-PVO method is basically an extension of PVO method [26] that is briefly discussed in introduction section. The I-PVO method first partitions the original image into several equal sized blocks (of size $(y \times z)$ pixels) where each block contains m=$(y \times z)$ pixels i.e., $(p_1, \ldots p_m)$, and then sorts the pixels of each block B_i in the ascending order to get $(p_{\pi(1)}, \ldots p_{\pi(m)})$ where $\{1, 2, \ldots, y\} \rightarrow \{1, 2, \ldots, z\}$ is the unique one-to-one mapping satisfying $\pi(i) < \pi(j)$ if $\pi(i) = \pi(j)$ and $i < j$. Here, the index 'i' is dropped to avoid any confusion. The method then calculates two prediction differences (D_{min} and D_{max}) using following equations:

$$D_{min} = X_s – X_t \tag{7}$$

$$D_{max} = X_u – X_v \tag{8}$$

where $s = \ min\ (\pi(1), \pi(2))$ & $t = \ max\ (\pi(1), \pi(2))$, and $u = \ min\ (\pi(m-1), \pi(m))$ & $v = \ max\ (\pi(m-1), \pi(m))$. Thus, the values of D_{min} and D_{max} lie in the range of −255 to 0 if $X_s \leq X_t$ and $X_u \leq X_v$, respectively, otherwise in the range of 1 to 255.

For embedding the secret data, the I-PVO method modifies pixels $p_{\pi(1)}$ & $p_{\pi(m)}$ by subtracting and adding the bit value respectively, when difference value is either '0' or '1', otherwise the pixel values is decreased and increased by 1 respectively, as follows.

$$p'_{\pi(1)} = \begin{cases} p_{\pi(1)} – b_1, & \text{if } D_{min} = 1, \\ p_{\pi(1)} – 1, & \text{if } D_{min} > 1, \\ p_{\pi(1)} – b_1, & \text{if } D_{min} = 0, \\ p_{\pi(1)} – 1, & \text{if } D_{min} < 0. \end{cases} \tag{9}$$

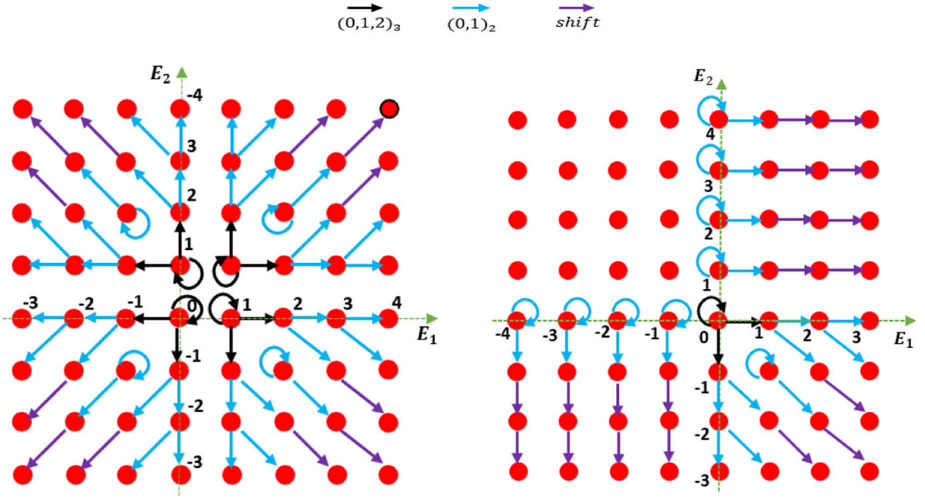
$$p'_{\pi(m)} = \begin{cases} p_{\pi(m)} + b_2, & \text{if } D_{max} = 1, \\ p_{\pi(m)} + 1, & \text{if } D_{max} > 1, \\ p_{\pi(m)} + b_2, & \text{if } D_{max} = 0, \\ p_{\pi(m)} + 1, & \text{if } D_{max} < 0. \end{cases} \tag{10}$$

where $p'_{\pi(1)}$ & $p'_{\pi(m)}$ refer to the modified lowest & highest valued pixels of the block and b_1 & b_2 refer to bits of the secret data, which can be either '0' or '1'. Thus, pixel $p_{\pi(1)}$ is either decreased or remains unchanged and the pixel $p_{\pi(m)}$ is either increased or remains unchanged. In this way, the order of pixels inside the block remains intact and ensures the lossless recovery of cover image after extraction of secret data. The method embeds secret data bits using two most frequent difference values, '0' or '1' which helps in achieving good EC while providing high-fidelity marked image.

2.4 Kumar et al.'s scheme [18]

Kumar et al. [18] recently discussed a high-capacity RDH using enhanced pairwise I-PVO (EPI-PVO). Similar to [29, 34], Kumar et al.'s scheme also transforms the host image into a chessboard pattern as shown in Fig. 1a and does the embedding in two passes. For this, the host image is first partitioned into 1 × 3 blocks (containing only dot pixels in the first pass and plus pixels in the second pass) by traversing in zig-zag order (spanning two rows). Next, the image blocks are categorized into smooth and complex category based on the correlation of rhombus mean of each pixel.

If the image block is a smooth one, the secret data is embedded in two layers. In the first layer, the embedding is done in pairwise manner using I-PVO method based on [43] where the value of minimum valued pixel is either decreased or remains unchanged and the value of maximum valued pixels is either increased or remains unchanged based on the bit value of the secret data. The modification of the prediction errors is done using enhanced 2D mapping mechanism inspired by Ou et al. [28]. The enhanced 2D mapping mechanism basically expands the error pair (0, 0) to only three pairs (0, 0), (0, −1) and (−1, 0), and thus only $\log_2 3$ bits are embedded into this pair. Moreover, the pair (−1, −1) is expanded to itself and (−2, −2) to embed 1-bit data in EPI-PVO, while this pair can be just shifted to (−2, −2) in the conventional pairwise I-PVO. The other pairs like (1,1), (0,1) and (1,0) are also similarly expanded as the pair (0,0). The complete illustrative expansion and shifting mechanism based on 2D mapping is shown in Fig. 3a. Thus, a significant performance improvement in the EC and image quality is seen. In the second layer, the embedding is done in such a way that most of the pixels modified in the first layer can be recovered while also enabling some additional embedding at the same time. For this, a new recovery based pairwise embedding strategy was discussed in which the pixels are arranged in ascending order using their rhombus mean. Next, the secret data is embedded using some defined 2D mapping as shown in Fig. 3b. Further, it additionally embeds the secret data in the medium pixel (ordered based on the mean sequence)



(a) 2D mapping mechanism for EPI-PVO [32]

(b) Pairwise recovery-based embedding strategy [32]

Fig. 3 Prediction Error Modification Mechanism on Kumar et al. [18] (**a**) 2D mapping mechanism for EPI-PVO [18] (**b**) Pairwise recovery-based embedding strategy [18]

using [34]. In the case that the block is a complex one, the secret data is embedded in pixel-wise manner using [29] so that the pixels which may individually belong to smooth region can be used to embed the secret data. Thus, the payload capacity is significantly increased in comparison to existing PVO based methods, however, the image quality is inferior at low embedding capacities.

2.5 Summary

Four RDH schemes are introduced above. Each of them represents pioneer work in the field of reversible data hiding as far as performance is concerned. In addition to aforementioned reviewed works, there has been introduced other noteworthy techniques in the domain of high-fidelity reversible data hiding which are briefly discussed as follows.

Kumar et al. [19] introduced a novel PVO-RDH scheme using a block extension strategy. This scheme basically extends the work of Peng et al. [31] to increase the EC. For this, it basically makes use of PEE strategy when the block is an extremely smooth block by extending the block size. The PEE strategy (with novel predictor) is used to embed the secret data into pixels which belong to extended block. The embedding in the extended block is done in two passes using two different predictors by keeping in mind the original pixel values and updated pixel values. Thus, the number of unused pixels such as the middle pixels in [31] are reduced which increases the EC in turn. In 2021, some new RDH schemes were introduced to further improve the embedding performance [15, 21, 22]. The work discussed in [15] is basically an extension of [18] to limit the modification of a pixel to ±1 while providing the good embedding capacity. LM-PVO method proposed by N. Kumar et al. [22] tries to exploit the image correlation by dividing the image into fixed size blocks and makes use of PVO and PEE schemes for embedding the information. In [21], R. Kumar et al. discussed an enhanced I-PVO method which tries to reserve a bin so that expansion can be done on the both sides for increasing the embedding capacity. However, the quality of the stego-image is deteriorated. Most of the PVO based research works have been surveyed in [8, 12–14, 18, 19, 23, 32, 35, 40, 43–45]. Qu et al. suggested a different way of increasing the EC by proposing a pixel-based PVO (PPVO) technique [32]. The PPVO introduces a sliding window-based image block creation method which populates a greater number of embeddable blocks. The embedding of secret data is done in pixel-by-pixel manner instead of block-by-block manner as in conventional method. In another work proposed by Weng et al. [40], the host image is partitioned into blocks which are categorized into smooth and complex categories based on neighbourhood pixels. Weng et al.'s scheme further sub-categorizes the smooth block into either low or moderately or high correlation blocks based on local complexity. Next, the secret data is adaptively embedded into the smooth blocks based on their sub-category - low, moderately, and high correlation. However, no embedding is done in the case of complex blocks to avoid large distortion due to lack of correlation. Thus, the scheme keeps a good balance between the EC and image quality.

The concept of pairwise embedding in PVO based RDH scheme for improving further the stego-image quality was introduced in [43] which is then extended by He et al. in [8]. The extended work disclosed by He et al. suggests a multi-pass PVO and pairwise embedding scheme [8] that shifting of one error in the pair can be used re-calculate other one based on block type - smooth block or normal block. Thus, the scheme provides additional EC and reduces distortion. To further, enhance the performance of pairwise PVO based RDH schemes, Zhang et al.'s introduced a location-based predictor [44] which takes location of pixels into consideration in addition to the pixel value orders. Zhang et al.'s scheme generates 2D PEH which is regular in shape and suitable for reversible embedding so that adaptive 2D mapping can be automatically applied. Therefore, the scheme successfully increases the marked image quality. However, it has been observed that there is still a dearth of RDH schemes which can exploit high EC property of PEE strategy and high-fidelity property of PVO based strategy. So, a novel adaptive RDH scheme using pairwise PVO and PEE scheme is proposed in this paper. The proposed RDH scheme adaptively selects between the pairwise PVO and PEE for embedding the secret data so that optimal performance can be achieved. The detailed discussion regarding the proposed scheme is presented in the next section.

3 Proposed scheme

In this section, the proposed adaptive RDH scheme using pairwise PVO and PEE is presented. Initially, the working of the proposed scheme is discussed. The adaptive RDH scheme is first introduced in Section 3.1. Next, host image scanning, block generation and block categorization methods are discussed in Section 3.2. Subsequently, implementation details of the proposed scheme are discussed in Section 3.3, which includes embedding algorithm of the proposed scheme. Finally, the exemplary illustration of the proposed scheme is described in Section 3.4.

3.1 Adaptive RDH scheme

The discussion of related works concludes that the PVO based RDH schemes generally provide high-fidelity images with limited EC, while the PEE based RDH schemes provide high EC with decent image quality. However, there has been no or little discussion related to RDH schemes which can utilize merits of both the PVO and PEE strategies. Though Kumar et al.'s RDH scheme makes use of both pairwise PVO and pairwise PEE schemes, yet the adaptive selection of PVO and PEE strategies based on block type is not truly explored. The proposed scheme tries to exploit benefits of both the strategies by their adaptive selection. For this, the proposed adaptive RDH scheme first divides the image block into inner and outer sub-blocks as shown in Fig. 4 and then categorizes inner sub-block into three categories namely smooth, moderately complex and highly complex, based on calculating standard deviation of outer sub-block. If the inner sub-block is a smooth block, then enhanced pairwise PEE is applied for embedding the secret data due to its high embedding efficiency in smooth regions than PVO based strategy. In the case of moderately complex inner sub-block, the enhanced pairwise PVO is applied for embedding the secret data due to its high embedding efficiency in

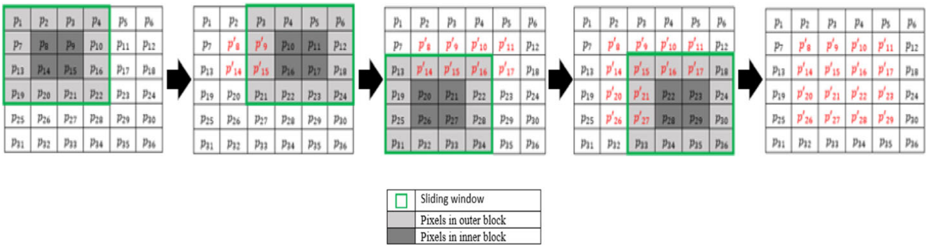


Fig. 4 Image Scanning and block generation strategy of the proposed adaptive RDH scheme

complex regions than PEE based strategy. However, the highly complex inner sub-blocks are left untouched because of high sensitivity towards change in pixel values. Thus, the pros of both the pairwise PVO and PEE are exploited while the cons which affect image quality due to embedding in highly complex blocks are avoided. In the next subsection, the host image scanning, block generation and block categorization strategy are disclosed.

3.2 Host image scanning, block generation and categorization of blocks

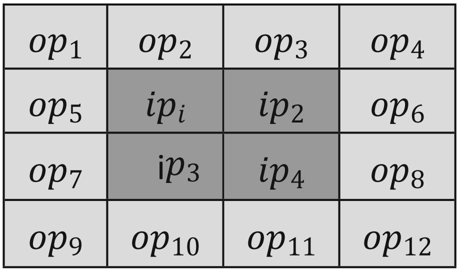
Inspired from [40], the proposed adaptive RDH scheme movably partitions the host image into number of 4 × 4 sized image blocks using a sliding window and embeds the secret message into image blocks as shown in Fig. 5.

Each image block is further partitioned into inner sub-block and outer sub-block as shown in Fig. 5. Border pixels N (shown in light gray in Fig. 5) of the block/window (op_1, op_2, op_3, op_4, op_5, op_6, op_7, op_8, op_9, op_{10}, op_{11} *and* op_{12}) are considered to form an outer sub-block and the remaining four pixels M (shown in dark gray in Fig. 5) of the block/window (ip_1, ip_2, ip_3, *and* ip_4) are considered to form an inner sub-block. Since the outer sub-block includes all surrounding pixels, the local context of inner sub-block pixels can be the best represented by outer sub-block pixels. So, it is assumed that the pixel distribution of outer sub-block can describe well the pixel distribution of inner sub-block. To verify the assumption, an experimental analysis is performed using histogram plots on Lena test image. For the analysis, two histograms are plotted, one of the histogram is plotted based on differences of mean of all the outer sub-block's pixels with respect to their corresponding mean of all inner sub-block's pixels, and the second histogram is plotted based on differences of standard deviation of all outer sub-block's pixels with respect to their corresponding standard deviation of all inner sub-block's pixels as shown in Fig. 5. The mean (μ_{op}) and standard deviation (σ_{op}) of outer sub-block are calculated using equations Eqs. (11) and (12), respectively.

$$\mu_{op} = \frac{\sum op_i}{N} \tag{11}$$

$$\sigma_{op} = \sqrt{\frac{\sum \left(op_i - \mu_{op}\right)^2}{N}} \tag{12}$$

where op_i refers to i^{th} pixel of outer sub-block. Similarly, the mean (μ_{ip}) & standard deviation (σ_{ip}) of inner sub-block are calculated using Eqs. (13) and (14), respectively.



op_1	op_2	op_3	op_4
op_5	ip_i	ip_2	op_6
op_7	ip_3	ip_4	op_8
op_9	op_{10}	op_{11}	op_{12}

Fig. 5 Image block for illustrative purpose

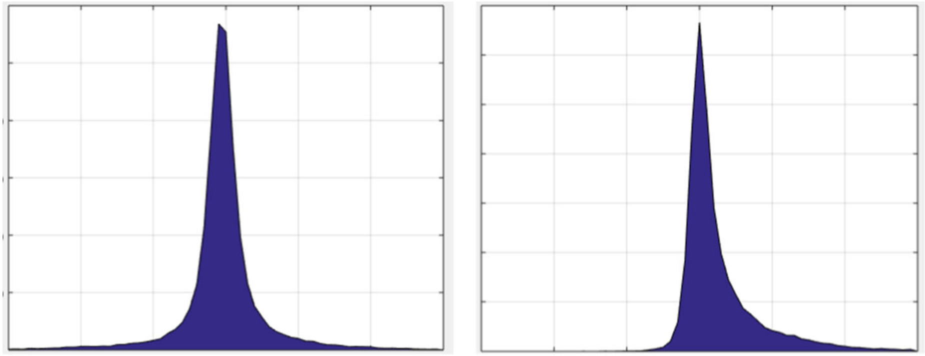
$$\mu_{ip} = \frac{\sum ip_i}{M} \tag{13}$$

$$\sigma_{ip} = \sqrt{\frac{\sum\left(ip_i - \mu_{ip}\right)^2}{M}} \tag{14}$$

where, ip_i refers to i^{th} pixel of inner sub-block. The peaks in both the plotted histograms are close to zero. This peak trend clearly suggest that the inner sub-blocks generally follow the characterstics of outer sub-blocks and thus the pixels of the outer-sub blocks can be used to determine the local context of inner sub-blocks Fig. 6.

Based on the determined pixel distribution of the outer sub-block, the inner sub-block is classified into either of the smooth, moderately complex or highly complex category. If the inner sub-block is a smooth one then pairwise PEE scheme [29] is applied. Since the inner sub-block has four pixels (as shown dark grey of Fig. 7a), two pairs can be formed (by considering diagonal pixels) from the inner sub-block. Then, pairwise PEE is applied in pairwise manner means embedding in the first pair (Fig. 7b) followed by embedding in the second pair (Fig. 7c) using 2D enhanced PEH shown in (Fig. 2b). In this way, both pairs are utilized for data embedding purpose. The implementation details of pairwise embedding are presented in sub-section 2B. Different from the conventional one-dimensional (1D), the pairwise embedding expands prediction error pairs such as (0, 0) to only three possible pairs (0, 0), (0, 1) and (1, 0) to embedded $\log_2 3$ bits, while discarding expansion from (0, 0) to (1, 1) to limit the changes to 1 only which minimizes distortion while shrinking EC a bit. However, it additionally exploits prediction error (1,1) to embed one-bit data by expanding (1,1) to either (1,1) or (2,2) based on the bit of secret data which in turn increases the EC.

In the case that the inner sub-block is a moderately complex one then pairwise PVO strategy is applied because PVO rearranges the pixels of the block in a sorted order that essentially turns the moderately complex sub-block into a smooth one and then embeds the



(a) Histogram of differences calculated using $\mu_{op} - \mu_{ip}$ corresponding to all generated image blocks

(b) Histogram of differences calculated using $\sigma_{op} - \sigma_{ip}$ corresponding to all generated image blocks

Fig. 6 Histogram plots for Lena Image to showcase the sharpness of peak (**a**) Histogram of differences calculated using $\mu_{op} - \mu_{ip}$ corresponding to all generated image blocks (**b**) Histogram of differences calculated using $\sigma_{op} - \sigma_{ip}$ corresponding to all generated image blocks

secret data. Before embedding, the pixels are sorted in ascending order to get sorted sequence ($P_{\pi(1)}$, $P_{\pi(2)}$, $P_{\pi(3)}$ $P_{\pi(4)}$), where $\{1, 2\} \rightarrow \{1, 2\}$ is the unique one-to-one mapping satisfying $\pi(i) < \pi(j)$ for $\pi(i) = \pi(j)$ and $i < j$. Next, two prediction differences D_{min} and D_{max} are computed using Eqs. (7) and (8) respectively. Now, the pixels of the inner sub-block are modified based on 2D mapping defined in Figs. 3a to embed the secret data bits. Thus, the high-fidelity image sub-blocks with decent amount of hidden secret data can be obtained even in the case of moderately complex sub-blocks. However, no change is carried out in the sub-block when the inner sub-block is highly complex sub-block so that high damages to the stego-image quality can be avoided.

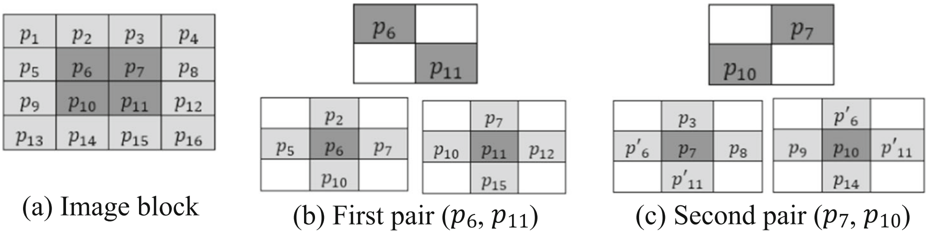
3.3 Detailed implementation

To recover the host image blindly and extract the hidden message, some auxiliary information is also embedded into the host image along with the secret data. This auxiliary information is embedded in the $16 + \log_2 L_{(CLM)} + L_{(CLM)} + \log_2 L_{(C)} + L_{(S)}$. LSBs of the border pixels are not incorporated in the actual data embedding process. The original LSBs are recorded in a binary sequence $S_{(LSB)}$. The details regarding this additional information are given below:

- *Thresholds (thr_1 & thr_2):* The user-defined thresholds thr_1 and thr_2 are required for determining category of inner sub-block based on calculation of standard deviation of outer sub-block. Therefore, a total of 16 bits are required for two threshold values.
- *Location Map:* To avoid the problem caused by underflow/overflow, a location map (LM) is constructed and embedded into the cover image. The LM is a binary sequence where each entry is dedicated to an individual block. For each inner sub-block say I_i, the corresponding entry in the LM is set as follows:

$$LM(i) = \begin{cases} 1 & if \max(I_i) = 255 \,\|\min(I_i) = 0 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

Then, the LM is compressed using arithmetic encoding so that the compressed location map (CLM) can be embedded without affecting the true EC much. If $L_{(CLM)}$ represent the length of the CLM, then its upper limit for a $P \times Q$ size image is calculated as $L_{(CLM)}$= $\left\lceil \log_2 \frac{(P-1)\times(Q-1)}{4} \right\rceil$. For example, if the image of size 512 × 512 pixels, the $L_{(CLM)}$ will be 16 bits.



(a) Image block

(b) First pair (p_6, p_{11})

(c) Second pair (p_7, p_{10})

Fig. 7 Illustration of selection of pixel pairs and determination of rhombus context (**a**) Image block (**b**) First pair (p_6, p_{11}) (**c**) Second pair (p_7, p_{10})

- *Index of last embedded block:* The coordinates of last block used for embedding are also used as auxiliary information. If the image of size 512 × 512 pixels, then 18 bits will be required to define the coordinates.
- Finally, the original LSBs of the border pixels denoted by $S_{(LSB)}$ are embedded after the secret payload in the next available block.

Next, the complete data embedding algorithm of the proposed adaptive RDH scheme is detailed as follows.

Algorithm: Data embedding

```
Inputs: Input Image Block $B_i = \{p_1, \ldots, p_{16}\}$, $thr_1$and $thr_2$: two user-defined thresholds (where $thr_1 < thr_2$), S: secret data in binary form
Output: Stego-block $B_i'$
Step 1: Divide block $(B_i)$ into inner sub-block and outer sub-block where inner sub-block $I_i = \{p_6, p_7, p_{10}, p_{11}\}$ and outer sub-block $O_i = \{p_1, p_2, p_3, p_4, p_5, p_8, p_9, p_{12}, p_{13}, p_{14}, p_{15}, p_{16}\}$
Step 2: Construct location map $(LM)$ for the block $(B_i)$ based on the pixel values of the inner sub-block $I_i$ and Eq. (15) so that the problem of overflow and underflow is avoided.
Step 3: Compute standard deviation $\sigma_i$ of outer sub-block $O_i$ using Eq. (14).
Step 4: If $\sigma_i <= thr_1$, means the inner sub-block $(I_i)$ is a smooth sub-block. In case of smooth inner sub-block,
    • Constitute two pairs of pixels where the first pair has pixels $\{p_6 \ and \ p_{11}\}$ and second pair has $\{ p_7 \ and \ p_{10}\}$ pixels.
    • Apply pairwise PEE [18] (defined in Section 2B) in pairwise manner (means embedding in the first pair followed by embedding in the second pair) to embed secret data into both the pixel pairs.
    • Output the stego sub-block $B_i'$ with modified pixel pairs of inner sub-block $I'_i = \{p_6', p_7', p_{10}', p_{11}'\}$ without modifying other pixels of the block.
Step 5: Else if $\sigma_i > thr_1$ & $\sigma_i <= thr_2$ means the inner sub-block $(I_i)$ is a moderately complex sub-block. In case of moderately complex inner sub-block,
    • Sort the pixels $\{p_6, p_7, p_{10}, p_{11}\}$ of moderately complex inner sub-block in ascending order to get sorted sequence $\{p_{\pi(1)}, p_{\pi(2)}, \ p_{\pi(3)}, \ p_{\pi(4)}\}$.
    • Calculate prediction errors ($D_{min}$ and $D_{max}$) using equations (7) and (8), respectively.
    • Perform data embedding according to pairwise PVO mappings shown in Fig. 5 and obtain the modified pixel pairs of lowest valued pixel and highest valued pixel using (9), (10) and (12) and (13).
    • Output the stego sub-block $B_i'$ with modified pixel pairs of Inner sub-block $I'_i = \{p_6', p_7', p_{10}', p_{11}'\}$ without modifying other pixels.
Step 6: Else $\sigma_i > thr_2$ means the inner sub-block $(I_i)$ is a highly complex sub-block. In case of highly complex sub-block, Skip the sub-block $B_i$ to avoid large distortion.
```

The extraction process is just the inverse of data embedding. The auxiliary information is extracted first, and then the embedded secret data is extracted and the cover pixels are recovered. The details are omitted for simplicity.

3.4 Exemplary illustration

To make the working of the proposed adaptive RDH scheme clearer, an exemplary illustration is presented. In the example, an input image of size 4 × 6 pixels as shown in Fig. 8 and a random sequence of secret data bits are taken as S = 1001. Additionally, two user defined thresholds are taken as i.e., $thr_1 = 5$ and $thr_2 = 10$. Firstly, the input image is scanned using the sliding window of size 4 × 4 (shown in green box of Fig. 8) to get the first image block,

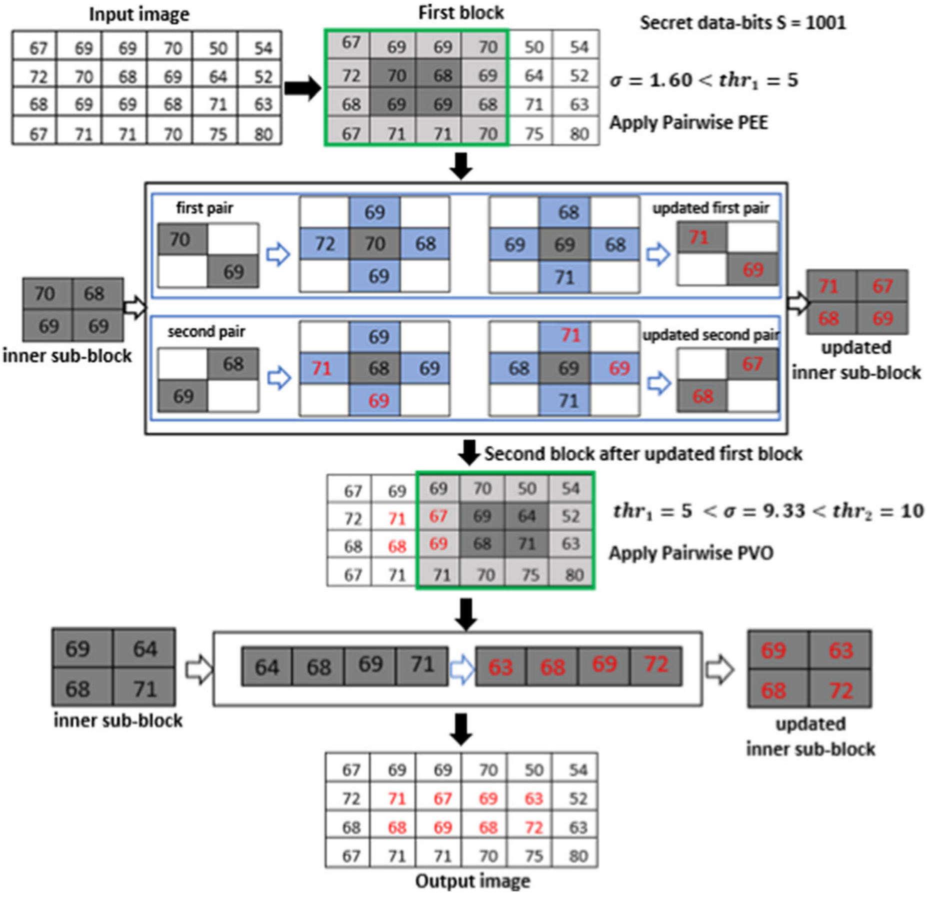


Fig. 8 An illustrative example of the proposed scheme

which is further partitioned into inner sub-block and outer sub-block, where pixels of outer sub-block are {67, 69, 69, 70, 72, 69, 68, 68, 67, 71,71,70} and pixels of inner sub-block are {70, 69, 69, 68}. Next, the standard deviation (σ) of pixels of outer block is calculated using Eq. (14). The calculated σ is 1.60, which is less than the first threshold $thr_1 = 5$, which means the sub-block is a smooth sub-block. In case of smooth sub-block, the pairwise PEE is applied for embedding the secret data in inner sub-block. To apply the pairwise PEE in the inner sub-block, two pairs are formed using the diagonal pixels of inner sub-block. The first pair of diagonal pixels is (70,69) and another pair of diagonal pixels is (68, 69). The embedding of secret data bits in the inner sub-block is done in pairwise manner means embedding in the first pixel pair followed by embedding in the second one.

For data embedding in the first pair, value of the first and second pixels in the first pair are predicted using rhombus context ({72,69,69,68} and {71,69,68,68}) respectively. Thus, the value of the predicted pair calculated using Eqs. (1) and (5) is (70, 69). Next, the prediction errors for the pixel pair are calculated using Eqs. (2) and (6). The pair of the calculated prediction error is (0, 0). Finally, the secret data is embedded into the pixel pair using 2D mapping which shown in Fig. 2b. Thus, the modified pixel pair after embedding two bits of secret message *S*, '10' are (71, 69). Now, the embedding in the second pair, (68, 69) is done in similar manner as the embedding was done in the first pair, but considering the updated block. In case of second pair, the pixel values for the second pair (68, 69) are predicted from rhombus context ({71,69,69,69}, {71,71,68,69}) respectively. Thus, the predicted pair calculated using Eqs. (1) and (5) is (70, 70) which implies that the prediction error pair will be (−2, −1) as per the Eq. (2) and (6). Now to embed the secret data as per the 2D mapping of Fig. 2b implies that no secret data bits can be embedded in second pixel pair and only the pixel values are shifted to (67,68) to ensure the reversibility. Thus, we get the final updated first block.

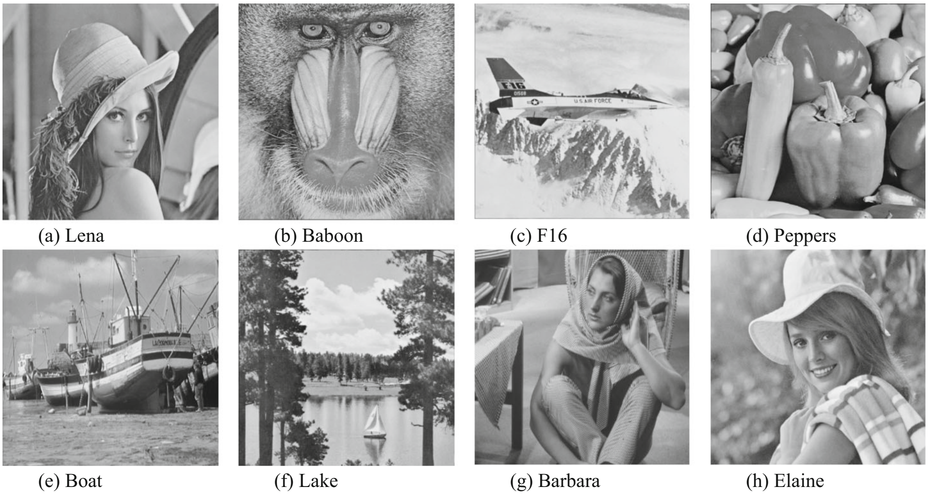
After processing the first block for data embedding, the next block is scanned by sliding the window two columns in forward direction. The obtained block is also partitioned into two sub-blocks - inner and outer sub-block - similar to previous block. Pixels which constitute the outer sub-block are {69, 70, 50, 54, 67, 52, 69, 63, 71,70, 75, 80} and the pixels of inner sub-block are {69, 71, 64, 68}. Now, standard deviation σ of the outer sub-block is calculated which is 9.33 as per the Eq. (14). The calculated standard deviation σ is greater than the first threshold $thr_1 = 5$ but less than the second threshold, thr_2= 10, which means the second sub-block is a moderately complex sub-block. In the case of moderately complex sub-block, the embedding is done using pairwise PVO by sorting the inner sub-block's pixels to get the sorted sequence as follows (64, 68, 69, 71). Next, a pair of prediction error from the first and second minimum valued pixels, and first and second maximum valued pixels is calculated as per the Eq. (7) and (8), respectively. Thus, the obtained error pair is (4, 1). Finally, the embedding of secret data is done using 2D mapping defined in Fig. 3a. As per the Fig. 3a, only one bit of the secret data, '0' can be embedded and the final updated sorted pixel set would be as (63, 68, 69, 72) for the second block which is correspondingly shown in Fig. 8. This way, the embedding in both the smooth and moderately complex blocks (except highly complex block) is illustratively presented. In case of highly complex block, the embedding is not performed so we have omitted the same from the exemplary discussion.

4 Experimental results and discussions

In this section, performance of proposed scheme is demonstrated using experimental results. The performance is measured using the two most popular parameters i.e., embedding capacity (EC) and peak-signal-to-noise ratio (PSNR), where EC refers to the number of secret bits embedded in the cover image and the Peak-signal-to-noise Ratio (PSNR) represents marked-image quality which is defined in terms of decibels (dB) and calculated as follows:

$$PSNR = 10log_{10}\left(\frac{255^2}{MSE}\right) \tag{16}$$

Here, MSE is the 'Mean Squared Error' i.e. the mean of the sum of squared differences between the original cover image and the marked image. The experiments are performed on eight standard grayscale images, each of size 512 × 512 pixels including Lena, Baboon, F16, Peppers, Boat, Lake, Barbara, and Elaine, which are downloaded from the USC-SIPI dataset [36]. These images are shown in Fig. 9. The proposed adaptive RDH scheme is implemented in MATLAB and the secret data is generated using pseudo-random number generators for the experimental purpose.


(a) Lena (b) Baboon (c) F16 (d) Peppers

(e) Boat (f) Lake (g) Barbara (h) Elaine

Fig. 9 Cover images of size 512 × 512 pixels used for experimental tests

The results of the proposed adaptive RDH scheme are comparatively analyzed against some of the closely related and latest state-of-the-art methods which include [7–9, 18, 29, 31, 34, 40, 43, 44].

The objective of the proposed adaptive RDH scheme is to achieve high EC along with good image quality. So, the techniques (based on both PEE and PVO strategies) which excels in both aspects, EC & visual quality are selected for fair comparative analysis. These techniques include Sachnev et al.'s method [34] which introduced the rhombus prediction based two pass embedding strategy and is pioneer work in the domain of PEE, Ou et al.'s Pairwise PEE method [29] which is the very first pairwise scheme, Peng et al.'s method [31] which introduced I-PVO technique and has been the pillar for advancements in the PVO domain, Weng et al.'s method [40] which disclosed adaptive PVO based on image block category, adaptive pairwise PVO method [43] which enhanced and adaptive pairwise embedding in PVO strategy, Multi-pass PVO and pairwise PEE method [8] which improves EC in PVO block in two-pass with differently applying embedding scheme for smooth block and normal block, and location dependent PVO method [44] which considers location of pixels along with PVO, and Kumar et al.'s EPI-PVO method [18] which utilizes both pairwise PVO and PEE methods for same block using two-pass embedding. The proposed scheme gets idea of sliding window-based image block creation from Weng et al.'s method. Further, the optimal use of both pairwise PVO and PEE strategies is inspired from Kumar et al.'s scheme. So, incorporation of these schemes along with others which are directed to achieve either higher EC or higher image quality or both in the comparative analysis, have been the need of the hour. Hence, the choice of these methods is practically reasonable to corroborate the performance of the proposed scheme.

The experimental results for the EC (in bits) versus image quality in terms of PSNR (dB) tradeoffs are presented in Fig. 10. These results have been taken by adjusting the user-defined thresholds i.e. thr_1 & thr_2 to suitable values for achieving optimal performance. The values of thresholds are iteratively decided (by user) to provide optimal performance considering his/her need and image characteristics. Initially, the value of thresholds is set to 1 and then it is incremented according to the capacity requirements. It has been observed that the iterative approach for determining a suitable value of thresholds plays a significant role in maintaining the high quality

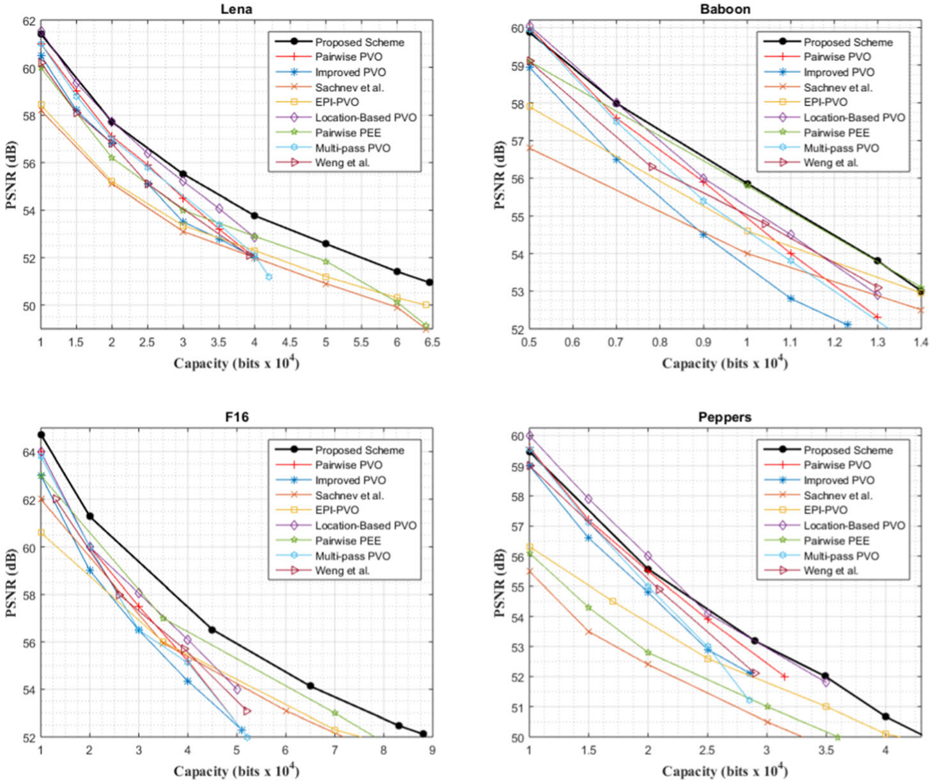


Fig. 10 EC versus PSNR of proposed method and existing methods [8, 13, 18, 29, 31, 34, 40, 43, 44]

of marked-images. It is evident from the presented experimental results that the proposed adaptive RDH scheme has superior embedding performance as it offers the best image quality irrespective of the EC. The stego-images after embedding the secret data are shown in Fig. 11.

For extensive analysis of PSNR at lower EC, the results are provided into Tables 1 & 2 for an EC of 10,000 bits and 20,000 bits, respectively. In Table 2, the experimental results for the baboon image are not mentioned as many methods don't provide 20,000 bits embedding capacity. Additionally, the PSNR results for existing methods [7, 9] for the images baboon and F16 are not mentioned as these schemes have not been tested by their respective authors on the baboon and F16 test images. This analysis is done to clearly highlight the superiority of the proposed scheme in the high-fidelity RDH domain as existing PVO based methods have dominating performance at lower embedding capacities. Form the tables, it is evident that the proposed scheme outperforms the existing methods in terms of visual quality for all test images. Among all test images, the highest PSNR obtained by the proposed scheme is for Airplane which possesses a smooth texture. The proposed scheme also does quite well for a complex image such as Baboon with PSNR of 55.86 dB at EC of 10,000 bits, which is again better than the existing methods. The main reason behind this superior performance is the adaptive selection of PVO and PEE schemes so that benefits of both the scheme can be reaped. Further, exploitation of the pairwise embedding helps in optimally utilizing the prevalent correlation among the spatially correlated prediction errors. Thus, we can conclude that the proposed scheme has the highest PSNR for corresponding EC.

To validate the visual robustness of the proposed scheme, the image intensity histograms are plotted as shown in Fig. 12. The left side histograms correspond to the original test images

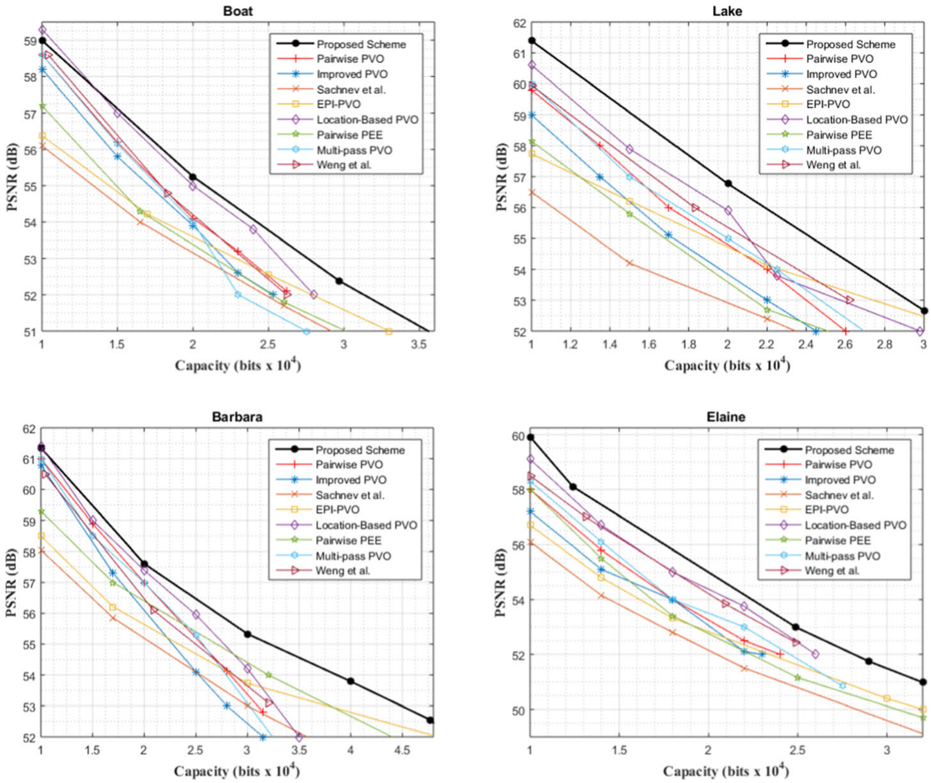
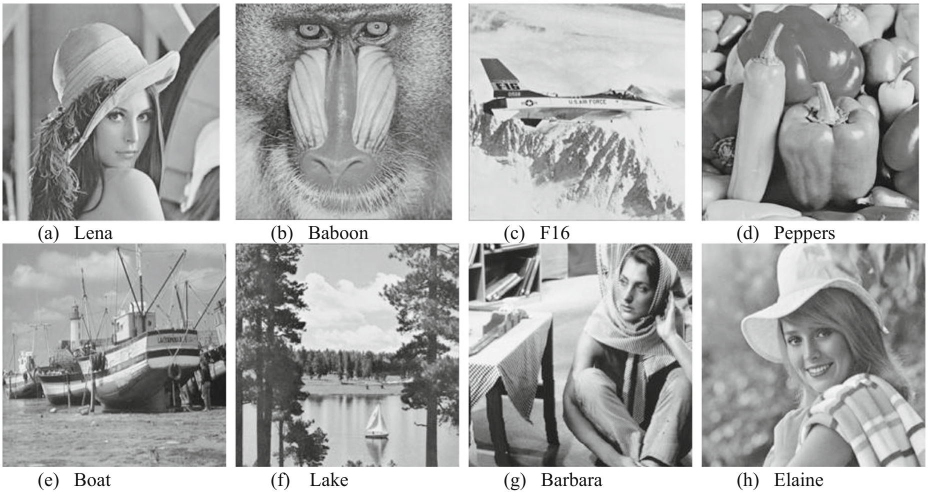


Fig. 10 (continued)


(a) Lena (b) Baboon (c) F16 (d) Peppers

(e) Boat (f) Lake (g) Barbara (h) Elaine

Fig. 11 Stego-images of size 512 × 512 pixels generated after experimental tests

Table 1 PSNR values of the proposed scheme and [7–9, 18, 29, 31, 34, 40, 43, 44] at an EC of 10,000 bits

Images/ Methods	Proposed Method	Multiple Pairwise PEE and Two-Layer Embedding (He et al.) [9]	Dual Pairwise Prediction-Error Expansion [7]	Enhanced pairwise PVO (Kumar et al.)	Multi-pass PVO based pairwise PEE	PVO based adaptive pairwise embedding (R. Ni.)	Location Dependent PVO	Weng et al.	Improved PVO (Peng)	Pairwise PEE (Ou 2013)
Lena	61.41	61.15	61.24	58.43	61.05	61.09	61.53	60.20	60.49	59.77
Baboon	55.86	–	–	54.5	54.65	54.90	55.19	56.30	53.58	55.23
F16	64.73	–	–	60.6	63.60	63.90	64.13	63.00	62.98	63.78
Peppers	59.47	59.75	59.81	56.3	59.56	59.63	59.97	59.00	58.88	56.25
Boat	58.99	59.04	59.14	56.4	58.70	58.66	59.32	58.40	58.27	57.56
Lake	61.39	60.62	60.65	57.7	60.01	59.74	60.76	59.95	58.81	58.67
Barbara	61.35	61.10	61.17	58.2	61.05	61.04	61.35	60.00	60.49	59.49
Elaine	59.92	58.86	58.90	56.7	58.30	58.07	59.18	58.50	57.28	58.10
Average	**60.39**	**60.07**	**60.15**	**57.35**	**59.61**	**59.62**	**60.17**	**59.41**	**58.84**	**58.60**

Table 2 PSNR values of the proposed scheme and [7–9, 18, 29, 31, 34, 40, 43, 44] at an EC of 20,000 bits

Images/ Methods	Proposed Method	Multiple Pairwise PEE and Two-Layer Embedding (He et al.)	Dual Pairwise Prediction-Error Expansion	Enhanced pairwise PVO (Kumar et al.)	Multi-pass PVO based pairwise PEE	PVO based adaptive pairwise embedding (R. Ni.)	Location Dependent PVO	Weng et al.	Improved PVO (Peng)	Pairwise PEE (Ou 2013)
Lena	57.71	57.45	57.51	55.00	57.20	57.30	57.74	56.80	56.57	56.29
F16	61.28	–	–	59.00	59.90	60.01	60.33	60.00	59.08	60.16
Peppers	55.57	55.89	56.00	53.80	55.60	55.40	55.98	55.30	54.78	52.86
Boat	55.25	54.84	55.00	53.50	54.30	54.20	54.97	54.25	53.84	53.37
Lake	56.77	55.74	55.85	55.00	55.10	54.80	55.63	57.00	53.54	53.76
Barbara	57.61	57.38	57.41	55.60	57.05	57.10	57.35	56.50	56.20	56.25
Elaine	54.80	54.06	54.18	53.00	53.70	53.40	54.33	54.08	52.65	52.92
Average	**56.99**	**55.88**	**55.99**	**54.98**	**56.12**	**56.03**	**56.61**	**56.27**	**55.23**	**55.08**

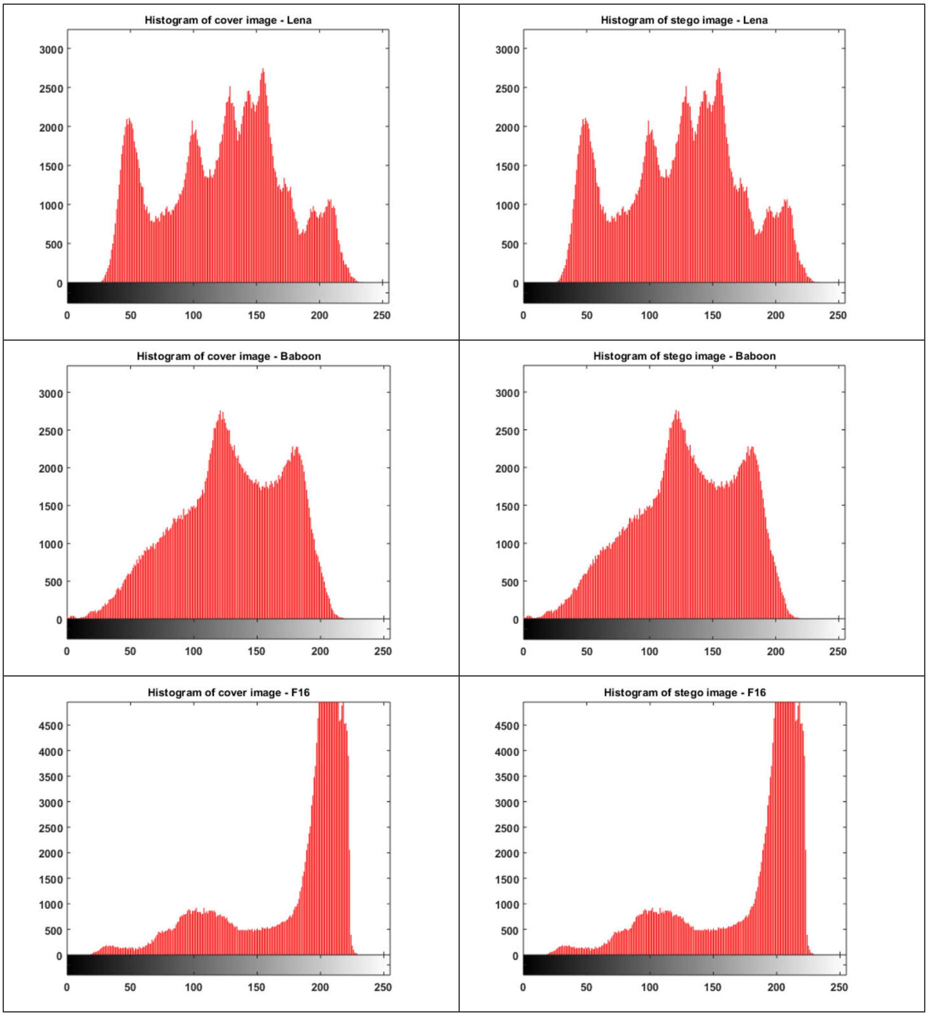


Fig. 12 Image histograms for all the test images to showcase the similarity of cover-images and stego-images

and the right-side ones are plotted considering the marked image which has the hidden of random secret data of 15,000 bits. From the naked eyes, the right-side histograms look copy of the corresponding histogram of left side, means there is no visible mark of the presence of data in the stego-images. So, these observations provide enough support to the fact that the proposed scheme offers higher embedding quality along with high EC without any perceptible change.

5 Conclusion

In this paper, an adaptive RDH scheme using pairwise PVO and PEE has been introduced. The proposed RDH scheme partitioned movably the host image into blocks. The

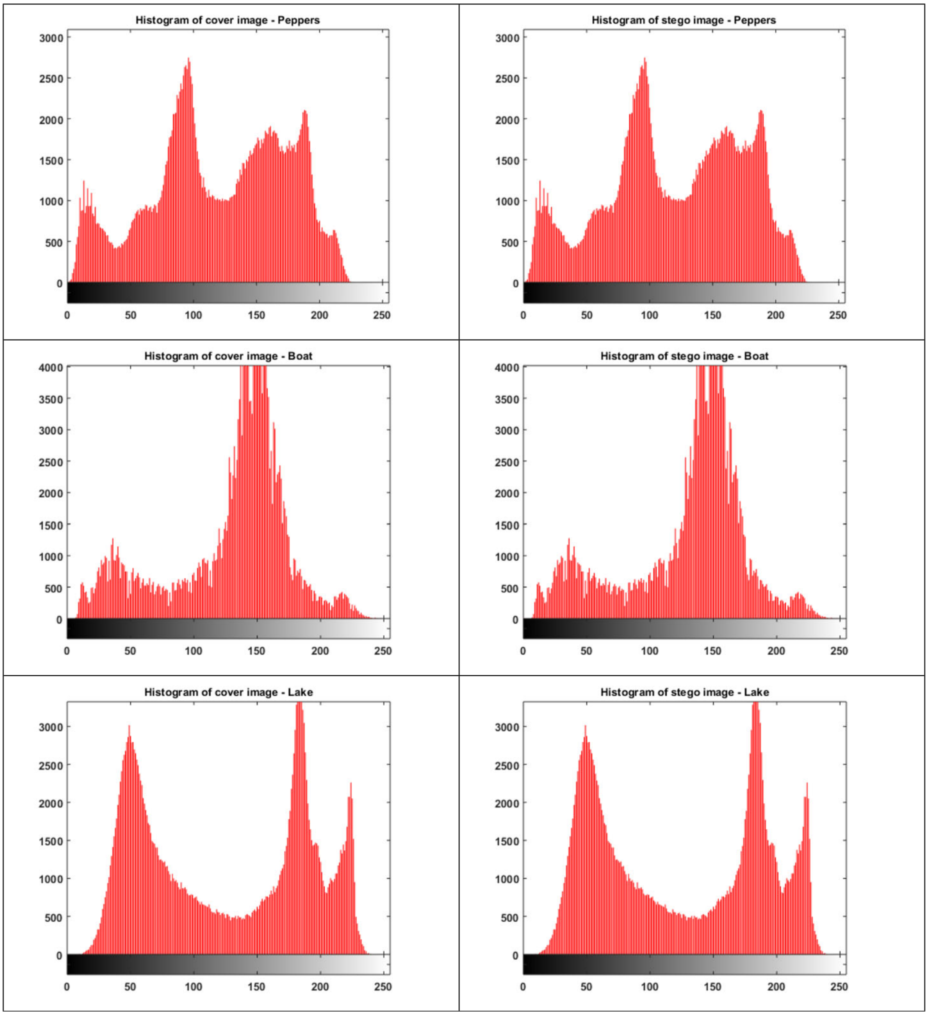


Fig. 12 (continued)

partitioning has enabled the data hider to exploit the spatial correlation and also allowed the extraction of independent rhombus context for embeddable pixels to embed optimally the secret data. Further, the adaptive selection of embedding strategy was done so that characteristics of both PVO and PEE strategies could be exploited. Consequently, the proposed scheme has resulted with higher PSNR than all of the existing related RDH schemes. More specifically, the average PSNR of the adaptive RDH scheme was 60.39 dB and 56.99 dB on 10 K and 20 K bits embedding capacity which was higher than the existing RDH schemes. Further, the experimental histograms provided in the context of visual robustness have proved that there is no visible mark of the presence of data in the marked-images. However, the process of finding the optimal value of both the thresholds was cumbersome as those values were found iteratively. Further, the performance was

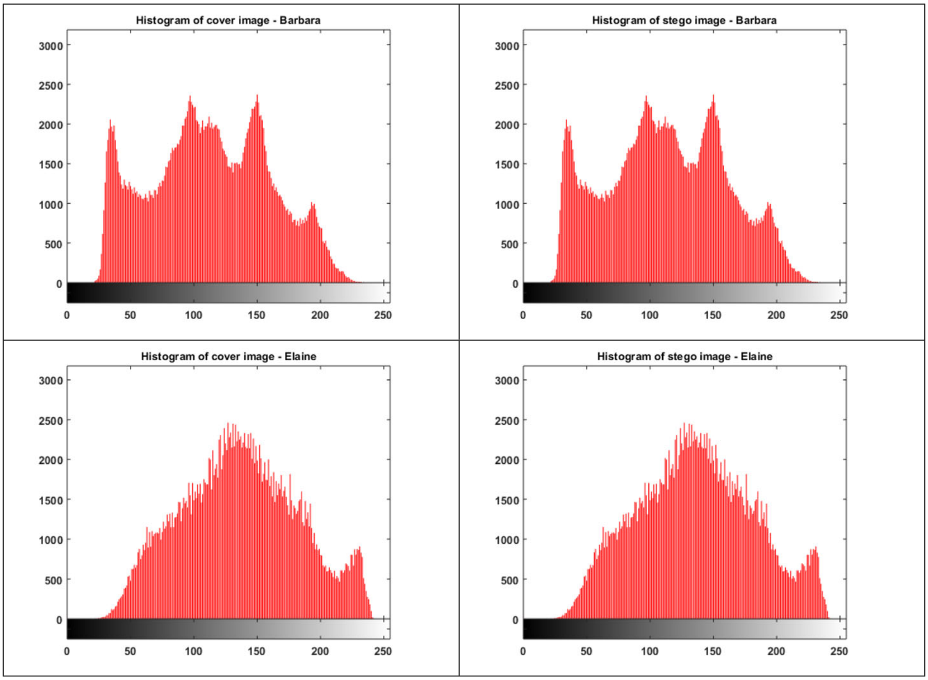


Fig. 12 (continued)

highly dependent on the performance of pairwise PEE and PVO methods. In the future work, the research can be directed to find the optimal values of threshold based on some image characteristics parameters such as entropy.

Acknowledgments This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1I1A3049788) and Brain Pool program funded by the Ministry of Science and ICT through the National Research Foundation of Korea (2019H1D3A1A01101687, 2021H1D3A2A01099390).

Data availability Supporting data and information can be found on the corresponding author's research gate profile.

Declarations

Conflict of interest There is no conflict of interest

References

1. Alattar AM (2004) Reversible watermark using the difference expansion of a generalized integer transform. IEEE Trans Image Proc 13(8):1147–1156

2. Dragoi I-C, Coltuc D (2016) Adaptive pairing reversible watermarking. IEEE Trans Image Proc 25(5): 2420–2422
3. Dragoi IC, Coltuc D, Caciula I (2015) Horizontal pairwise reversible watermarking. Signal Processing Conference (EUSIPCO) 2015 23rd European. pp. 56–60
4. Dragoi IC, Coltuc D, Coanda HG (2017) Adaptive pairwise reversible watermarking with horizontal grouping. Signals Circuits and Systems (ISSCS) 2017 International Symposium on. pp. 1–4
5. Fridrich J, Goljan M, Du R (2002, 2002) Lossless Data Embedding—New Paradigm in Digital Watermarking, EURASIP J. Adv Signal Proc (2). https://doi.org/10.1155/S1110865702000537
6. Gadamsetty S, Rupa CH, Anusha CH, Iwendi C, Gadekallu TR (2022) Hash-based deep learning approach for remote sensing satellite imagery detection. Water 14(5):707
7. He W, Cai Z (2021) Reversible data hiding based on dual pairwise prediction-error expansion. IEEE Trans Image Proc 30:5045–5055
8. He W, Xiong G, Weng S, Cai Z, Wang Y (2018) Reversible data hiding using multi-pass pixel-value-ordering and pairwise prediction-error expansion. Inf Sci 467:784–799
9. He W, Xiong G, Wang (2022) Reversible data hiding based on multiple pairwise PEE and two-layer embedding. Sec Commun Netw. https://doi.org/10.1155/2022/2051058
10. Hong W, Chen TS, Shiu CW (2009) Reversible data hiding for high quality images using modification of prediction errors. J Syst Softw 82(11):1833–1842
11. Hou J, Ou B, Tian H, Qin Z (2021) Reversible data hiding based on multiple histograms modification and deep neural networks. Signal Process Image Commun 92:116118
12. Hui S, Jingning H, Yaowu X (2014) An optimized template matching approach to intra coding in video/image compression. Proceedings of SPIE - The International Society for Optical Engineering. 9029. https://doi.org/10.1117/12.2040890.
13. Kamstra L, Heijmans HJAM (2005) Reversible data embedding into images using wavelet techniques and sorting. IEEE Trans Image Proc 14:2082–2090
14. Kaur G, Singh S, Rani R, Kumar R (2020) A comprehensive study of reversible data hiding (RDH) schemes based on pixel value ordering (PVO). Arch Computat Methods Eng
15. Kaur G, Singh S, Rani R, Kumar R, Malik A, (2021) High-quality reversible data hiding scheme using sorting and enhanced pairwise PEE. IET image processing. 1–15
16. Kumar R, Chand S (2016) A reversible high capacity data hiding scheme using pixel value adjusting feature. Multimed Tools Appl 75(1):241–259
17. Kumar R, Chand S (2018) A reversible data hiding scheme using pixel location. Int Arab J Inf Technol 15(4):763–768
18. Kumar R, Jung K-H (2020) Enhanced pairwise I-PVO-based reversible data hiding scheme using rhombus context. Inf Sci 536:101–119. https://doi.org/10.1016/j.ins.2020.05.047
19. Kumar R, Kim DS, Lim SH, Jung KH (2019) High-Fidelity reversible data hiding using block extension strategy. In: 2019 34th international technical conference on circuits/systems, computers and communications (ITC-CSCC). IEEE, JeJu, Korea (South), pp 1–4
20. Kumar R, Kumar N, Jung K-H (2020) Color image steganography scheme using gray invariant in AMBTC compression domain. Multimedia System Sign Process 31, 1145, 1162.
21. Kumar R, Kumar N, Jung K-H (2020) I-PVO based High Capacity Reversible Data Hiding using Bin Reservation Strategy. Multimedia Tools App, Springer 79:22635–22651
22. Kumar N, Kumar R, Caldelli R (2021) Local moment driven PVO based reversible data hiding. IEEE Signal Proc Lett 28:1335–1339
23. Lee C-F, Shen J-J, Wu Y-J, Agrawal S (2020) PVO-based reversible data hiding exploiting two-layer embedding for enhancing image Fidelity. Symmetry 12:1164
24. Li X, Yang B, Zeng T (2011) Efficient reversible watermarking based on adaptive prediction-error expansion and pixel selection. IEEE Trans Image Proc 20(12):3524–3533
25. Li X, Li B, Yang B, Zeng T (2013) General framework to histogram-shifting based reversible data hiding. IEEE Trans Image Proc 22(6):2181–2191
26. Li X, Li J, Li B, Yang B (2013) High-fidelity reversible data hiding scheme based on pixel-value ordering and prediction-error expansion. Signal Process 93(1):198–205
27. Malik A, Singh S, Kumar R (2018) Recovery based high capacity reversible data hiding scheme using even-odd embedding. Multimed Tools Appl 77(12):15803–15827
28. Ni Z, Shi YQ, Ansari N, Su W (2006) Reversible data hiding. IEEE Transact Circuits Syst Video Technol 16(3):354–362
29. Ou B, Li X, Zhao Y, Ni R, Shi Y-Q (2013) Pairwise prediction-error expansion for efficient reversible data hiding. IEEE Trans Image Proc 22(12):5010–5021

30. Ou B, Li X, Zhang W, Zhao Y (2019) Improving pairwise PEE via hybrid-dimensional histogram generation and adaptive mapping selection. IEEE Trans Circuits Syst Video Technol 29(7):2176–2190. https://doi.org/10.1109/TCSVT.7610.1109/TCSVT.2018.2859792
31. Peng F, Li X, Yang B (2014) Improved PVO-based reversible data hiding. Dig Signal Proc 25:255–265
32. Qu X, Kim HJ (2015) Pixel-based pixel value ordering predictor for high-fidelity reversible data hiding. Signal Process 111:249–260
33. Reddy Gadekallu T, Srivastava G, Liyanage M, et. al. Hand gesture recognition based on a Harris hawks optimized convolution neural network. Computers & Electrical Engineering, vol. 100, Article ID 107836
34. Sachnev V, Kim HJ, Nam J, Suresh S, Shi YQ (2009) Reversible watermarking algorithm using sorting and prediction. IEEE Transac Circuits Syst Video Technol 19(7):989–999
35. Tan TK, Boon CS, Suzuki Y (2006) Intra Prediction by Template Matching. International conference on image processing, Atlanta, GA, 2006, pp. 1693–1696 https://doi.org/10.1109/ICIP.2006.312685
36. The USC-SIPI Image Database (n.d.) [Online]. Available: http://sipi.usc.edu/database
37. Thodi DM, Rodriguez JJ (2007) Expansion embedding techniques for reversible watermarking. IEEE Trans Image Proc 16(3):721–730
38. Tian J (2003) Reversible data embedding using a difference expansion. IEEE Transac Circuits Syst Video Technol 13(8):890–896
39. Weng S, Zhao Y, Pan JS, Ni R (2008) Reversible watermarking based on invariability and adjustment on pixel pairs. IEEE Signal Proc Lett 15(1):721–724
40. Weng S, Liu Y, Pan JS, Cai N (2016) Reversible data hiding based on flexible block-partition and adaptive block-modification strategy. J Vis Commun Image Represent 41:185–199
41. Weng SW, Tan WL, Ou B, Pan JS (2021) Reversible data hiding method for multi-histogram point selection based on improved crisscross optimization algorithm. Inf Sci 549:13–33
42. Wu H-T, Huang J (2012) Reversible image watermarking on prediction errors by efficient histogram modification. Signal Process 92(12):3000–3009
43. Wu H, Li X, Zhao Y, Ni R (2019) Improved reversible data hiding based on PVO and adaptive pairwise embedding. J Real-Time Image Proc 16(3):685–695
44. Zhang T, Li X, Qi W, Guo Z (2020) Location-based PVO and adaptive pairwise modification for efficient reversible data hiding. IEEE Transac Inform Forensics Sec 15:2306–2319. https://doi.org/10.1109/TIFS.2019.2963766
45. Zhao WQ, Yang BL, Gong SZ (2018) A higher efficient reversible data hiding scheme based on pixel value ordering. J Inf Hiding Multimed Signal Process 9:918–928

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

Scheduling of Energy Storage System (ESS) for Electricity Distribution Companies (DISCOMs)

Chetan Gusain
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
chetangusain95@gmail.com

Madan Mohan Tripathi
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
mmtripathi@dce.ac.in

Uma Nangia
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
umanangia@dce.ac.in

2022 IEEE 10th Power India International Conference (PIICON) | 978-1-6654-5930-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/PIICON56320.2022.10045220

Abstract—**With the growth of distributed energy resources (DERs), there has been an overall increase in power generation. Because of this, the type of power produced fluctuates, contributing to grid instability. To lessen this unplanned power outage and avoid system disruptions, the electricity distribution companies (DISCOMs) give their schedules for 96-time blocks daily. The present study investigates the amount of Deviation Charges (DC) administered by DISCOMs in 15-minute time blocks due to the Deviation Settlement Mechanism (DSM). After examining the power deviation curve, it was observed that implementing Energy Storage System (ESS) at the distribution side and then facilitating their energy scheduling is the best practice to minimize the overall financial impact of the DISCOMs. Therefore, a simulation-based tool was proposed that includes Battery Control Algorithm (BCA) and Battery Degradation Model (BDM) for Kolkata DISCOM using real-time deviation data of 29 days. The results indicate that the proposed simulation-based model of ESS increased the project's financial viability and the battery's expected lifetime.**

Keywords—Deviation settlement mechanism, Energy storage system, Battery degradation, DISCOMs, Energy scheduling

I. Introduction

Today, power distribution companies (DISCOMs) have undergone huge financial problems [1] due to substantial changes in the 15-minutes power deviation. In India, poor reliability of power supply has been addressed, making it difficult for all the system operators to manage and adhere to the scheduled power in the Day-Ahead-Market (DAM) [2]. As a result, various electricity DISCOMs suffer severe deviation penalties. The Deviation Settlement mechanism (DSM) is a new regulatory solution to reduce Deviation Charges (DC) introduced by Central Electricity Regulatory Commission (CERC) in 2014 [3], [4]. There are several other options for meeting the given scheduled power; one is to link to Energy Storage System (ESS) [5]. The popularity of ESS has been growing over the past several years due to the rising demand, and it is now gradually emerging in several energy industries [6].

Currently, rapidly growing energy consumption and fast penetration of renewable energy (RE) are processing a threat to electrical infrastructure due to changing load patterns and rising demand over time [7]. Thus, higher inequalities in actual and scheduled power in the DISCOMs brings the most significant challenge that power utilities manage. Also, utility companies delay making the investment necessary to upgrade or replace their outdated electrical SCADA accounting network [8]. Therefore, this research paper's main novelty is to implement the application of ESS for power utilities on the distribution side, which can reduce deviation penalties due to an imbalance in the system paired with the subsequent frequency range [9], [10], [11]. Additionally, this research aims to develop a battery-linked deviation settlement tool that intends to evaluate the technical viability of large-scale grid-connected ESS while incorporating various ESS parameters like battery life, battery degradation, and battery size. The research article provides a comprehensive overview of the ESS requirements emerging for grid-scale applications [12] in the Indian grid code to shift the power deviation to lower the net deviation costs and use the benefits of deviation management.

The present research paper is divided into five sections. The literature review section gives an overview of the current state of the Indian power market and the structure of the country's electricity market, which covers both day-ahead and real-time energy transactions. Additionally, the stacking of grid applications of ESS and cost-reduction economic analysis of energy storage has been researched to assess their effects on the project's viability. The materials and methods section includes a thorough discussion of historical data regarding power schedules and deviations, which paints a realistic picture of the use of battery storage for the project implementation in Kolkata DISCOM as a case study [13], [14]. The section's methodology explains the ESS sampling every 30 seconds, which assists in controlling the DISCOMS to lower the deviation penalties incurred in each 15-minute time block and then generate revenue through ESS's deviation settlement costs. Recommendations for the choice of energy storage are given in the results and discussion section, which displays the percentage decrease in DISCOM's penalty when employing ESS. Further processing of the project is determined in the conclusion and future scope.

II. Literature Review

A. Current Indian Power Market

The power market scenario has been changing due to the rise of per-capita energy demand caused by the gradual increase in the global population [15]. According to the 2021-22 India Economic Survey [16], this favourable shift in the market has encouraged numerous foreign investors to increase their investments in various industries, including power production, grid network/or transmission, and battery storage. With a goal to reach 175 GW and 450 GW of RE capacity by 2022 and 2030, respectively, India is now in the lead with the second-fastest expanding power industry in the world [17]. In addition, the development of better infrastructure, effective energy management, and digitalization has resulted in a more intelligent grid management system that directly streamlines the rapid growth of the Indian Power sector.

According to the CEA report, the total installed capacity as on 30.09.2022 is 407.32 GW with 117.6 GW of installed renewable energy (RE) capacity [18], as shown in *Fig. 1.* With

978-1-6654-5930-3/22/$31.00 ©2022 IEEE

this, India is ranked 4th in the world for the installed capacity of both wind and solar and 3rd for the total installed renewable energy (RE) capacity. India has adopted a "one nation, one grid" policy, which refers to coordinating all five regional grids into one national grid at a single frequency [19]. However, this impacts India's entire energy trading market. This policy was implemented in response to the excess growth of demand and generation. According to the CEA data, 1491.90 billion units (BU) of electricity produced in the year (2021-22) compared to 1381.90 billion units (BU) in the year (2020-12) [20], representing an average yearly growth rate of roughly 7.96%, shown in *Fig. 2* . Over 80% of the surplus energy produced was supplied by coal.

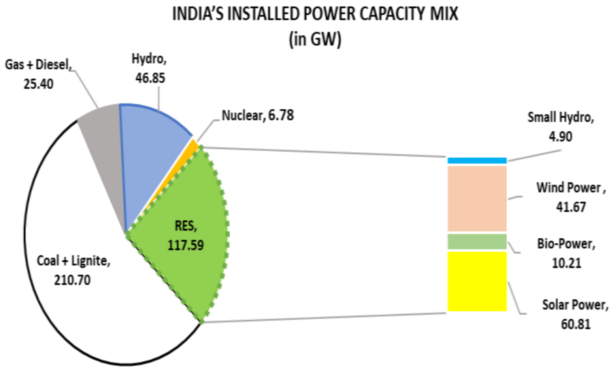


Fig. 1. Total installed capacity as on 30.09.2022

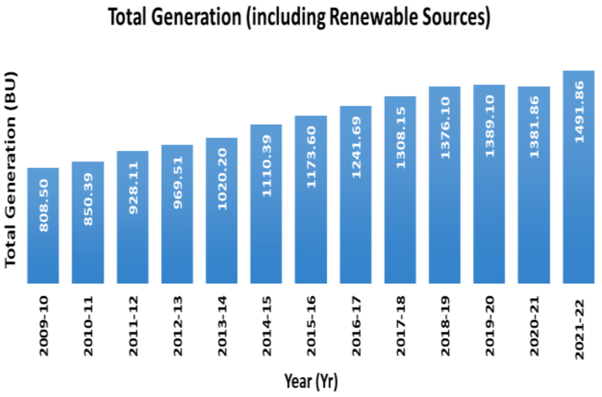


Fig. 2. Electricity generation in India from 2009 to 2021

B. Structure of the Indian Electricity Market

The Electricity Act of 2003 was a significant catalyst for the growth of the Indian power market because it allowed for a multi-buyer or multi-seller framework while de-licensing the power generating industry, which resulted in the introduction of power traders to the Indian market for the first time. In 2004, it created the Open-Access method, which gives utilities the power to select a provider from any region of India. The market introduced a new feature in 2012 called the 15-minute average time block for selling power, with the primary goal being to uphold grid discipline and reinforce the corresponding frequency change. The former unscheduled interchange (UI) penalties were replaced by the Deviation settlement method, which CERC implemented in 2014. In order to boost the momentum of power trading, remove congestion at the level of both production and distribution, and improve grid frequency stabilization, new improvements have been made, as well as the introduction of an ancillary services mechanism in 2016. Another change is the Electricity Act of 2018 [21], which intends to improve supply competition on the distribution side by allowing several service operators to deliver electricity to utilities to create a dynamic market structure and also offer incentives to RE generators. It encouraged the Direct Benefit Transfer (DBT) for electricity grants and managed DISCOM's losses.

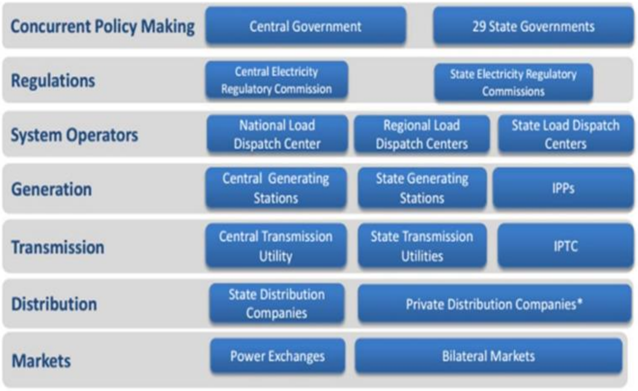

Concurrent Policy Making
Central Government
29 State Governments
Regulations
Central Electricity Regulatory Commission
State Electricity Regulatory Commissions
System Operators
National Load Dispatch Center
Regional Load Dispatch Centers
State Load Dispatch Centers
Generation
Central Generating Stations
State Generating Stations
IPPs
Transmission
Central Transmission Utility
State Transmission Utilities
IPTC
Distribution
State Distribution Companies
Private Distribution Companies*
Markets
Power Exchanges
Bilateral Markets

Fig. 3. Structure of India's power sector

Since electricity comes within the concurrent subject matter of the Indian constitution, the policy-making at the top of the hierarchy is separated into two parts: the central regime and the state regime. The Indian power market is divided into numerous segments. National Load Dispatch Center (NLDC) maintains the scheduling and dispatching of power on a national level, divided into five Regional Load Dispatch Centers (RLDCs) [22]. At the state level that the 33 States Load Dispatch Centers handle, it has been responsible for managing this entire system, considering the system operators that operate the National Grid (SLDCs), as shown in *Fig. 3*.

III. MATERIALS AND METHODS

A deviation in the 15-minute time block occurs due to the imbalance, particularly between the schedule drawl and actual drawl, resulting from the rise in electricity demand, particularly on the distribution side for the state utilities. This deviation significantly impacts the frequency-linked deviation settlement charges in the distribution licensee area. There are three categories for the overall deviation charges. The first charges are a result of deviation at each overdrawl and receiving compensation for each underdrawl, the second charges are a result of other deviation that is solely related to the frequency that is restricted between (49.85-50.05 Hz), and the third charges are a result of sustained deviation.

A. Case Study of Kolkata DISCOM

In this paper, the case study of load statistics of Kolkata DISCOM with 96-time blocks is considered. The daily load variation curve is drawn in *Fig. 4* between Mar. 11 to Apr. 04, 2019, and each time block lasts for 15 minutes. It is a daily historical data set that includes both the current schedule and the schedule for the day before. The DSM modification tightens the grid frequency in a given range for improved grid security and stability. Also, the IEX rates for a specific time block are checked correspondingly.

According to the given load data of Kolkata DISCOM, frequency-linked charges of the DSM are characterized in two ways. First is overdrawl: if the frequency is below 49.85 Hz, deviation charges and additional deviation charges are imposed. If the frequency is between 49.85 Hz and 50 Hz, deviation charges or additional deviation charges are assessed depending on the frequency. If the frequency exceeds 50.05 Hz, deviation charges and additional deviation charges are nullified. The second is under drawl, where deviation charges are imposed if the frequency is less than 49.85 Hz. And no additional deviation charges are imposed if the frequency is between 49.85 Hz and 50 Hz. Only additional charges linked to ACP rates at 50 Hz are imposed if the frequency is more significant than 50.05 Hz.

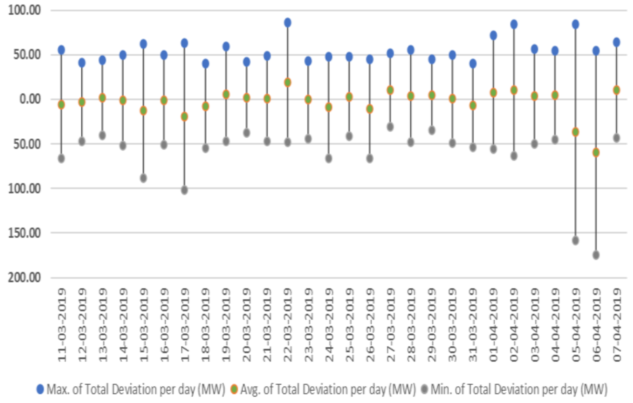


Fig. 4. Daily deviation pattern (MW) in given time-interval

B. Methodology for Deviation Settlement

After the scheduled and actual generation data is extracted, the following formula is used to determine the deviation and accompanying settlement penalties:

- Deviation Calculation:

Deviation (Pd) is the difference between the actual (Pa) andforecasted schedule (Ps). It is calculated in MW using *(1)* and *Table I* :

$$\Delta P_D = P_A - P_S \qquad (1)$$

Where;

PA = Actual power in the DISCOMs

PS = Scheduled Power in the DISCOMs

Table I. Calculation of additional deviation charges within slabs as per regulation (Over drawl):

When 12% <= 150 MW	**When 12% <= 150 MW**	**Additional deviation charges**
12-15%	150-200 MW	20% of Normal Deviation Charges on average freq. of that block
15-20%	200-250 MW	40% of Normal Deviation Charges on average freq. of that block
Above 20%	Above 250 MW	100% of Normal Deviation Charges on average freq. of that block

C. Need for ESS at the Kolkata DISCOM

Energy storage, which temporarily stores energy and uses it to offset the daily power variation generated to reduce the deviation penalties imposed on the Kolkata DISCOM, is essential in this situation. The complete interpolation and scheduling of 30-second sampling of ESS in real-time data aid erroneous forecasting systems as well as energy trading, which DISCOM used to resolve working interruptions caused by unpredictability and uncertainty. Therefore, the purpose of adopting ESS charging and discharging is to entice DISCOMs to engage in the real-time market response program to manage the imbalance and reduce the net accounting deviation charge. The ESS's function in this situation is to charge the battery. At the same time, under-drawl beneficiaries are present at the stated provided state of charge (SOC) and discharge the battery to the DISCOMs when over-drawl charges are present. The spreadsheet is designed for Deviation Management (DM) tool. The primary purpose of introducing the tool is to assess the technical feasibility of ESS at DISCOMs. DM tool uses the utility's 15-minute time block of the day to analyze the advantages by reducing deviation charges (DC). The DM tool is a platform that allows users to interact with findings to visualize them and do different types of analysis depending on battery characteristics, technological input, etc.

D. Approach adopted for Deviation Management (DM) Tool

General Assumptions:

- For the past 29 days of historical data, DISCOM has had access to a power schedule and an actual drawl.
- The frequency-linked DSM rate is evaluated using the IEX rates every 15 minutes (INR/kWh).
- Battery's primary application is to participate in a real-time market reaction that manages imbalances and lowers net accounting deviation charging.

Data Preparation:

- The tool takes data for DISCOM schedule load in intervals of 15 minutes which operates in real-time.
- The model may contain real-time frequency following the time steps specified for power deviation
- The Indian Energy Exchange (IEX) rates and area clearing price (ACP) imported for various dates and time blocks, including underdrawl and overdrawl
- According to ESS, various parameters, such as DoD and degradation characteristics, are used to anticipate the project life energy throughput from the battery.

Model Methodology:

- The model is an interactive Excel-based algorithm tool.
- A more effective manner to convey the computations and quick visualization in the dashboard to make the presentation of the results easier.
- The model analyses the battery's charging and discharging mode capacity depending on input data like DoD, SoC, etc.
- Based on the operating depth of discharge (DoD) and the number of cycles over a given time, the model forecasts the battery technology's cycle life.

Operating Assumptions:

- Depending on factors, the preliminary battery size may change. Any calculations considering a battery size are displayed in the results chapter.
- When there is a drawl, the battery is charged by the grid.

i. Initial State of Charge = 100% is an assumption for battery size.
ii. Li-ion batteries have a charging/discharging efficiency of 95%, compared to VRLA batteries' 80%.
iii. Depth of Discharge: 70–90%, with operating DoD to alter and displayed in outcomes.

IV. Results and Discussion

A. *Feasible storage technology selection*

In order to decrease the deviation for DAM, this project plans to build a methodology that allows DISCOMs to compute real-time charges. Limiting the deviation is better than applying for management deviation settlement and reducing the costs and penalties for severely deviating from the schedule. In order to support this real-time application, it is necessary to select the optimum battery option that not only reduces penalties but also, through automated charging and discharging, increases the demands of the DISCOMs. Here for a better study of the outcome, historical data are used to quantify the number of penalties. In order to synchronize battery applications with grid-scale, an excel sheet was built to grasp based on a 96-time block and make a choice regarding its rated capacity. Here, the data is extrapolated for 30 seconds to assess the actual prevalence of ESS in real-time and act appropriately about the 15-minute block to make the study more quantitative.

Table II. Monthly summary and Distribution of DSM penalties

Date	Total Penalties			Net DSM (Rs.)
	Charge of Dev. (Rs.)	Addl. charge of Dev. (Rs.)	Penalty for cont. one dir. Dev. (Rs.)	
11.03.2019	615631	77120	492504	11,85,255
12.03.2019	-75694	21421	0	-54,273
13.03.2019	18490	9678	3698	31,866
14.03.2019	-142090	36786	28418	-76,886
15.03.2019	50941	93110	0	1,44,051
16.03.2019	-177932	34428	0	-1,43,504
17.03.2019	16520	106208	0	1,22,728
18.03.2019	-36665	35831	0	-834
19.03.2019	3943	38249	1577	43,769
20.03.2019	-158240	78419	0	-79,821
21.03.2019	-137081	150450	0	13,369
22.03.2019	211024	148118	42205	4,01,347
23.03.2019	146520	32258	0	1,78,778
24.03.2019	220489	131453	0	3,51,942
25.03.2019	138590	0	0	1,38,590
26.03.2019	-335178	74630	0	-2,60,548
27.03.2019	138960	76815	55584	2,71,359
28.03.2019	737303	314899	294921	13,47,123
29.03.2019	65747	52548	13149	1,31,444
30.03.2019	470058	87041	94012	6,51,111
31.03.2019	23030	160744	4606	1,88,380
01.04.2019	1432253	169607	1145803	27,47,663
02.04.2019	513834	67270	411068	9,92,172
03.04.2019	237900	63120	95160	3,96,180
04.04.2019	-153916	28945	0	-1,24,971
05.04.2019	-116294	77446	0	-38,848
06.04.2019	61790	122444	0	1,84,234
07.04.2019	22095	215672	0	2,37,767
	37,92,028	25,04,710	26,82,705	89,79,443

The primary goal is to quantify the amount of deviation and the related penalties for a DISCOM operating in a certain Indian region. The penalties are quantified in various ways, so it is simple to show the types of charges that often occur daily, as demonstrated in *Table II*. Along with this, the utilities that are frequency-linked carried with (-) sign. When the frequency is assumed to be higher than 50.05 Hz, the deviation occurs to be the fewer penalties arising in a 15-minute time interval. It receives the negative back as revenue in the ESS system that the DISCOMs facilitate on the latter in higher deviation-linked frequency.

B. *Performance at charging and discharging*

The simulation results of the state of charge (SoC) before and after battery performance with a 5000 kWh capacity, shown in *Fig. 5*, along with hourly analysis for a day.

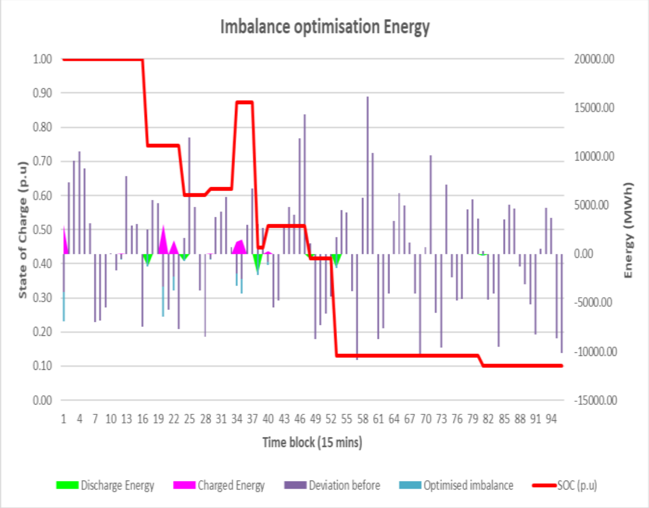


Fig. 5. Hourly analysis of imbalance energy before and after the simulation

According to the Battery Control Algorithm (BCA) logic, the battery begins to discharge as soon as a peak rate in the afternoon at a certain time of day goes up, as shown below in *Fig. 6*. During which battery starts reducing the DSM penalties of DISCOMs.

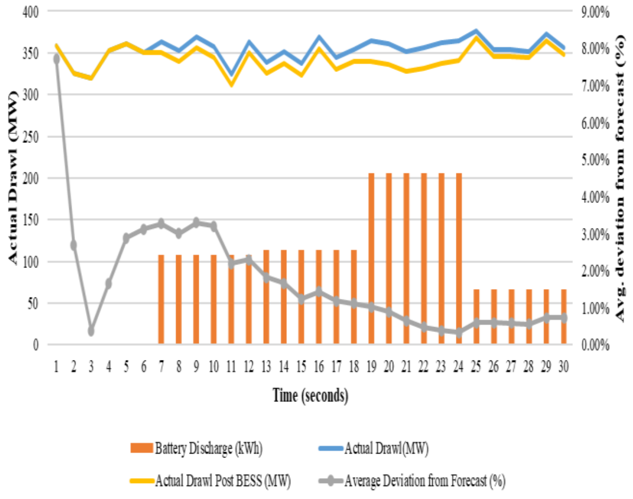


Fig. 6. Battery Discharging control result for 1st block (during OD)

However, to take advantage of peak rates, the C-rate for discharging at peak rates must differ from that of off-peak rates. The battery is shown to be charging more cheaply and off-peak. To account for the losses that cause additional deterioration, the battery is seen to be inactive during mid-peak hours, as shown below in *Fig. 7*.

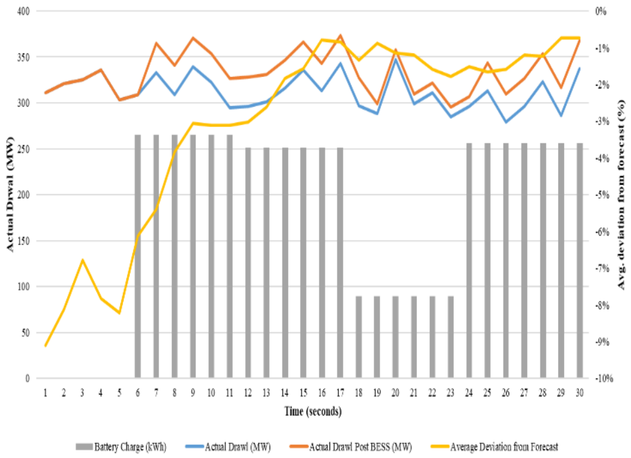


Fig. 7. Battery Charging control result for 1st block during UD (Post BESS Operation)

V. Discussions and Future Work

The project aims to construct an excel-based model to assess the technical viability of establishing an energy storage system for managing deviation at DISCOMs. According to the report, with the current battery pricing, one of the DISCOMs' challenging responsibilities is determining the extent of ESS for managing deviations. Therefore, combining different applications and activating the ancillary market should be crucial for greater technical viability, which does not just work with a single business to buy and sell as it often did in the energy trading market. Additionally, various restrictions prevent it from being used in a useful way by DISCOM beneficiaries. The first difficulty facing today's communities is the cost of installing such a system since utilities are always concerned about their investment's return. Additionally, the main novelty of the paper is the utility-scale optimization of Application stacking and usage of Energy storage for wholesale market participation in the day-ahead market and real-time dispatch trading.

References

[1] RMI and NITI, *Turning around the power distribution sector : learnings and best practices from reforms*. 2021.

[2] Central Electricity Regulatory Commission, “Monthly report on short-term transactions of electricity in India - March 2018,” no. April, pp. 1–27, 2018.

[3] B. Koul, K. Singh, and Y. S. Brar, “Deviation Settlement Mechanism and Its Implementation in Indian Electricity Grid,” *Lect. Notes Electr. Eng.*, vol. 768, pp. 237–246, 2022, doi: 10.1007/978-981-16-2354-7_22/COVER.

[4] CERC, “Deviation Settlement Mechanism and Related Matters.” 2014.

[5] F. Braeuer, J. Rominger, R. McKenna, and W. Fichtner, “Battery storage systems: An economic model-based analysis of parallel revenue streams and general implications for industry,” *Appl. Energy*, vol. 239, pp. 1424–1440, Apr. 2019, doi: 10.1016/J.APENERGY.2019.01.050.

[6] D. Sen, “Battery Storage: An Enabler for Utility and Grid Dynamics,” pp. 149–156, 2022, doi: 10.1007/978-3-030-76221-6_22.

[7] V. P. Wright, “World Energy Outlook.,” pp. 23–28, 1986.

[8] M. Y. Shabalov, Y. L. Zhukovskiy, A. D. Buldysko, B. Gil, and V. V. Starshaia, “The influence of technological changes in energy efficiency on the infrastructure deterioration in the energy sector,” *Energy Reports*, vol. 7, pp. 2664–2680, 2021, doi: 10.1016/j.egyr.2021.05.001.

[9] J. Satre and S. Deshmukh, “Deviation Settlement Mechanism Linked with Market Price in Indian Power Sector,” *2018 IEEE Int. WIE Conf. Electr. Comput. Eng. WIECON-ECE 2018*, vol. limi, pp. 168–171, Dec. 2018, doi: 10.1109/WIECON-ECE.2018.8783024.

[10] S. R. Sabbanwar, R. J. Satputaley, and V. B. Borghate, “A Review on Strategies to Control Frequency Variation in Deviation Settlement Mechanism,” *Proc. 2021 IEEE 2nd Int. Conf. Smart Technol. Power, Energy Control. STPEC 2021*, 2021, doi: 10.1109/STPEC52385.2021.9718731.

[11] P. Gupta and Y. P. Verma, “Optimisation of deviation settlement charges using residential demand response under frequency-linked pricing environment,” *IET Gener. Transm. Distrib.*, vol. 13, no. 12, pp. 2362–2371, Jun. 2019, doi: 10.1049/iet-gtd.2018.7116.

[12] M. T. Lawder *et al.*, “Battery energy storage system (BESS) and battery management system (BMS) for grid-scale applications,” *Proc. IEEE*, vol. 102, no. 6, pp. 1014–1030, 2014, doi: 10.1109/JPROC.2014.2317451.

[13] R. K. Das, “An O verview of CESC ’ s 315 kWh Grid Connected Battery Energy Storage System,” pp. 1–6, 2022.

[14] N. Kane, L. Wasan, and B. Karunakaran, “Experience of Battery Energy Storage System by TATA Power DDL,” pp. 295–307, 2022, doi: 10.1007/978-981-16-8727-3_32.

[15] I. Energy Agency, “India Energy Outlook 2021 World Energy Outlook Special Report”, Accessed: Nov. 07, 2021. [Online]. Available: www.iea.org/t&c/

[16] W. A. Lewis, “Economic Survey,” *Econ. Surv.*, 2013, doi: 10.4324/9781315016702.

[17] “Report of Expert Group on 175 GW RE by 2022,” 1386.

[18] C. E. Authority, R. Project, and M. Division, “September-2022 1,” pp. 1–32, 2022.

[19] Department of Agriculture & Farmers Welfare, “Annual Report 2021-22,” *Minist. Agric. Farmers Welf. Gov. India*, pp. 1–307, 2021.

[20] C. E. Authority, “(Draft) Generation Vol- I Government of India Ministry of Power Central Electricity Authority,” no. 4, 2022.

[21] “Draft Central Electricity Regulatory Commission (Deviation Settlement Mechanism and related matters) (Fifth Amendment) Central Electricity Regulatory Commission New Delhi”.

[22] C. O. S. Patricia, “Evolution of System Operation and Emergence of POSOCO as an Independent Institution in India,” vol. 3, no. 2, p. 6, 2021.

2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT)
Manipal University, Jaipur. Jan 19-20, 2023

2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT) | 978-1-6654-5357-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCT56969.2023.10075964

Seismic Lithology Interpretation using Attention based Convolutional Neural Networks

Vineela Chandra Dodda
Department of ECE
SRM University AP
Andhra pradesh,India
vineelachandra_dodda@srmap.edu.in

Lakshmi Kuruguntla
Department of ECE
SRM University AP
Andhra pradesh,India
lakshmi_kuruguntla@srmap.edu.in

Shaik Razak
Department of ECE
SRM University AP
Andhra pradesh, India
rajak_shaik@srmap.edu.in

Anup Mandpura
Department of Electrical engineering
Delhi Technological University
New Delhi, India
kanup@dtu.ac.in

Sunil Chinnadurai
Department of ECE
SRM University AP
Andhra pradesh, India
sunil.c@srmap.edu.in

Karthikeyan Elumalai
Department of ECE
SRM University AP
Andhra pradesh, India
karthikeyan.e@srmap.edu.in

***Abstract*—Seismic interpretation is essential to obtain information about the geological layers from seismic data. Manual interpretation, however, necessitates additional pre-processing stages and requires more time and effort. In recent years, Deep Learning (DL) has been applied in the geophysical domain to solve various problems such as denoising, inversion, fault estimation, horizon estimation, etc. In this paper, we propose an Attention-based Deep Convolutional Neural Network (ACNN) for seismic lithology prediction. We used Continuous Wavelet Transform (CWT) to obtain the time-frequency spectrum of seismic data which is further used to train the network. The attention module is used to scale the features from the convolutional layers thus prioritizing the prominent features in the data. We validated the results on blind wells and observed that the proposed method had shown improved accuracy when compared to the existing basic CNN.**

***Index Terms*—Deep learning, Lithology prediction, seismic data, Interpretation**

I. Introduction

Seismic Exploration is a cost effective method used to identify the hydrocarbons and minerals in the earth layers'. The seismic exploration process includes the four basic steps: acquisition, processing, inversion and interpretation. The acquisition step is used to get the raw seismic data from the earth layers, processing step is utilized to attenuate the noise present in the raw data. Whereas inversion step is used to retrieve the physical properties of the earth layers from the seismic reflection data and finally the last step is the interpretation which is used to transform the data in seismic sections into geological information. In [1], authors defined that the seismic interpretation acts as an interface between the exact processed seismic data and inexact geological data. The seismic interpretation process converts the velocity and time of subsurface reflecting layers into depth form which can translate the seismic data into geological images [2]. There are three different types of interpretations; structural interpretation, stratigraphic interpretation and lithological interpretation. Structural interpretation is used to get the structural maps of sub-surface layers with respect to observed arrival times of reflected data. Stratigraphic interpretation gives the pattern of reflections in sub-surface layer. Whereas lithological interpretation is used to identify the potential hydrocarbon-bearing zones in each layer of the earth's sub-surface. Accurate and detailed information of the structural, stratigraphic and lithology interpretation depends on the seismic attributes [3]. There are different types of seismic attributes which includes: instantaneous amplitude, frequency, phase and bandwidth. There are various parameters in Interpretation such as Volume of shale, Porosity, permeability, hydrocarbon saturation, effective porosity, and fluid content which are the important parameters that should be considered for any type of interpretation.

From literature, we infer that methods for seismic lithology prediction can be broadly classified into two; conventional and data-driven methods. In conventional methods, interpretaion is based on physical relation between the seismic attributes and the geological targets [4]–[8]. Whereas in data-driven methods, the statistical relationship between seismic attributes and geological factors is derived using machine learning techniques [9] [10]. Recently, with the rapid development of GPU's and huge availability of data, the data-driven methods gained popularity in various areas such as computer-vision, natural language processing, image and speech recognition etc. Machine learning (ML) techniques can learn nonlinear correlations implicitly from a large number of labels and are usually not constrained as traditional methods, which are based on signal processing theories or physics-based equations. In reservoir characterization, ML techniques such as Support Vector Machine (SVM), Random forests and Deep Learning (DL) has been used to predict permeability, porosity, saturation and volume of sand, shale etc etc [10]–[13]. Here, the

TABLE I
SUMMARY OF LITERATURE WORK FOR SEISMIC LITHOLOGY PREDICTION

Ref.No	problem	Network/Algorithm
[10]-[13]	Seismic liquefaction potential, Prediction of porosity and water saturation, Seismic event classification	SVM
[16]	Salt body classification	CNN
[17]	High resolution 3D porous media reconstruction	GAN
[18]	Seismic facies prediction	DCAE
[19]	Sand shale classification	CWT-CNN

problem of prediction can be either regression, classification or clustering. Artificial Neural networks (ANN) have been used to predict reservoir properties however ANN have issues like over fitting, parameter selection and vanishing gradient problem. On contrary, DL has advantages such as usage of pre-trained models, novel activation functions, GPU functionality etc over ANN. There are many architectures in DL; feed forward neural networks, CNN, Generative Adversarial Network (GAN) and Recurrent neural Network (RNN) [14], [15]. In [16], CNN is used to extract attributes and perform salt body classification from seismic data.

Further in [17], GANs are used to reconstruct high resolution 3D porous media at different scales. In [18], Deep Convolutional Auto encoder (DCAE), an unsupervised approach is used to predict the seismic facies from the prestack seismic data. Whereas in [19], CWT-CNN was used to predict seismic lithology. Furthermore, various network architectures are compared and found that Continuous Wavelet Transform (CWT)-CNN had better performance compared to the other architectures. A brief summary of the literature for seismic lithology prediction is given in TABLE I. In all the aforementioned methods, accuracy and time to train the network are the important factors to consider while interpreting about the seismic lithology.

Therefore, to improve the accuracy and reduce the time for training, we used attention based CNN for seismic lithology prediction in specific sand and shale classification. The wavelet transformed data is used to train the network thus making use of full frequency spectrum of the data. Inspired by the human nature i.e., we add an attention mechanism called squeeze and excitation block to the CNN architecture which further increase the attention and overall performance of seismic lithology prediction. Moreover, the added attention mechanism can model the global information and can handle long-range dependencies [20]. We compared the proposed method with the existing methods and observed an increase in the accuracy of prediction.

In section II, we give the proposed methodology which describes about the network architecture followed by the results in III. Finally, the conclusion is mentioned in the section IV.

II. Methodology

A. Overview of Network Architecture

Deep Learning (DL) techniques are based on learning and recognizing the relationships in the data, similar to operations inside the human brain. From literature, we observe that the commonly used neural network architectures in seismic data denoising and reconstruction are either based on Fully Connected Networks (FCN) or convolutional neural networks.

CNN, on the other hand, is employed in a wide range of real-world applications. CNN can recognize the patterns in the data, which is important for classification, object identification, and natural language processing. The CNN architecture is build by stacking three main layers: Convolution, Pooling and Fully-connected layer. CNNs operate by utilizing layers with filters positioned along each dimension gathering particular information about the visual field. In the training process of CNN, the filter is shifted over the input volume's width and height during the forward pass to generate an activation map that contains the responses of that filter at each spatial point. The stacking of conv-pool layers is used to identify complex features from previous layers. The pooling layers are used to get downsampled data and reduce the computational complexity, while preserving the input features. The output is now flattened using fully-connected layer and back propagation is performed using chosen loss function on each training cycle for the specified number of epochs.

B. Proposed Network Architecture

In our work, we propose to use Attention based CNN for seismic lithology prediction.The ACNN consists of two series of input, convolution, pooling, fully connected, attention and output layers.The input layer is fed with CWT spectrum data which are 2D matrices, where one sample is shown in Fig. 2. The output from feature maps of convolution layers are fed as an input to the fully connected layer which aids in classification. After each series of convolution layer, attention module is used to scale the feature maps. In general, the channels of the network are given equal priority while obtaining the feature maps. This mechanism is altered with the addition of attention module which weighs the channels adaptively. The input to the attention module is convolutional layer and number of channels in that particular layer. Then average pooling is used to squeeze each channel to obtain single value. To introduce non-linearity, we use two fully connected layers with an activation function. Finally, the output of attention module is obtained by multiplying the output of excitation layers with the input of the attention module which is given as input to the next convolution layer. In addition, the advantage of attention module is to avoid the vanishing gradient problem and to provide better explainability about the model. In order to preserve the features, the final convolution layer is not followed by

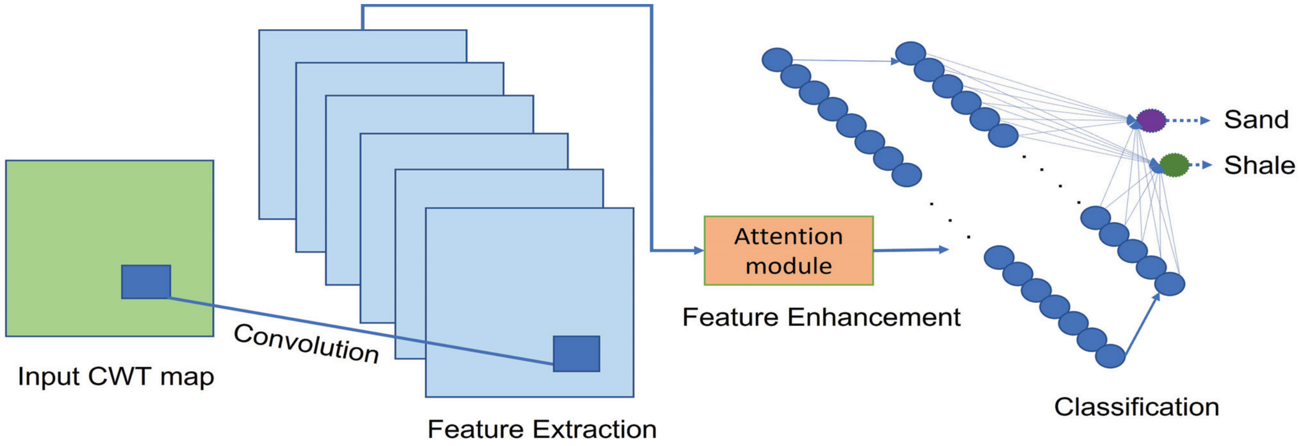


Fig. 1. Proposed ACNN architecture

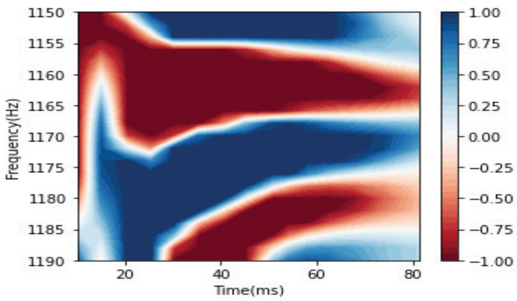


Fig. 2. Sample of 2D spectrum map obtained from Continuous Wavelet Transform (CWT)

activation and pooling layers. We used softmax to calculate the probabilities of the output.

III. Numerical Results

In this section, we demonstrate the various numerical results obtained using proposed ACNN method on blind wells. The dataset (well and post-stack seismic data) is from the eastern slope of the Chuanxi Depression in the Sichuan Basin, SW China.

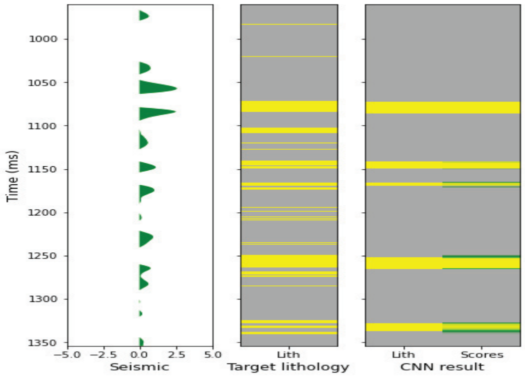


Fig. 3. Results of ACNN on blind test well JS7

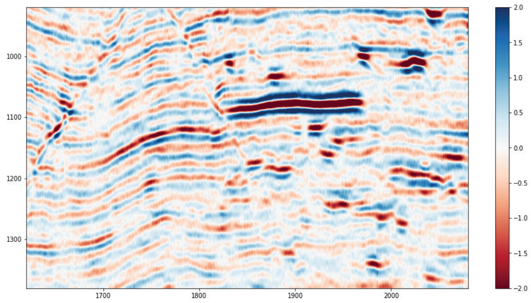


Fig. 4. post-stack seismic data

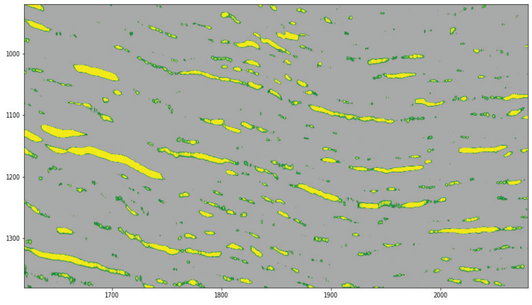


Fig. 5. Sand probability and lithology

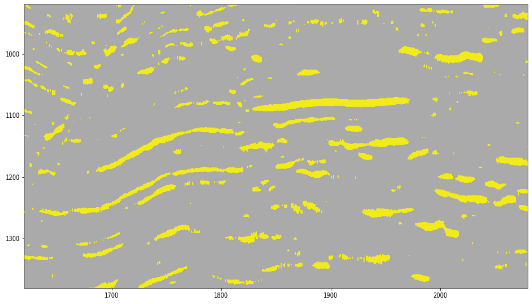


Fig. 6. Sand probability of well JS7

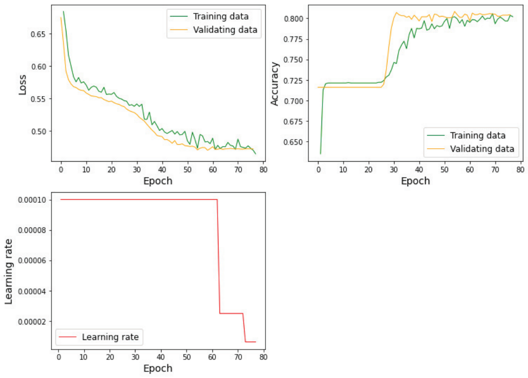


Fig. 7. Analysis plots of Epochs Vs Accuracy, Loss, Learning rate

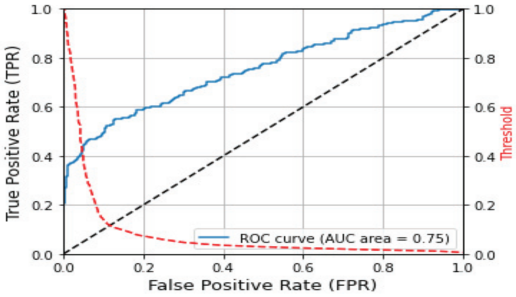


Fig. 8. Plots of True Positive and False positive rates

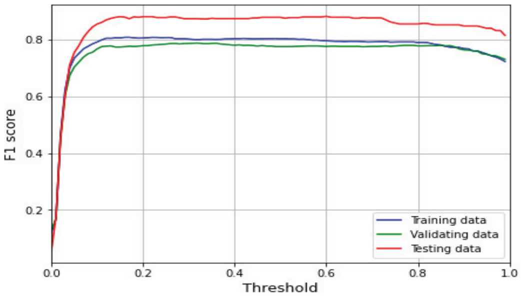


Fig. 9. F1 scores of training, validation and test data

This river-delta region is made up of tightly packed sands. A total of 3720 samples from 13 wells were used for modelling. The complete data set is separated into three subsets: training, validation, and blind testing. The model is trained using the validation and training data sets. The blind testing data set i.e., well JS7, is used for the final model evaluation. We used 40% of data as validation and remaining as training data after checking model performance with various train-test ratios.The ratio of sand samples is very less compared to the ratio of shale. Hence, the minority sand category is over sampled to two times to balance the training and validation set. Fig. 4 shows the post-stack seismic data for which CWT is computed and fed as input to train the network. Figs. 3,5,6 shows the predicted sand probabilities with the proposed ACNN for well JS7. The model is built using Tensorflow deep learning library using python 3.8 on server with configuration Intel® Xeon® Silver 4216 CPU @2.10 GHz (2 processors) with 256 GB RAM, 64-bit operating system. The ACNN network is trained with batch size of 40 and with adaptive learning rate i.e., when the validation error stops decreasing, the learning rate also decreases by four. The epochs is set as 1000, however if learning rate does not decrease two times, training is stopped. We used Binary cross entropy as the cost function and Adam as the optimization algorithm during the training of network. We tuned various hyper parameters and selected the optimal parameters to train the network. aZXSDZ

The performance of the model is evaluated based on confusion matrix, which is widely used metric for classification. There are four values in confusion matrix; TN, FN, FP and TP. From these values, we calculate, accuracy, recall, precision and F1 scorewhich are defined as follows.

$$\begin{aligned} Accuracy &= TP + TN/(P + N) \\ recall &= TP/TP + FN \\ precision &= TP/(TP + FP) \\ F1 &= 2.TP/(2.TP + FN + FP) \end{aligned} \tag{1}$$

Figs. 7, 8 shows the analysis for number of epochs versus learning rate, loss and accuracy. We manually tuned the hyper parameters in the proposed network whereas in future research work, we would explore on automatic tuning of hyper parameters. In this paper, we used F1 score to validate the results as shown in Fig. 9. The higher the f1 score the better the classification model. The accuracy of the proposed ACNN model is 89.67% whereas for the basic CNN, the accuracy is 85.55%.

IV. Conclusion

In this paper, we proposed an attention based CNN for seismic lithology prediction. In our proposed method, we perform classification of sand and shale from the post-stack seismic data. The attention module is used to better retrieve the features and improve the accuracy of prediction. The continuous wavelet transform is computed for the input seismic data and the obtained time-frequency maps are used to train the network. This aids in considering the full frequency features from the data. Moreover, the affect of various hyper parameters in the training process of convolutional neural network is analyzed. The efficacy of the proposed deep learning method is observed using various metrics such as accuracy, precision, recall and F1 score. We observed that the proposed method has achieved higher accuracy compared to the existing CNN method.

Acknowledgement

Authors would like to acknowledge the financial support from Science and Engineering Research Board, Department of Science and Technology, India through Core Research Grant, Ref: CRG/2019/001234. The authors would like to thank Prof. Yangkang Chen at the University of Texas at Austin, for providing research data.

References

[1] L. R. Denham, “Seismic interpretation,” *Proceedings of the IEEE*, vol. 72, no. 10, pp. 1255–1265, 1984.

[2] N. Ahmad, S. Khan, and A. Al-Shuhail, “Seismic data interpretation and petrophysical analysis of kabirwala area tola (01) well, central indus basin, pakistan,” *Applied Sciences*, vol. 11, no. 7, p. 2911, 2021.

[3] M. Anees, “Seismic attribute analysis for reservoir characterization,” in *10th Biennial International Conference and Exposition*, 2013.

[4] D. A. Cooke and W. A. Schneider, “Generalized linear inversion of reflection seismic data,” *Geophysics*, vol. 48, no. 6, pp. 665–676, 1983.

[5] B. Russell and D. Hampson, “Comparison of poststack seismic inversion methods,” in *SEG Technical Program Expanded Abstracts 1991*. Society of Exploration Geophysicists, 1991, pp. 876–878.

[6] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, “Multiscale seismic waveform inversion,” *Geophysics*, vol. 60, no. 5, pp. 1457–1473, 1995.

[7] M. Bosch, T. Mukerji, and E. F. Gonzalez, “Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review,” *Geophysics*, vol. 75, no. 5, pp. 75A165–75A176, 2010.

[8] Q. Zeng, Y. Guo, R. Jiang, J. Ba, H. Ma, and J. Liu, “Fluid sensitivity of rock physics parameters in reservoirs: Quantitative analysis,” *Journal of seismic exploration*, vol. 26, no. 2, pp. 125–140, 2017.

[9] P. S. Schultz, S. Ronen, M. Hattori, and C. Corbett, “Seismic-guided estimation of log properties (part 1: A data-driven interpretation methodology),” *The Leading Edge*, vol. 13, no. 5, pp. 305–310, 1994.

[10] D. P. Hampson, J. S. Schuelke, and J. A. Quirein, “Use of multiattribute transforms to predict log properties from seismic data,” *Geophysics*, vol. 66, no. 1, pp. 220–236, 2001.

[11] M. Pal, “Support vector machines-based modelling of seismic liquefaction potential,” *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 30, no. 10, pp. 983–996, 2006.

[12] S. Na’imi, S. Shadizadeh, M. Riahi, and M. Mirzakhanian, “Estimation of reservoir porosity and water saturation based on seismic attributes using support vector regression approach,” *Journal of Applied Geophysics*, vol. 107, pp. 93–101, 2014.

[13] A. Reynen and P. Audet, “Supervised machine learning on a network scale: Application to seismic event classification and detection,” *Geophysical Journal International*, vol. 210, no. 3, pp. 1394–1409, 2017.

[14] Y. LeCun, Y. Bengio *et al.*, “Convolutional networks for images, speech, and time series,” *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[16] A. Waldeland and A. Solberg, “Salt classification using deep learning,” in *79th eage conference and exhibition 2017*, vol. 2017, no. 1. European Association of Geoscientists & Engineers, 2017, pp. 1–5.

[17] L. Mosser, O. Dubrule, and M. J. Blunt, “Reconstruction of three-dimensional porous media using generative adversarial neural networks,” *Physical Review E*, vol. 96, no. 4, p. 043309, 2017.

[18] F. Qian, M. Yin, X.-Y. Liu, Y.-J. Wang, C. Lu, and G.-M. Hu, “Unsupervised seismic facies analysis via deep convolutional autoencoders,” *Geophysics*, vol. 83, no. 3, pp. A39–A43, 2018.

[19] G. Zhang, Z. Wang, and Y. Chen, “Deep learning for seismic lithology prediction,” *Geophysical Journal International*, vol. 215, no. 2, pp. 1368–1387, 2018.

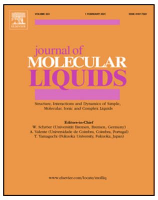
[20] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

Journal of Molecular Liquids 380 (2023) 121705

Contents lists available at ScienceDirect

Journal of Molecular Liquids

journal homepage: www.elsevier.com/locate/molliq



Selective picomolar detection of carcinogenic chromium ions using silver nanoparticles capped via biomolecules from flowers of *Plumeria obtusa*

Samiksha Shukla, Mohan Singh Mehata *

Laser Spectroscopy Laboratory, Department of Applied Physics, Delhi Technological University, Delhi 110042, India

ARTICLE INFO

Article history:
Received 3 September 2022
Revised 12 March 2023
Accepted 20 March 2023
Available online 24 March 2023

Keywords:
Biosynthesized silver nanoparticles
Plumeria obtusa
Absorption spectra
Sensing
Antibacterial activity

ABSTRACT

Biocompatible silver nanoparticles (Ag-NPs) were synthesized by employing an eco-friendly approach of green chemistry with flower extract of *Plumeria obtusa*. Flavonoids are the main phytochemical constituents found in flowers of *Plumeria obtusa*, out of which rutin is the most abundant and acts as a controller. The synthesized Ag-NPs were studied using UV–vis spectroscopy, X-ray diffractometry (XRD), transmission electron microscopy (TEM), fourier transform infrared (FTIR) analysis, zeta potential and dynamic light scattering (DLS) analysis. The Ag-NPs showed their characteristic surface plasmon resonance (SPR) peak at around 430 nm. This peak depended on various physicochemical parameters such as extract concentration, reaction time, temperature and pH. The formation of the FCC lattice of Ag-NPs with an average particle size of 13 ± 1 nm was confirmed. The zeta potential of –22.7 mV of Ag-NPs indicated their stability in colloidal suspensions. The biosynthesized Ag-NPs were used for developing a sensing method for carcinogenic hexavalent chromium ions (Cr^{6+}) in various aqueous mediums that can be utilized in the diagnosis of any contamination in drinking water or food by cancer-causing Cr^{6+} ions with a very efficient limit of detection (LoD) of 95 ± 2 pM (recorded at pH $\sim$ 7.2), which is the lowest reported value for green synthesized nanomaterials and also overcome limitations such as a lack of sensitivity, accuracy, low detection limit and proper explanation. In addition, the antibacterial action of Ag-NPs was investigated against a gram-positive bacterium, *Staphylococcus aureus*. Thus, based on the results, the Ag-NPs synthesized in this study can be resorted to applications like biosensing and biomedicine.

1. Introduction

Due to the growing environmental crisis, the necessity of researching greener sustainable technologies has been increasing. In the past decade, certain ways have been developed to use nanomaterials as ideal prospects for overcoming environmental and ecological challenges, such as wastewater contamination due to chemical dyes, heavy metal ions, microbes, and bacteria [1–8]. Typical synthesis routes of nanomaterials involve either using hazardous chemicals like hydrazine [9] or expensive mechanical methods [10] that consume immense amounts of energy, space, and time. Thus, implementing green alternative synthesis routes lowers the costs and hassle during production and leads to nanomaterials that are non-toxic in nature and eco-friendly.

Silver (Ag) has been valued as a rare precious noble metal alongside gold but is much more abundant comparatively. It is highly reflective, ductile, malleable, exhibits high electrical as well as thermal conductivity, and has a lustrous property. Ag has been used in medical sciences since the classical age [11]. However, the future scope of Ag for diagnostic, medicinal and therapeutic practices lies strictly in nanoscales as the nanoparticles of Ag act as photocatalysts [12,13], antibacterial, antimicrobial, antioxidant, antiplatelet, antidiabetic, antithrombotic agents [14–19] and in targeted drug delivery systems [20,21]. When the interaction of free electrons at the Ag-NPs surface is effectuated at certain wavelengths, the electrons oscillate collectively, giving rise to SPR. Because of this property, Ag-NPs show exceptional absorption and scattering efficiency of light. Thus, the synthesis process of Ag-NPs involves a visual indicator, i.e., the colorless precursor solution of silver nitrate ($AgNO_3$) becoming completely brown owing to SPR as Ag-NPs are formed [22].

Green or biological methods of synthesis of Ag-NPs employ either plants or microorganisms instead of synthetic chemical precursors. The mechanism governing green synthesis has not been explained precisely yet, but the various biocompounds like flavonoids, phenolic acids, terpenoids, proteins, polyols, and alkaloids are considered to reduce Ag^+ ions from precursor solution to Ag^0

* Corresponding author.
E-mail address: msmehata@gmail.com (M.S. Mehata).

https://doi.org/10.1016/j.molliq.2023.121705
0167-7322/© 2023 Elsevier B.V. All rights reserved.

and then cause their capping to form nanoparticles [23,24]. In some cases, aldehydes, ketones, and sugars (aldoses) act as reducing agents [25]. The stabilization of the reduced nanoparticles is caused by proteins, anthracenes, and flavanones [26]. Biosynthesis methodologies also include the usage of microorganisms, which is comparatively difficult as it requires maintaining delicate cell cultures [27]. Therefore, biosynthesis using extracts of plants is the most viable alternative to synthetic synthesis procedures.

Plumeria obtusa is an abundantly available plant, popular for its highly fragrant flowers of bright white color. This plant is commonly known as *Champa* and has been believed to be of great religious, ornamental, pharmacological, and medicinal importance, especially in India. Apart from the use of flowers of *Plumeria Obtusa* for fragrance and decorative purposes, they are actively used in the treatment of diabetes mellitus and aromatherapy [28]. It contains various phytochemicals such as flavonoids, phenolic acids, triterpenoids, steroids, and iridoids [29,30]. The major phytochemical constituent found in flowers of *Plumeria Obtusa* is rutin (500.3 ± 32.0 μg per g of dry flower extract) [31]. Rutin (quercetin-3-rhamnosyl glucoside) is a bioflavonoid with high antioxidant properties allowing it to act as a reducing agent by donating electrons [32]. For this study, the flower extract of *Plumeria obtusa* plant is prepared to be used as a reducer and stabilizer for preparing Ag-NPs. The mechanism behind the biosynthesis of Ag-NPs from *Plumeria obtusa* can be given in Fig. 1.

The current contemporary lifestyle of humans often leads to exposure to many different kinds of toxic heavy metals, including carcinogens such as chromium. These carcinogenic metals are used commercially on a wide scale and cause DNA and cell damage, oxidative stress and cancer-related diseases [33]. Chromium (Cr) exists commonly in 4 oxidation states, including elemental (0), divalent (+2), trivalent (+3), and hexavalent (+6) forms. Out of these, the hexavalent state of Cr is highly carcinogenic due to the mechanism by which it gets reduced, that is, by releasing reactive radicals of hydroxyl, which cause cellular damage and DNA binding [34]. Any manifestation, inhalation, or contact with carcinogens through contaminated food or water items poses increased cancer risks. Since there are no sure-shot inexpensive treatments for fatal diseases like cancer, early detection of carcinogens is essential for cancer prevention. For a long time, the detection method of carcinogens depended on rodents which soon became inefficient [35,36]. Current detection techniques for hexavalent Cr^{6+}, like isotope dilution analysis [37], atomic absorption spectrometry [38,39], and ion chromatography [40] are complicated and chronophagous. Therefore, developing effective, quick, and simple detection techniques for Cr^{6+} ions is essential, especially in liquid assemblies like water or food. Thus, chemo-sensors or optical-sensors based on nanoparticles of noble metals are being actively researched for sensing various materials [41,42].

In this study, we have successfully prepared Ag-NPs using the biomolecules present in the flower extract of *Plumeria obtusa* and investigated an ultrasensitive detection method of heavy metal ions using the biosynthesized Ag-NPs. The selectivity of the detection process towards hexavalent chromium ions makes them an ideal prospect for developing sensors for detecting carcinogenic chromium ions in aqueous mediums. Their activity against the gram-positive bacteria *S. aureus* has also been investigated, which opens up numerous possibilities regarding biological applications of biosynthesized Ag-NPs.

2. Experimental:

2.1. Materials

A precursor solution of $AgNO_3$ (1.0 mM) was prepared by dissolving 17 mg of $AgNO_3$ salt into 100 mL of ultrapure (UP)/deionized water (resistivity 18.2 MΩ·cm). The flowers of the *Plumeria obtusa* plant were picked from the campus of Delhi Technological University, Delhi. The metal salts nickel perchlorate hexahydrate, copper perchlorate hydrate, cobalt perchlorate hydrate, zinc perchlorate hexahydrate, aluminium perchlorate nonahydrate, potassium dichromate, cadmium chloride hydrate, lead perchlorate trihydrate, iron perchlorate hydrate, mercury nitrate monohydrate for Ni^{2+}, Cu^{2+}, Co^{2+}, Zn^{2+}, Al^{3+}, Cr^{6+}, Cd^{2+}, Pb^{2+}, Fe^{3+}, Fe^{2+} and Hg^{2+} metal ions respectively, were bought from Sigma Aldrich.

2.2. Preparing flower extract

The flowers of *Plumeria Obtusa* were plucked and put inside an oven for about 2 hrs at 90 °C for drying. The dehydrated flowers were pulverized using porcelain mortar and pestle. 1 gm of this powder was added to 20 mL UP water and stirred at 60 °C for 20 min. The resulting liquid was filtered to obtain an extract of dark yellowish color. This extract was used throughout this study for preparing samples of Ag-NPs, stored in a cool and dark environment.

2.3. Synthesizing Ag-NPs from prepared flower extract

For synthesizing Ag-NPs, the precursor solution of $AgNO_3$ (10 mL) was augmented with the prepared flower extract (0.5 mL) and stirred for 20 min at 250 rpm at 70 °C. The obtained solution was initially colorless but gradually changed to a reddish-brown tint; no further change in color intensity was observed after 60 min. A schematic representation of the preparation of plant extract and the synthesis process by green route is given in Fig. 2.

2.4. Instrumentation

Lambda™ 750 UV/VIS/NIR spectrometer from Perkin Elmer, Bruker's D-8 Advanced, Morgagni 268D, and Zetasizer Nano ZS by Malvern were employed for obtaining the absorption spectra, XRD pattern, TEM images, DLS particle size analysis, and zeta potential,

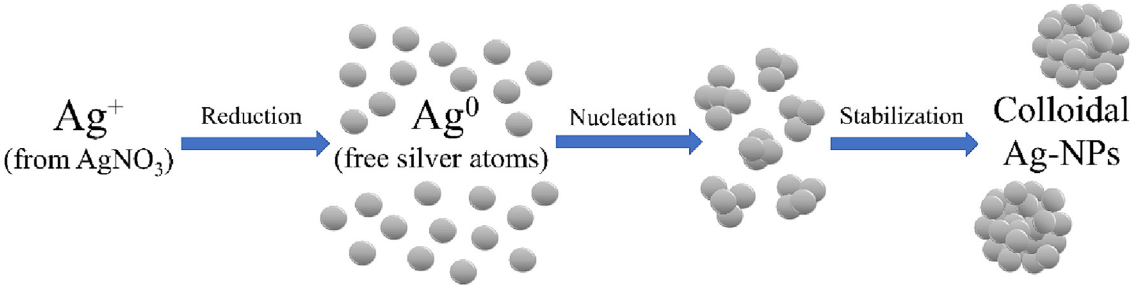


Fig. 1. A pictorial illustration of the mechanism behind biosynthesizing Ag-NPs.

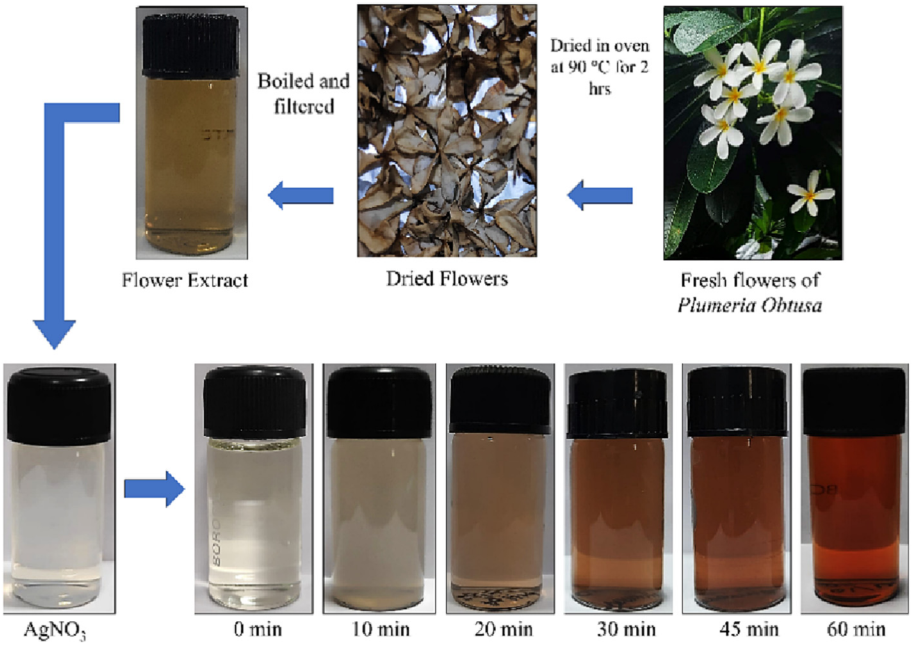


Fig. 2. Biosynthesis of Ag-NPs from prepared flower essence of *Plumeria Obtusa.*

respectively. Fourier transform infrared (FTIR) spectra were measured with a FT-IR spectrometer (Perkin Elmer Spectrum™ 3).

3. Results and discussion

The absorption spectra were recorded for the prepared flower essence, $AgNO_3$, and Ag-NPs in the range of 250–800 nm (Fig. 3a) since Ag-NPs show a characteristic peak around 430 nm corresponding to their SPR. Any broadening or shift of the SPR peak towards the red or blue end of the wavelength spectrum due to changes in extract concentration used, reaction time, pH, and temperature were studied, indicating changes in particle size or shape [43]. The same has been represented in Fig. 3 (a, b, c, d and e). A single SPR peak suggests the production of Ag-NPs of spherical conformation [44].

3.1. Influence of amount of flower extract on the synthesis

The impact of the amount of flower extract used in synthesizing Ag-NPs was inspected by taking four different concentration ratios (v/v), i.e., 0.5:20, 1.0:20, 1.5:20, 2.0:20 (mL:mL) of extract to precursor solution of $AgNO_3$. As the flower extract increases, a clear rise in absorption intensity and redshift are observed in the SPR band (Fig. 3b). The rise in absorption intensity with an increase in the amount of extract used is due to the rise in the production of Ag-NPs due to the availability of more reducing and capping agents from flower extract, and the redshift indicates an increase in particle size [45].

3.2. Influence of time

After adding flower extract to $AgNO_3$, the solution changed from colorless to a pale yellow. The gradual color change can be observed by naked eye and can be taken as a visual indicator of the synthesis process. The absorption spectra were recorded for 60 min. As time proceeds, the synthesis reaction between the phytochemicals of flower extract and $AgNO_3$ also proceeds forward, and hence, the absorption peak rises (Fig. 3c). The prepared flower extract ultimately reduced Ag^+ ions to metallic Ag^0 in about 60 min and capped them to form Ag-NPs, as no color change was observed beyond this.

3.3. Influence of temperature

The biosynthesized Ag-NPs were refrigerated (as cold as 15 °C) and heated (as hot as 55 °C) to observe how temperature variations influence the size distribution or polydispersity of Ag-NPs. Theoretically, the particle size is believed to decrease with increasing temperature because nucleation is favored at higher temperatures [46]. However, no noteworthy changes in absorption spectra were observed in this case (Fig. 3d), reflecting the high stability of prepared Ag-NPs to temperature variations.

3.4. Influence of pH

The pH of biosynthesized Ag-NPs was varied between 3 and 11 to observe the corresponding changes in the size or shape of particles. Changing the pH of the reaction alters the electrical charges present on various phytochemicals and biomolecules, which alters the rate of reduction and capping in the biosynthesis of Ag-NPs [47]. On increasing the pH from 3 to 11, a red shift is observed in the SPR peak from 416 to 445 nm (Fig. 3e), suggesting an increase in the diameter of Ag-NPs. As the particle size increases, the amount of energy necessary for exciting surface plasmons reduces, thus, causing a redshift [48]. Besides affecting particle size and the position of the SPR peak, pH variation also altered the absorption intensity. With the increase in pH, the absorption intensity also rose significantly, indicating better reduction and capping of Ag-NPs at alkaline pH values [2,49].

3.5. XRD pattern

Thin films were prepared by the drop coating method from biosynthesized Ag-NPs for XRD analysis, given in Fig. 4. Bragg's four major reflection peaks were observed at 38.19°, 46.17°, 64.96° and 76.54° representing lattice planes (111), (200), (220) and (311), confirming the lattice has a FCC

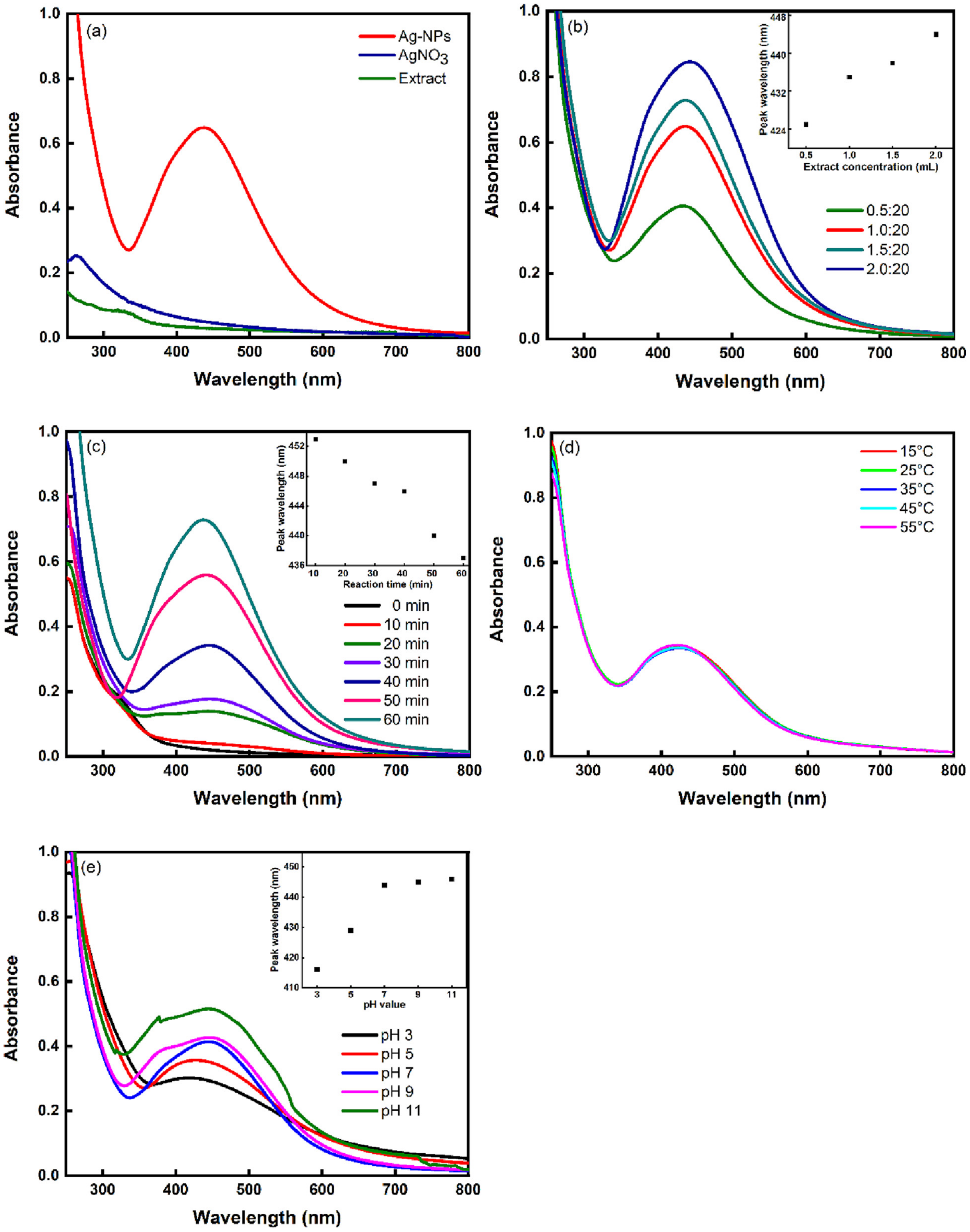


Fig. 3. Absorption spectra for colloidal Ag-NPs along with $AgNO_3$ and prepared flower extract (a) influence of variation of extract concentration (b), reaction time (c), temperature (d), and pH (e).

(face-centered cubic) structure [50]. On comparing these observed peaks with the peak positions given in the JCPDS database for bulk silver (card no. 04-0783), very tiny displacements may be observed that indicate the existence of strain on the formed crystal lattice, which is a typical attribute of nanocrystals [51]. Using obtained XRD pattern, the average crystallite size was also estimated, which came out to be 20 nm by adopting the Debye-Scherrer formula [52]. Since the crystallite size is $\sqrt{2}$ times the particle diameter for a FCC crystal; the particle diameter comes out to be ~14 nm.

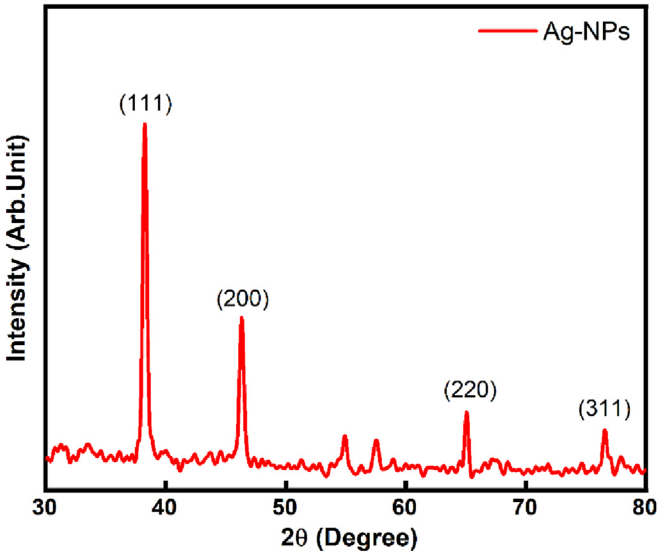


Fig. 4. XRD pattern of Ag-NPs thin film.

3.6. FTIR analysis

Chemical compositional analysis of Ag-NPs and flower extract of *Plumeria Obtusa* was studied via FTIR spectroscopy. On exposing the flower extract and Ag-NPs sample to IR radiation, the atomic vibrations in the bonds within functional groups present in the flower extract and adsorbed on the Ag-NPs surface begin to get impacted. These changes in atomic vibrations and stretching are detected as signals and peaks in the spectrum, thus, giving a structural fingerprint of the present functional groups. FTIR spectra of the flower extract and colloidal Ag-NPs are represented in Fig. 5. The flower extract of *Plumeria Obtusa* showed some major peaks due to O—H and N—H stretching at 3660 cm^{-1} and 3376 cm^{-1}, C—H and N—H stretching at 2981 and 2894 cm^{-1} and C=C stretching at 1628 cm^{-1}, respectively. These are attributed to the presence of flavonoids, triterpenoids, alkaloids, and phenolic acids [53]. Similar bands present in the spectrum of Ag-NPs indicated the adsorption of these compounds on the Ag-NPs surface [54], thus, suggesting their role in the reduction, capping and stabilizing of Ag-NPs.

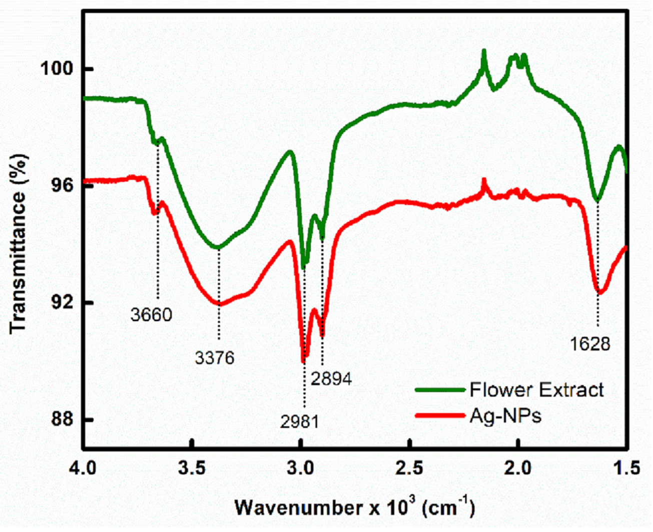


Fig. 5. FTIR spectra for flower extract and Ag-NPs.

3.7. Zeta potential and size analysis

In nanosuspensions, creating an electrical double film at the surface leads to zeta potential directly related to its potential stability, i.e., its ability to resist coagulation. Ideally, a zeta potential value of ±30 mV is considered strongly anionic or cationic [55]. The inset of Fig. 6 shows that the zeta potential of Ag-NPs comes out to be −22.7 mV (average observation of three repetitions).

A negative zeta potential value indicates that the stabilizing agents controlling the resulting morphology and diameter of Ag-NPs, are anionic in nature [56]. Fig. 6 shows the size distribution intensity fetched from DLS (approximately 71 nm). Since this is the hydrodynamic size, i.e., it includes the diameter of the core Ag-NPs plus the adsorbed biomolecules on the surface, this average particle size may be assumed to be much bigger than the actual size of Ag-NPs present in the prepared samples.

3.8. Morphology

To determine the morphological characteristics of biosynthesized Ag-NPs, TEM analysis was done. The sample of Ag-NPs was kept in a sonicator for 15 min. After sonicating, they were coated on a mesh grid made of copper for TEM analysis. The observed image of Ag-NPs is given in Fig. 7(a), from which their spherical shape can be confirmed. Subsequently, a particle size distribution curve was plotted in Fig. 7(b) to determine the average diameter of Ag-NPs to be 13 nm.

3.9. Selective sensing property towards carcinogenic Cr^{6+}

The novelty of the work relies upon the application of the biosynthesized silver nanoparticles, i.e., an ultrasensitive detection method of heavy metal ions using the biosynthesized Ag-NPs, especially hexavalent chromium. Though attempts have been made to detect the presence of hexavalent chromium ions via biosynthesized Ag-NPs, but each method has encountered limitations such as a lack of sensitivity, accuracy, low detection limit, and proper explanation of the results [57–60]. To study the potentiality of biosynthesized Ag-NPs to sense heavy metals, aqueous solutions (1 μM) of various metal ions (Ni^{2+}, Cu^{2+}, Co^{2+}, Zn^{2+}, Al^{3+}, Cr^{6+}, Cd^{2+}, Pb^{2+}, Fe^{3+}, $Fe^{2+\cdot}$ Hg^{2+}) were prepared in deionized

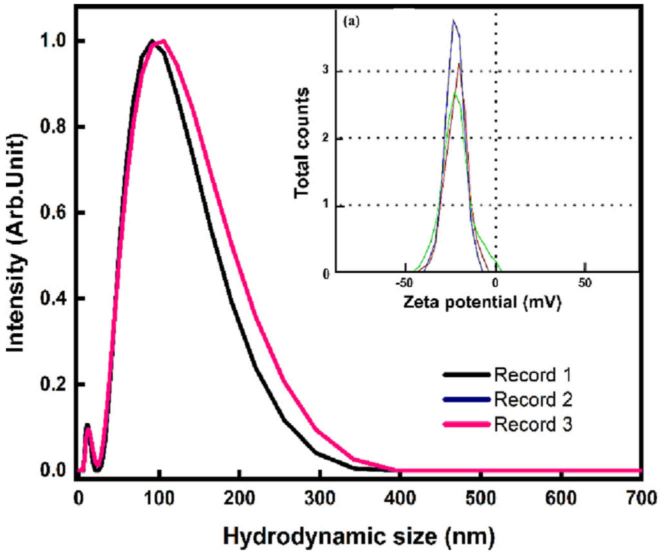


Fig. 6. Hydrodynamic size distribution by intensity repeated for three records (Records 1, 2 and 3) of colloidal Ag-NPs. Inset represents the apparent zeta potential (mV) of colloidal Ag-NPs.

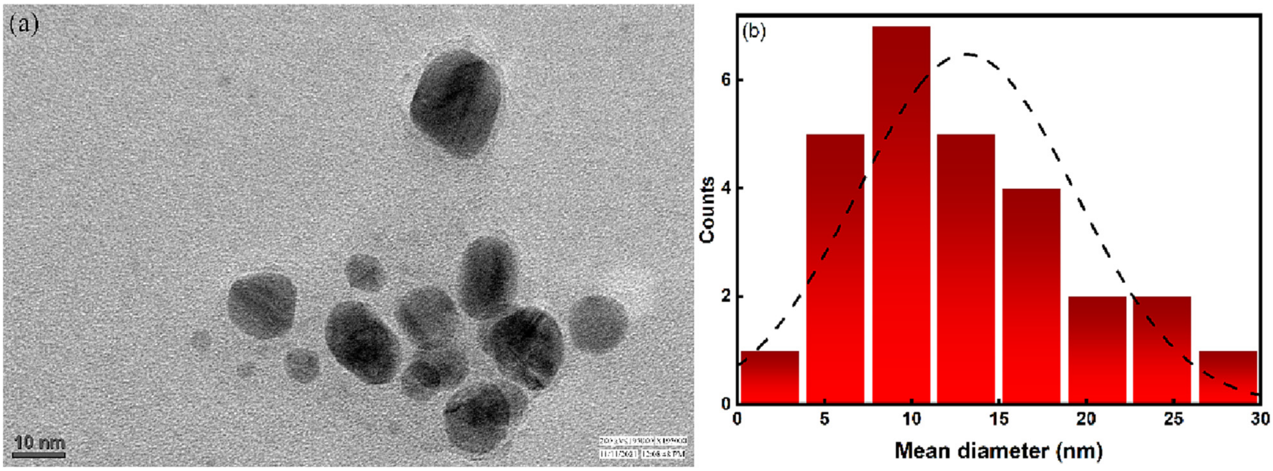


Fig. 7. TEM image (a) along with mean diameter distribution (b) of AgNPs.

water, to be added to Ag-NPs. On being added to the Ag-NPs sample, the effect on its color and absorption spectra was recorded. No visible color change was recorded for any metal ions except Cr^{6+}, as shown in Fig. 8. This selectivity makes Ag-NPs ideal for the colorimetric detection of carcinogenic Cr^{6+} in aqueous mediums.

The absorption spectra were also recorded to study consequent changes after treating Ag-NPs with different metal ions, as shown in Fig. 9(a). On the addition of Cr^{6+} in Ag-NPs, a major fall in absorbance (absorption intensity) of the SPR band of Ag-NPs is seen. Another peak rises in the absorption spectrum at 377 nm, a characteristic peak of dichromate ions. For rest metal ions, minor insignificant shifts in peak absorption intensities are recorded; a comparative bar diagram to highlight the same has been represented in the inset of Fig. 9(a). To formulate controlled selective sensing of Cr^{6+} ions, Ag-NPs were treated with different concentrations of Cr^{6+} ions (0.3 nM to 166.6 nM), and their resulting absorption spectra were recorded, as given in Fig. 9(b). The reciprocal of absorbance of the SPR peak of Cr^{6+} treated Ag-NPs was plotted as a function of the concentration of Cr^{6+} and was found to be linear with a correlation factor of 0.98, given in the inset of Fig. 9(b). The slope of this linear graph was used to calculate LoD using the formula $3.3 \times (\sigma/\text{slope})$, where σ is the standard deviation (obtained from the ten repeated absorption spectra of the pure Ag-NPs sample). The LoD came out to be 95 ± 02 pM, which is the lowest LoD ever achieved for detecting Cr^{6+} via biosynthesized nanomaterials, to the best of our knowledge. The LoD value reported in this work is in the range of picomolar, which has not been reported for any other biosynthesized silver nanoparticles yet for the significant range of concentration, 0.3 nM–166.6 nM. A comparative summary to support the same has been given in Table 1.

Section 3.4 discusses the measure of the stability of Ag-NPs in different pH conditions. Still, to experimentally verify if the sensing mechanism would work in aqueous mediums other than deionized water, the applicability of the detection method of Cr^{6+} via Ag-NPs was studied in different aqueous mediums. Two different water samples (tap water and Yamuna river water) were tested for the same to conclude that detecting Cr^{6+} via Ag-NPs works effectively in different aqueous mediums, as given in Fig. 9(c and d). The sensing mechanism works in three different aqueous mediums with different pH values, and the LoD of each is almost the same. The respective LoD values calculated for the three aqueous mediums, i.e., tap water (~6.8 pH), deionized water (~7.2 pH), and Yamuna river water (~7.7 pH), are 97, 95 and 117 pM, respectively. Considering that, there is a slight difference in the LoD value with such changes in pH value. As long as the Ag-NPs sample shows a significant SPR peak, the sensing mechanism would continue to work in different aqueous mediums having different pH values.

The proposed mechanism behind the selectivity of Ag-NPs towards sensing Cr^{6+} ions may be attributed to the contrasting electrochemical characteristics of Ag^0 and Cr^{6+}. A metal that has a greater reduction potential, i.e., Cr^{6+} (+1.33 V), can oxidize a metal with a lower reduction potential, i.e., Ag^0 (+0.80 V), whereas Cr^{2+} (−0.91 V), Cr^{3+} (−0.74 V) or any of the metals mentioned above cannot. Thus, Cr^{6+} ions cause the oxidation of Ag^0 to Ag^+. Moreover, Ag^+ ions show strong selectivity and binding affinity to dichromate $(Cr_2O_7)^{2-}$ ions, leading to the formation of $Ag_2Cr_2O_7$ [58]. This causes the SPR band of Ag-NPs to diminish.

3.10. Antibacterial action

Due to the increased use of antibiotics these days, the problem of multidrug resistance [61] has arisen, and to solve the same, new antibacterial strategies are required. Silver has been valued as an antimicrobial agent [62] since the classical ages. Since nanoparticles have greater surface area per unit volume, they facilitate contact with the bacterial surface and enhance its antibacterial action.

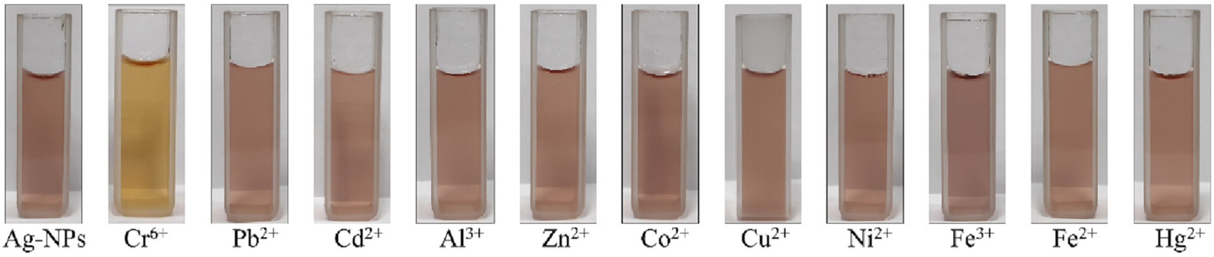


Fig. 8. Color changes in colloidal Ag-NPs after adding different metal ions (166.6 nM).

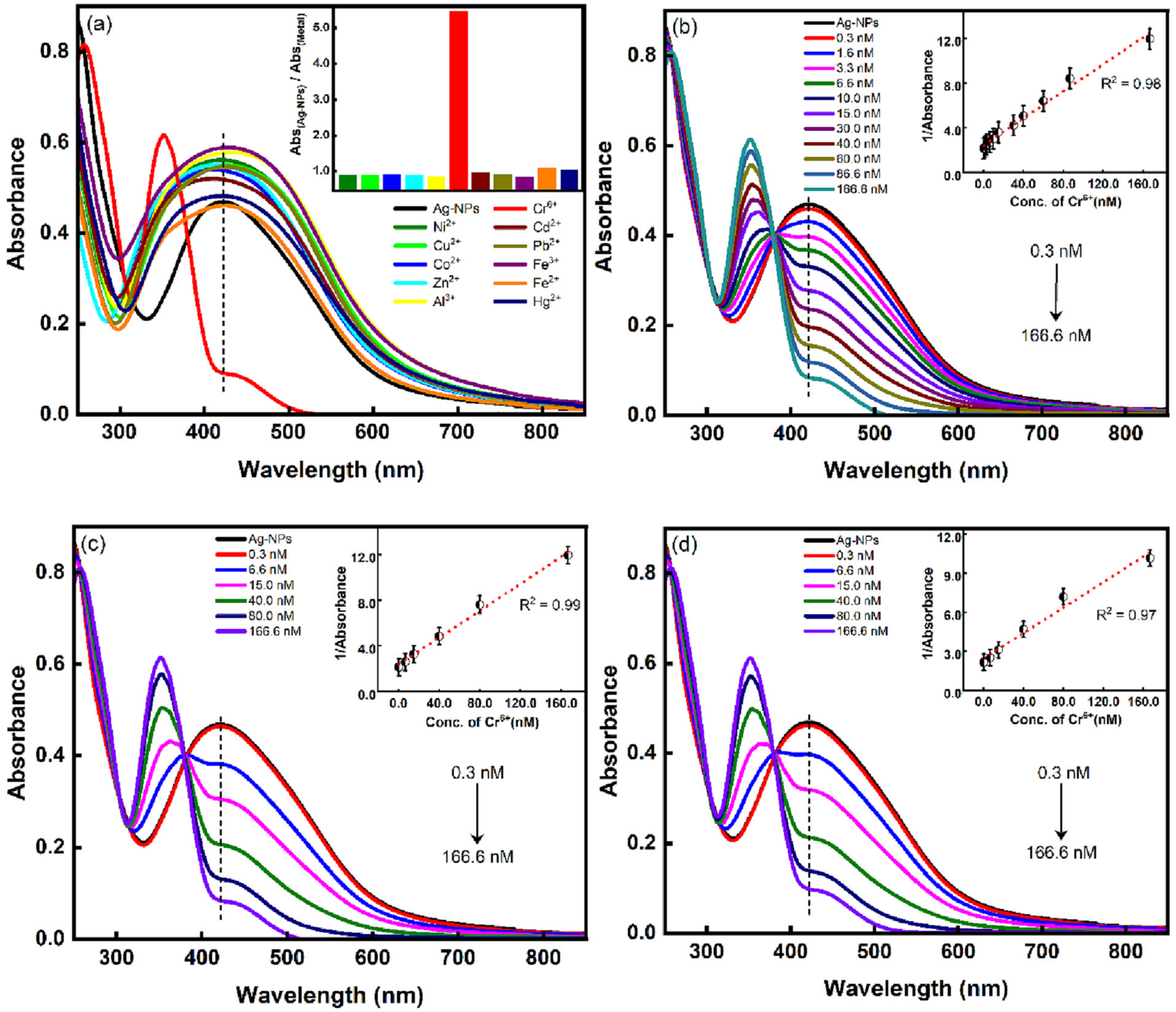


Fig. 9. Absorption spectra of colloidal Ag-NPs post addition of a fixed concentration of various metal ions, inset represents a bar diagram for the ratio of peak absorbances of different metal ions to pure Ag-NPs (a), successive addition of Cr^{6+} ions in Ag-NPs dispersed in aqueous medium of deionized water (b), tap water (c) and Yamuna river water (d). Insets of Fig. b, c, and d show the linear plots of (1/absorbance) of Cr^{6+} treated Ag-NPs as a function of the concentration of Cr^{6+}.

Table 1
Comparison of previously reported nanosensors for Cr^{6+} detection.

Probe used	Detection method	Linear Range	LoD	Reference
Biosynthesized Ag-NPs	SPR based	0.3–166.6 nM (0.09–49.1 ppb)	95 pM (0.028 ppb)	This work
Chemically synthesized Ag-NPs	SPR based	10^{-3}–10^{-9} M	1 nM	[57]
Biosynthesized Ag-NPs	SPR based	10–100 ppm	0.1 ppm	[58]
Biosynthesized Ag-NPs	Colorimetric	–	1 μM	[59]
Biosynthesized Ag-NPs	Fluorescence-based	–	1.15 nM	[60]
Tartaric acid capped Ag-NPs	SPR based	10–100 μg/L	3 μg/L	[65]
Ascorbic acid capped Ag-NPs	SPR based	7.0×10^{-8}–1.84×10^{-6} M	5×10^{-8} M	[66]
Electrodeposited Au-NPs	Voltammetric	10 μg/L–5 mg/L	5 μg/L	[67]
DTT functionalized Au-NPs	SPR based	100–600 nM	20 nM	[68]
Glutathione - CdTe QDs	Fluorescence based	0.01–1.00 μg/mL	0.008 μg/mL	[69]
Tb/acac/PAM composite NPs	Fluorescence based	5–600 ng/mL	0.8 ng/mL	[70]
Tetraphenylbenzosilole derivative (TPBS-C)	Electroluminescence based	10^{-12}–10^{-4} M	0.83 pM	[71]
Metal-organic complex of Cu(II)	Fluorescence-based	–	74.4 pM	[72]
TiO_2@Ag-NPs substrate	Surface enhanced Raman scattering	10 nm–2 μM	1.45 nM	[73]

This was investigated using the method of disk diffusion, where the antibacterial effect of Ag-NPs was compared with that of $AgNO_3$ and flower extract against a gram-positive bacteria *S. aureus*. As shown in Fig. 10, the inhibition zone obtained for Ag-NPs is much bigger than that obtained for Ag-NO_3 and flower extract, thus, proving that Ag-NPs show enhanced antibacterial properties. Table 2 shows the inhibition zones obtained after treating strains of *S. aureus* at 37 °C for 48 incubation hours.

Flower extract has some antibacterial action since *Plumeria obtusa* is a medicinal plant whose flowers show antimicrobial

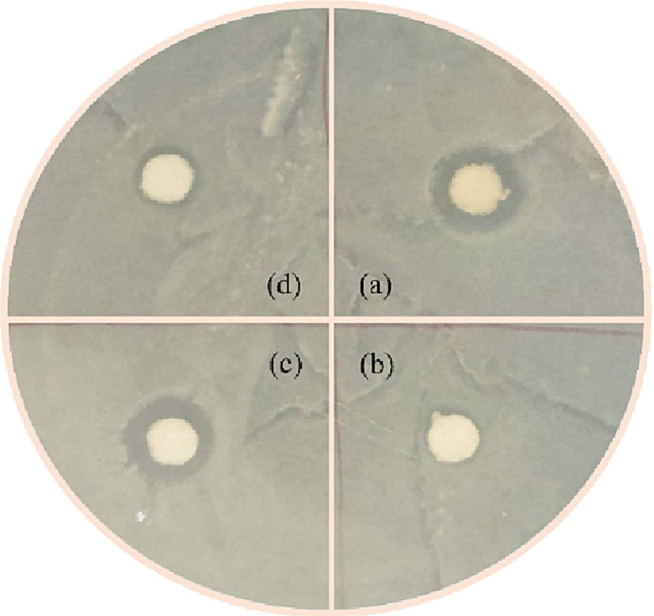


Fig. 10. Antibacterial activity of Ag-NPs (a), flower extract (b), Ag-NPs (c), and Ag-NO_3 against *S. aureus* (d).

Table 2
Inhibition zone in *S. aureus* treated with flower extract, Ag-NPs and $AgNO_3$

Sample	Inhibition Zone (mm)
Ag-NPs	10.8
Flower Extract	6.7
$AgNO_3$	7.7

activity in different solvents [63]. However, this is much less than the antibacterial action observed in the case of Ag-NPs. Similarly, $AgNO_3$ also shows some antibacterial activity. Metallic or bulk silver is mostly inert, but the antibacterial action of $AgNO_3$ is accredited to the very high reactivity and binding affinity of ionized Ag particles to the proteins, DNA, and RNA of bacterial cells, thus causing their malfunction. The mechanism behind the enhanced antibacterial activity of Ag-NPs is along these lines, Ag-NPs interact with protein and DNA, penetrate the cell wall and reach the center of the bacterial cells. To protect the DNA in nucleus, reactive oxygen, and nitrogen species start to agglomerate and give rise to high oxidative stress in many cell parts, including the nucleus [14]. Ag-NPs strike the cell division and chain of respiration of the bacteria, leading to complete cell rupture [64].

4. Conclusion

Biocompatible spherical Ag-NPs of the mean diameter of 13 nm were successfully synthesized, where flower extract of *Plumeria Obtusa* served as reducer and stabilizer. The procured Ag-NPs were characterized and confirmed using different characterization techniques. The synthesis yield was significantly dependent on various physicochemical criteria such as pH, temperature, time, and added extract amount. Biosynthesized Ag-NPs were very sensitive to Cr^{6+} ions in aqueous mediums with a very efficient LoD of 95 ± 02 pM (recorded at pH $\sim$ 7.2). The synthesized Ag-NPs also enhanced the antibacterial action over $AgNO_3$ countered to a gram-positive bacterium, *S. aureus* and showed an inhibition zone of 10.8 mm. Along with having unique optical and chemical properties, low cost, one-step simple eco-friendly synthesis procedure, and biocompatibility, the synthesized Ag-NPs show excellent selectivity towards detecting a carcinogen (Cr^{6+}) and appreciable antibacterial action (against *S. aureus*), proposing to be highly useful in the fields of biosensing and biomedicine.

CRediT authorship contribution statement

Samiksha Shukla: Formal analysis, Data curation, Investigation, Writing - original draft. **Mohan Singh Mehata:** Resources, Funding acquisition, Supervision, Conceptualization, Writing - review & editing.

Data availability

All the data generated or analyzed in this study are included in this article.

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.

Acknowledgement

Author SS sincerely thanks the Laser-spectroscopy Lab members for their help and especially Mr. Aryan for his help in executing the antibacterial activity.

References

[1] M.S. Mehata, Green route synthesis of silver nanoparticles using plants/ginger extracts with enhanced surface plasmon resonance and degradation of textile dye, Mater. Sci. Eng. B 273 (2021) 115418–115427, https://doi.org/10.1016/j.mseb.2021.115418.
[2] S. Jain, M.S. Mehata, Medicinal plant leaf extract and pure flavonoid mediated green synthesis of silver nanoparticles and their enhanced antibacterial property, Sci. Rep. 7 (2017) 15867, https://doi.org/10.1038/s41598-017-15724-8.
[3] Aryan, Ruby, M.S. Mehata, Green synthesis of silver nanoparticles using *Kalanchoe pinnata* leaves (life plant) and their antibacterial and photocatalytic activities, Chem. Phys. Lett. 778 (2021) 138760–138770, http://dx.doi.org/10.1016/j.cplett.2021.138760.
[4] P. Sharma, M.K. Singh, M.S. Mehata, Sunlight-driven MoS_2 nanosheets mediated degradation of dye (crystal violet) for wastewater treatment, J. Mol. Struct. 1249 (2022) 131651–131662, https://doi.org/10.1016/j.molstruc.2021.131651.
[5] P. Sharma, M.S. Mehata, Rapid sensing of lead metal ions in an aqueous medium by MoS_2 quantum dots fluorescence turn-off, Mater. Res. Bull. 131 (2020), https://doi.org/10.1016/j.materresbull.2020.110978.
[6] M.K. Singh, M.S. Mehata, Enhanced photoinduced catalytic activity of transition metal ions incorporated TiO_2 nanoparticles for degradation of organic dye: absorption and photoluminescence spectroscopy, Opt. Mater. 109 (2020) 110309–110317, https://doi.org/10.1016/j.optmat.2020.110309.
[7] M.K. Singh, M.S. Mehata, Phase-dependent optical and photocatalytic performance of synthesized titanium dioxide (TiO_2) nanoparticles, Optik 193 (2019) 163011–163022, https://doi.org/10.1016/j.ijleo.2019.163011.
[8] A. Verma, M.S. Mehata, Controllable synthesis of silver nanoparticles using Neem leaves and their antimicrobial activity, J. Radiat. Res. Appl. Sci. 9 (2016) 109–115, https://doi.org/10.1016/j.jrras.2015.11.001.
[9] A. Shenava, Synthesis of silver nanoparticles by chemical reduction method and their antifungal activity, Int. Res. J. Pharm. 4 (2013) 111–113, https://doi.org/10.7897/2230-8407.041024.
[10] S. Iravani, H. Korbekandi, S.V. Mirmohammadi, B. Zolfaghari, Synthesis of silver nanoparticles: chemical, physical and biological methods, Res. Pharm. Sci. 9 (2014) 385–406. http://www.ncbi.nlm.nih.gov/pubmed/26339255.
[11] S. Medici, M. Peana, V.M. Nurchi, M.A. Zoroddu, Medical uses of silver: history, myths, and scientific evidence, J. Med. Chem. 62 (2019) 5923–5943, https://doi.org/10.1021/acs.jmedchem.8b01439.
[12] F.A. Alharthi, A.A. Alghamdi, N. Al-Zaqri, H.S. Alanazi, A.A. Alsyahi, A. El Marghany, N. Ahmad, Facile one-pot green synthesis of Ag-ZnO Nanocomposites using potato peel and their Ag concentration dependent photocatalytic properties, Sci. Rep. 10 (2020) 20229, https://doi.org/10.1038/s41598-020-77426-y.
[13] M. Parthibavarman, S. Bhuvaneshwari, M. Jayashree, R. Boopathi Raja, Green synthesis of silver (Ag) nanoparticles using extract of apple and grape and with

enhanced visible light photocatalytic activity, Bionanoscience 9 (2019) 423–432, https://doi.org/10.1007/s12668-019-0605-0.

[14] Aryan Ruby, M.S. Mehata, Surface plasmon resonance allied applications of silver nanoflowers synthesized from: *Breynia vitis-idaea* leaf extract, Dalt. Trans. 51 (2022) 2726–2736, https://doi.org/10.1039/d1dt03592d.

[15] S. Shrivastava, T. Bera, S.K. Singh, G. Singh, P. Ramachandrarao, D. Dash, Characterization of antiplatelet properties of silver nanoparticles, ACS Nano 3 (2009) 1357–1364, https://doi.org/10.1021/nn900277t.

[16] B. Pant, P.S. Saud, M. Park, S.-J. Park, H.-Y. Kim, General one-pot strategy to prepare Ag–TiO_2 decorated reduced graphene oxide nanocomposites for chemical and biological disinfectant, J. Alloy. Compd. 671 (2016) 51–59, https://doi.org/10.1016/j.jallcom.2016.02.067.

[17] 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.

[18] P. Le Thi, Y. Lee, H.J. Kwon, K.M. Park, M.H. Lee, J.-C. Park, K.D. Park, Tyrosinase-mediated surface coimmobilization of heparin and silver nanoparticles for antithrombotic and antimicrobial activities, ACS Appl. Mater. Interfaces 9 (2017) 20376–20384, https://doi.org/10.1021/acsami.7b02500.

[19] H. Agarwal, S.V. Kumar, S. Rajeshkumar, Antidiabetic effect of silver nanoparticles synthesized using lemongrass (*Cymbopogon citratus*) through conventional heating and microwave irradiation approach, J. Microbiol. Biotechnol. Food Sci. 7 (2018) 371–376, https://doi.org/10.15414/jmbfs.2018.7.4.371-376.

[20] E. Liu, M. Zhang, H. Cui, J. Gong, Y. Huang, J. Wang, Y. Cui, W. Dong, L. Sun, H. He, V.C. Yang, Tat-functionalized Ag-Fe_3O_4 nano-composites as tissue-penetrating vehicles for tumor magnetic targeting and drug delivery, Acta Pharm. Sin. B 8 (2018) 956–968, https://doi.org/10.1016/j.apsb.2018.07.012.

[21] H. Gao, Progress and perspectives on targeting nanoparticles for brain drug delivery, Acta Pharm. Sin. B 6 (2016) 268–286, https://doi.org/10.1016/j.apsb.2016.05.013.

[22] G. Yamal, P. Sharmila, K.S. Rao, P. Pardha-Saradhi, Inbuilt potential of YEM medium and its constituents to generate Ag/Ag_2O nanoparticles, PLoS One 8 (2013) e61750–e61760, https://doi.org/10.1371/journal.pone.0061750.

[23] Y. Park, Y.N. Hong, A. Weyers, Y.S. Kim, R.J. Linhardt, Polysaccharides and phytochemicals: a natural reservoir for the green synthesis of gold and silver nanoparticles, IET Nanobiotechnol. 5 (2011) 69–78, https://doi.org/10.1049/iet-nbt.2010.0033.

[24] Y. Zhou, W. Lin, J. Huang, W. Wang, Y. Gao, L. Lin, Q. Li, L. Lin, M. Du, Biosynthesis of gold nanoparticles by foliar broths: roles of biocompounds and other attributes of the extracts, Nanoscale Res. Lett. 5 (2010) 1351–1359, https://doi.org/10.1007/s11671-010-9652-8.

[25] S.S. Shankar, A. Rai, B. Ankamwar, A. Singh, A. Ahmad, M. Sastry, Biological synthesis of triangular gold nanoprisms, Nat. Mater. 3 (2004) 482–488, https://doi.org/10.1038/nmat1152.

[26] J. Huang, Q. Li, D. Sun, Y. Lu, Y. Su, X. Yang, H. Wang, Y. Wang, W. Shao, N. He, J. Hong, C. Chen, Biosynthesis of silver and gold nanoparticles by novel sundried *Cinnamomum camphora* leaf, Nanotechnology 18 (2007), https://doi.org/10.1088/0957-4484/18/10/105104.

[27] A. Bala, G. Rani, A review on phytosynthesis, affecting factors and characterization techniques of silver nanoparticles designed by green approach, Int. Nano Lett. 10 (2020) 159–176, https://doi.org/10.1007/s40089-020-00309-7.

[28] S. Semenya, M. Potgieter, L. Erasmus, Ethnobotanical survey of medicinal plants used by Bapedi healers to treat diabetes mellitus in the Limpopo Province, South Africa, J. Ethnopharmacol. 141 (2012) 440–445, https://doi.org/10.1016/j.jep.2012.03.008.

[29] T. Bihani, P. Tandel, J. Wadekar, *Plumeria obtusa* L.: a systematic review of its traditional uses, morphology, phytochemistry and pharmacology, Phytomedicine Plus 1 (2021), https://doi.org/10.1016/j.phyplu.2021.100052.

[30] M. Saleem, N. Akhtar, N. Riaz, M.S. Ali, A. Jabbar, Isolation and characterization of secondary metabolites from *Plumeria obtusa*, J. Asian Nat. Prod. Res. 13 (2011) 1122–1127, https://doi.org/10.1080/10286020.2011.618452.

[31] O. Kaisoon, S. Siriamornpun, N. Weerapreeyakul, N. Meeso, Phenolic compounds and antioxidant activities of edible flowers from Thailand, J. Funct. Foods 3 (2011) 88–99, https://doi.org/10.1016/j.jff.2011.03.002.

[32] J. John, C.T. Aravindakumar, S. Thomas, Green synthesis of silver nanoparticles using phyto-constituents of *Ficus Auriculata Lour.* leaf extract: mechanistic approach, SAJ Biotechnol 4 (2018).

[33] H.S. Kim, Y.J. Kim, Y.R. Seo, An overview of carcinogenic heavy metal: molecular toxicity mechanism and prevention, J. Cancer Prev. 20 (2015) 232–240, https://doi.org/10.15430/JCP.2015.20.4.232.

[34] A.D. Dayan, A.J. Paine, Mechanisms of chromium toxicity, carcinogenicity and allergenicity: review of the literature from 1985 to 2000, Hum. Exp. Toxicol. 20 (2001) 439–451, https://doi.org/10.1191/096032701682693062.

[35] I.B. Weinstein, The scientific basis for carcinogen detection and primary cancer prevention, Cancer 47 (1981) 1133–1141, https://doi.org/10.1002/1097-0142(19810301)47:5+<1133::aid-cncr2820471312>3.0.co;2-7.

[36] J. Ashby, R.S. Morrod, Detection of human carcinogens, Nature 352 (1991) 185–186, https://doi.org/10.1038/352185a0.

[37] R. Nusko, K.G. Heumann, Cr(III)/Cr(VI) speciation in aerosol particles by extractive separation and thermal ionization isotope dilution mass spectrometry, Fresenius, J. Anal. Chem. 357 (1997) 1050–1055, https://doi.org/10.1007/s002160050303.

[38] M. Sperling, S. Xu, B. Welz, Determination of chromium (III) and chromium (VI) in water using flow injection on-line preconcentration with selective adsorption on activated alumina and flame atomic absorption spectrometric detection, Anal. Chem. 64 (1992) 3101–3108, https://doi.org/10.1021/ac00048a007.

[39] S. Balasubramanian, V. Pugalenthi, Determination of total chromium in tannery waste water by inductively coupled plasma-atomic emission spectrometry, flame atomic absorption spectrometry and UV-visible spectrophotometric methods, Talanta 50 (1999) 457–467, https://doi.org/10.1016/S0039-9140(99)00135-6.

[40] E.J. Arar, J.D. Pfaff, Determination of dissolved hexavalent chromium in industrial wastewater effluents by ion chromatography and post-column derivatization with diphenylcarbazide, J. Chromatogr. A 546 (1991) 335–340, https://doi.org/10.1016/S0021-9673(01)93031-6.

[41] N.D. Nguyen, T. Van Nguyen, A.D. Chu, H.V. Tran, L.T. Tran, C.D. Huynh, A label-free colorimetric sensor based on silver nanoparticles directed to hydrogen peroxide and glucose, Arab. J. Chem. 11 (2018) 1134, https://doi.org/10.1016/j.arabjc.2017.12.035.

[42] H.K. Choi, M.-J. Lee, S.N. Lee, T.-H. Kim, B.-K. Oh, Noble metal nanomaterial-based biosensors for electrochemical and optical detection of viruses causing respiratory illnesses, Front. Chem. 9 (2021) 605, https://doi.org/10.3389/fchem.2021.672739.

[43] S.L. Smitha, K.M. Nissamudeen, D. Philip, K.G. Gopchandran, Studies on surface plasmon resonance and photoluminescence of silver nanoparticles, Spectrochim. Acta A: Mol. Biomol. Spectrosc. 71 (2008) 186–190, https://doi.org/10.1016/j.saa.2007.12.002.

[44] B.J. Wiley, S.H. Im, Z.Y. Li, J. McLellan, A. Siekkinen, Y. Xia, Maneuvering the surface plasmon resonance of silver nanostructures through shape-controlled synthesis, J. Phys. Chem. B 110 (2006) 15666–15675, https://doi.org/10.1021/jp0608628.

[45] K.C. Lee, S.J. Lin, C.H. Lin, C.S. Tsai, Y.J. Lu, Size effect of Ag nanoparticles on surface plasmon resonance, Surf. Coatings Technol. 202 (2008) 5339–5342, https://doi.org/10.1016/j.surfcoat.2008.06.080.

[46] S. Shukla, A. Masih, Aryan, M.S. Mehata, Catalytic activity of silver nanoparticles synthesized using *Crinum asiaticum* (Sudarshan) leaf extract, Mater. Today:. Proc. 56 (2022) 3714–3720, https://doi.org/10.1016/j.matpr.2021.12.468.

[47] M.M.H. Khalil, E.H. Ismail, K.Z. El-Baghdady, D. Mohamed, Green synthesis of silver nanoparticles using olive leaf extract and its antibacterial activity, Arab. J. Chem. 7 (2014) 1131–1139, https://doi.org/10.1016/j.arabjc.2013.04.007.

[48] O.A. Yeshchenko, I.M. Dmitruk, A.A. Alexeenko, A.V. Kotko, J. Verdal, A.O. Pinchuk, Size and temperature effects on the surface Plasmon resonance in silver nanoparticles, Plasmonics 7 (2012) 685–694, https://doi.org/10.1007/s11468-012-9359-z.

[49] I. Fernando, Y. Zhou, Impact of pH on the stability, dissolution and aggregation kinetics of silver nanoparticles, Chemosphere 216 (2019) 297–305, https://doi.org/10.1016/j.chemosphere.2018.10.122.

[50] R.I. Priyadharshini, G. Prasannaraj, N. Geetha, P. Venkatachalam, Microwave-mediated extracellular synthesis of metallic silver and zinc oxide nanoparticles using macro-algae (*Gracilaria edulis*) extracts and its anticancer activity against human PC3 cell lines, Appl. Biochem. Biotechnol. 174 (2014) 2777–2790, https://doi.org/10.1007/s12010-014-1225-3.

[51] M.J. Hÿtch, A.M. Minor, Observing and measuring strain in nanostructures and devices with transmission electron microscopy, MRS Bull. 39 (2014) 138–146, https://doi.org/10.1557/mrs.2014.4.

[52] A.O. Bokuniaeva, A.S. Vorokh, Estimation of particle size using the Debye equation and the Scherrer formula for polyphasic TiO_2 powder, J. Phys. Conf. Ser. 1410 (2019), https://doi.org/10.1088/1742-6596/1410/1/012057.

[53] R.N. Oliveira, M.C. Mancini, F.C.S. de Oliveira, T.M. Passos, B. Quilty, R.M. da, S. M Thiré, G.B. McGuinness, FTIR analysis and quantification of phenols and flavonoids of five commercially available plants extracts used in wound healing, Matéria (Rio J.) 21 (2016) 767–779, https://doi.org/10.1590/S1517-707620160003.0072.

[54] K.H. Oh, V. Soshnikova, J. Markus, Y.J. Kim, S.C. Lee, P. Singh, V. Castro-Aceituno, S. Ahn, D.H. Kim, Y.J. Shim, Y.J. Kim, D.C. Yang, Biosynthesized gold and silver nanoparticles by aqueous fruit extract of *Chaenomeles sinensis* and screening of their biomedical activities, Artif. Cells, Nanomed. Biotechnol. 46 (2018) 599–606, https://doi.org/10.1080/21691401.2017.1332636.

[55] S. Pattanayak, M.M.R. Mollick, D. Maity, S. Chakraborty, S.K. Dash, S. Chattopadhyay, S. Roy, D. Chattopadhyay, M. Chakraborty, *Butea monosperma bark* extract mediated green synthesis of silver nanoparticles: characterization and biomedical applications, J. Saudi Chem. Soc. 21 (2017) 673–684, https://doi.org/10.1016/j.jscs.2015.11.004.

[56] S. Singh, A. Bharti, V.K. Meena, Structural, thermal, zeta potential and electrical properties of disaccharide reduced silver nanoparticles, J. Mater. Sci. Mater. Electron. 25 (2014) 3747–3752, https://doi.org/10.1007/s10854-014-2085-x.

[57] A. Ravindran, M. Elavarasi, T.C. Prathna, A.M. Raichur, N. Chandrasekaran, A. Mukherjee, Selective colorimetric detection of nanomolar Cr (VI) in aqueous solutions using unmodified silver nanoparticles, Sensors Actuators, B Chem. 166–167 (2012) 365–371, https://doi.org/10.1016/j.snb.2012.02.073.

[58] M. Ismail, M.I. Khan, K. Akhtar, M.A. Khan, A.M. Asiri, S.B. Khan, Biosynthesis of silver nanoparticles: a colorimetric optical sensor for detection of hexavalent chromium and ammonia in aqueous solution, Phys. E Low-Dimensional Syst. Nanostruct. 103 (2018) 367–376, https://doi.org/10.1016/j.physe.2018.06.015.

[59] C.K. Balavigneswaran, T.S.J. Kumar, R.M. Packiaraj, S. Prakash, Rapid detection of Cr(VI) by AgNPs probe produced by *Anacardium occidentale* fresh leaf

extracts, Appl. Nanosci. 4 (2014) 367–378, https://doi.org/10.1007/s13204-013-0203-3.

[60] F.A.M. Alahdal, M.T.A. Qashqoosh, Y.K. Manea, M.A.S. Salem, R.H. Khan, S. Naqvi, Ultrafast fluorescent detection of hexavalent chromium ions, catalytic efficacy and antioxidant activity of green synthesized silver nanoparticles using leaf extract of *P. austroarabica*, Environ. Nanotechnology, Monit. Manag. 17 (2022) 100665, https://doi.org/10.1016/j.enmm.2022.100665.

[61] S. Parmar, H. Kaur, J. Singh, A.S. Matharu, S. Ramakrishna, M. Bechelany, Recent advances in green synthesis of AgNPs for extenuating antimicrobial resistance, Nanomaterials 12 (2022) 1115, https://doi.org/10.3390/nano12071115.

[62] M. Riaz, U. Sharafat, N. Zahid, M. Ismail, J. Park, B. Ahmad, N. Rashid, M. Fahim, M. Imran, A. Tabassum, Synthesis of biogenic silver nanocatalyst and their antibacterial and organic pollutants reduction ability, ACS Omega 7 (2022) 14723–14734, https://doi.org/10.1021/acsomega.1c07365.

[63] N. Ali, D. Ahmad, J. Bakht, Antimicrobial activity of different solvent extracted samples from the flowers of medicinally important *Plumeria obstusa*, Pak. J. Pharm. Sci. 28 (2015) 195–200. http://www.ncbi.nlm.nih.gov/pubmed/25553696.

[64] M. Rai, A. Yadav, A. Gade, Silver nanoparticles as a new generation of antimicrobials, Biotechnol. Adv. 27 (2009) 76–83, https://doi.org/10.1016/j.biotechadv.2008.09.002.

[65] K. Shrivas, S. Sahu, G.K. Patra, N.K. Jaiswal, R. Shankar, Localized surface plasmon resonance of silver nanoparticles for sensitive colorimetric detection of chromium in surface water, industrial waste water and vegetable samples, Anal. Methods 8 (2016) 2088–2096, https://doi.org/10.1039/c5ay03120f.

[66] X. Wu, Y. Xu, Y. Dong, X. Jiang, N. Zhu, Colorimetric determination of hexavalent chromium with ascorbic acid capped silver nanoparticles, Anal. Methods 5 (2013) 560–565, https://doi.org/10.1039/c2ay25989c.

[67] G. Liu, Y.Y. Lin, H. Wu, Y. Lin, Voltammetric detection of Cr(VI) with disposable screen-printed electrode modified with gold nanoparticles, Environ. Sci. Tech. 41 (2007) 8129–8134, https://doi.org/10.1021/es071726z.

[68] F. Tan, X. Liu, X. Quan, J. Chen, X. Li, H. Zhao, Selective detection of nanomolar Cr(vi) in aqueous solution based on 1,4-dithiothreitol functionalized gold nanoparticles, Anal. Methods 3 (2011) 343–347, https://doi.org/10.1039/c0ay00534g.

[69] L. Zhang, C. Xu, B. Li, Simple and sensitive detection method for chromium(VI) in water using glutathione-capped CdTe quantum dots as fluorescent probes, Microchim. Acta 166 (2009) 61–68, https://doi.org/10.1007/s00604-009-0164-0.

[70] L. Wang, G. Bian, L. Dong, T. Xia, S. Hong, H. Chen, Selective fluorescence determination of chromium (VI) in water samples with terbium composite nanoparticles, Spectrochim. Acta - Part A: Mol. Biomol. Spectrosc. 65 (2006) 123–126, https://doi.org/10.1016/j.saa.2005.09.042.

[71] J. Guo, W. Feng, P. Du, R. Zhang, J. Liu, Y. Liu, Z. Wang, X. Lu, Aggregation-induced electrochemiluminescence of tetraphenylbenzosilole derivatives in an aqueous phase system for ultrasensitive detection of hexavalent chromium, Anal. Chem. 92 (2020) 14838–14845, https://doi.org/10.1021/acs.analchem.0c03709.

[72] R.K. Mondal, S. Dhibar, P. Mukherjee, A.P. Chattopadhyay, R. Saha, B. Dey, Selective picomolar level fluorometric sensing of the Cr(VI)-oxoanion in a water medium by a novel metal-organic complex, RSC Adv. 6 (2016) 61966–61973, https://doi.org/10.1039/c6ra12819j.

[73] W. Zhou, B.-C. Yin, B.-C. Ye, Highly sensitive surface-enhanced Raman scattering detection of hexavalent chromium based on hollow sea urchin-like TiO_2@Ag nanoparticle substrate, Biosens. Bioelectron. 87 (2017) 187–194, https://doi.org/10.1016/j.bios.2016.08.036.

Semi-Flexible Diversified Circularly Polarized Millimeter-Wave MIMO Antenna for Wearable Biotechnologies

Rakesh N. Tiwari, Vikrant Kaim, *Member, IEEE,* Prabhakar Singh, Taimoor Khan, *Sr. Member, IEEE and* Binod Kumar Kanaujia, *Sr. Member, IEEE*

***Abstract*—This article presents a novel semi-flexible two port multi-input-multi-output (MIMO) antenna with dual circular polarization (CP) in the millimeter-wave frequency band (24-31 GHz) for wearable applications in biotechnologies. The radiating circular patch and ground plane are embedded with multiple sectoral slots and moon-shaped slots, respectively, for broad impedance bandwidth and CP property. The E_θ and E_ϕ fields are also classified as RHCP and LHCP at 27.10 GHz and 29.53 GHz, respectively. The microstrip line feeds the MIMO antenna, which is printed on a semi-flexible substrate with an overall electrical length of $2.9\lambda_0 \times 1.8\lambda_0 \times 0.04\lambda_0$ at 24 GHz. The antenna is designed for homogeneous space conditions (off-body) with a flat profile. Its on-body performance capability top-up with bending analysis is examined on the surface (chest, hand, and leg) of the anatomical Gustav human body model. Further, the MIMO diversity parameters are also evaluated in the CST Microwave Studio Simulator. Later on, the simulated radiation characteristics are testified through in-vitro measurement of the fabricated MIMO antenna in free-space and chest mimicking gel-based phantom. Lastly, to ensure human shielding, the SAR analysis is mapped with the communication link margin to select the input power.**

***Index Terms*— Broadband antenna, circular polarization, LHCP-RHCP, MIMO antenna, MIMO diversity, millimeter-wave, on-body antenna, semi-flexible, uplink-downlink communication.**

I. Introduction

THE wearable wireless technologies are a growing interest for many researchers globally due to their various body-centric applications in the medical/healthcare field, navigation, infotainment, sports, armed forces, Internet-of-Things (IoT), etc. [1]–[5]. The broad applicability of such wearable technologies led to tremendous growth in device usage from 20 million (in 2015) to 187.2 million (in 2020) [6]. Wearable on-body communication devices and microstrip patch antennas are the first choices of the research fraternity due to their conformability, lightweight, convenient design, robust profile, and ease to customize [7], [8].

The flexibility of wearable antenna is also one of the desirable characteristics of its easy placement on curved body surfaces [9]. In [10]–[12], liquid materials, elastomeric PDMS substrate, conductive fabrics, and mesh-like structures are utilized to obtain conformability for the on-body antennas. Although these materials facilitate the person's mobility, such techniques still render the cumbersome integration and soldering of the antenna components. Also, the bending effects of the antenna significantly deteriorate the communication link robustness due to reflections and scatterings of the multipath EM waves while propagating on the curved body surface [13], [14]. The dent in the reliable communication link dramatically reduces the data rate.

As the bending issues in wearable antennas are inevitable, in view of the same, the multiple-input-multiple-output (MIMO) antennas can overcome this menace. Nevertheless, compact wearable antennas can compromise the mutual coupling between the antenna elements, and the MIMO designs with improved isolations are generally recommended for enhanced performance. There are many isolation improvement techniques reported in the literature [15]–[21] but do not fit for compact on-body antennas. The optimal placement of the antenna elements in the MIMO configuration can achieve the desirable miniaturization in the antenna structure [22]. A linearly polarized 2-port MIMO antenna design is reported in [23] for wearable applications in the lower ISM band. This design utilizes the ground plane as radiating element with an impedance bandwidth of 20%. In [24], a two-port wearable MIMO antenna design is reported on the jeans substrate. This

R. N. Tiwari is with the Department of Electronics and Communication Engineering, Madanapalle Institute of Technology & Science, Madanapalle, Andhra Pradesh-517325, India (e-mail: srakeshnath@gmail.com).

V. Kaim, and B. K. Kanaujia are with the School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi–110067,India, and B. K. Kanaujia is also with the Dr. B.R. Ambedkar National Institute of Technology, Jalandhar (Punjab)–144011, India (e-mail: vikran16_sit@jnu.ac.in, bkkanaujia@ieee.org).

P. Singh is with the Division of Physics, School of Basic and Applied Sciences, Galgotias University, Greater Noida, Uttar Pradesh–203201, India (e-mail: prabhakarsingh3@gmail.com).

T. Khan is with the Department of Electronics and Communication Engineering, NIT Silchar, Assam, India (e-mail: ktaimoor@gmail.com).

TABLE I
COMPARISON OF THE PROPOSED ANTENNA WITH THE REPORTED ANTENNAS

References (Year)	[6] **(2017)**	[25] **(2018)**	[26] **(2019)**	[27] **(2020)**	[28] **(2020)**	**This work (2022)**
Antenna type	Microstrip Loop	Microstrip Patch	Microstrip EBG	Circular Patch	Microstrip Patch	Microstrip Patch
Profile	Planar (SISO)	Planar (SISO)	Planar (MIMO)	Planar (SISO)	Planar (SISO)	Planar (MIMO)
Substrate layer	Single	Single	Dual	Three	Single	Single
Conformal	×	√	√	×	√	√
Bending analysis	×	√	√	×	×	√
Elements	*NA*	*NA*	2	*NA*	*NA*	2
Frequency (GHz)	60	28, 36	24	69.6	28	27.10, 29.53
Polarization	Linear	Linear	Linear	Linear	Linear	RHCP, LHCP
Area (mm^2)	14 × 10.5	11 × 12	27 × 27	3.5 × 3	14.96 × 17.45	36 × 22.5
Volume (mm^3)	169.05	13.2	370.33	11.445	130.53	411.48
Sim. Enclosure	On-body	On-body	On-body	On-body	On body	On body
ARBW (GHz)	×	×	×	×	×	0.35, 0.94
On-body gain (dBi)	12.0	8.0-8.76	6.0	6.6 (peak)	7.4	6.1
Rad. Efficiency (%)	63	90	80.5	-	80	> 80
Application	Biomedical	Biomedical	Biomedical	Biomedical	Non biomedical (5G)	Biomedical
1-g SAR (W/Kg)	–	–	–	–	–	1.64, 2.18 (peak)
Link Margin	×	×	×	×	×	Uplink / Downlink

EBG: Electromagnetic Bandgap ; *NA*: *not applicable;* –: *not given*

antenna covers the frequency range from 2.74 to 12.33 GHz. A high impedance surface-based MIMO antenna is presented in [29] with an operating frequency from 2.4 to 2.49 GHz. It is noted that the MIMO designs aforementioned here demonstrate linear polarization characteristics. However, the MIMO designs with circular polarization (CP) increase the communication robustness and reduce the polarization loss and mitigate multipath interference, especially during the mobility of the person wearing the wearable device [30]. A CP wearable antenna in [9] is designed using flexible polydimethylsiloxane (PDMS) and silver nanowires. In [31], [32], wideband wearable CP filtering antennas are reported. In [22], a wideband CP MIMO antenna is designed for wearable biotelemetric devices. Indeed, the reported CP antennas can sustain reliable communication links and reduce polarization mismatch losses. However, they have been restricted to the single-input-single-output (SISO) port configuration and further limited with the limited bandwidth of the 3-dB axial ratio (AR). Furthermore, all the reported LP and CP wearable antennas are not designed to operate in the mm-wave frequency band. Most of them have not considered the antenna conformability and coupling reduction between the antenna elements and safety issues, which are essential for wearable wireless technologies. In [6], the loop based microstrip antenna is proposed at the frequency of 60 GHz. A flexible antenna designed using a substrate of liquid crystal polymer working in the mm-wave (24-40 GHz) was reported in [25]. In [27], a microstrip line-fed circular patch operating in the mm-wave range (69.6 GHz) was proposed. In [28], a microstrip patch antenna printed on fingernail was reported in the mm-wave range (28 GHz), but the authors did not perform SAR analysis. All the reported antennas are designed for biomedical applications in the millimeter range, are linearly polarized, and are restricted to the SISO port only. Moreover, most of the antennas have not done bending and SAR analysis. While designing the wearable antenna, it is critical to ensure low levels of SAR value by minimizing the unwanted back radiations from the antenna [33]. In [26], [34] adopted the extra layer of electromagnetic bandgap structures at the backside of the antenna to lower the back radiations, which are usually transformed into heat by the lossy tissues. Additionally, to the best of the author's knowledge, only the demonstrated work in [26] has reported the semi-flexible EBG-backed mm-wave (24 GHz) two-port MIMO antenna for wearable applications. Certainly, the reported MIMO antenna leads to reduced heat absorption; nonetheless, the antenna could not achieve vertical compactness due to its multiplayer profile; and was also restricted to the single operating frequency band with linear polarization and narrow impedance bandwidth.

In this paper, for the first time, a novel semi-flexible broadband wearable two-port MIMO antenna with vertical compactness due to a single layer profile is reported in the millimeter wave frequency spectrum (24.14–30.48 GHz). The proposed mm-wave MIMO design is based on the loading of several sectoral slots at the periphery of the circular radiating patch. Each sectoral slot in the patch excites the corresponding resonance mode. The frequency of each resonance mode can be controlled by slot dimensions and its position in the patch. However, in the proposed design, all the sectoral patch fins have the same dimensions, making the resonance overlap leading to the wideband performance. Although the proposed MIMO antenna is a bit larger in size; but, still proves its novelty with a strong candidature in terms of diversified CP property, broad impedance bandwidth, easy-to-use geometry with comparable on-body radiation performance (gain and efficiency) and low SAR relative to state-of-the-art mm-wave antennas, as shown in Table I. Moreover, the lean availability of research articles in the literature similar to the proposed work on mm-wave bands further justifies the importance of the proposed wearable MIMO antenna. The *mm*-wave bands are also license-free, compatible with the growing 5G technologies, and predominantly highly resistant to atmospheric attenuation [6]. Moreover, in the high-frequency band, the wearable antenna can easily satisfy the foremost requirement, i.e., energy efficiency, without compromising the requirements such as broad bandwidth, robust wireless link, high gain, channel capacity, and data rates to fulfill the body-centric applications beyond 6 GHz.

This article consists of seven sections, after the introduction, organized as follows: Section II presents the mm-wave MIMO antenna design in homogeneous space, followed by its operating mechanism and CP property. Section III analyses the bending effects of the antenna, and section IV reports the antenna performance at different locations on the anatomical body model. This section further compares the near-field and far-field radiation parameters in simulation and measurement scenarios. Section V discribes the numerical calculation of MIMO diversity parameters, and section VI shows the safety validation of the proposed MIMO antenna through SAR evaluation and further checks the communication link capacity. Section VII winds up the article with a conclusion.

II. Off-Body Antenna Performance

A. *mm-wave MIMO antenna design:*

A wideband *mm*-wave 2-element MIMO antenna consisting of two circular-shaped radiating patches is designed on one side of the substrate, as shown in Fig. 1. Each antenna element is embedded with eleven sectoral slots and fed by a 50 Ω transmission line, and the width is calculated using the basic antenna design equations. The proposed antenna is printed on a semi-flexible material (RT/duroid 5880) having a relative permittivity (ϵ_r) 2.2, loss tangent (tan δ) 0.0004, and thickness (h) 0.508 mm. Fig. 1(a) presents the front view of the antenna design, in which a slotted circular patch of radius r_2 is fed with a microstrip line. All the metallic sectoral patch fins have the slant length l_1 ($AB = AC$) with an arc angle θ_1; except the fin (arc angle θ_3) connected with the feedline. The sectoral slots have dimensions l_2 ($DE = DF$) with an arc angle θ_2. A defected ground plane (Fig. 1(b)) embedded with two moon-shaped slots (with radius r) is printed on the bottom side of the substrate. The overall footprints of the proposed MIMO antenna design are 22.5×36.0×0.508 mm^3, and the optimized dimensions are given in Table II.

B. *MIMO antenna operating mechanism:*

The design iteration of the proposed MIMO antenna structure is shown in Fig. 2. The corresponding simulated S-parameters, axial-ratio (AR), and total gain in the far-field are presented in Fig. 3 (a and b). Initially, simulation in electromagnetic (EM) software CST Microwave Studio is iterated with design-1, consisting of two circular-shaped radiating patches connected with the respective feedlines, and the ground plane is intact. The single element of design-1 provides the two resonant peaks at 25.7 and 29 GHz, as validated from the impedance plot in Fig. 4. The resonant frequency (f_r) is calculated using the following equations (1-2) [29] by treating the radiating patch as a circular waveguide [35], [36].

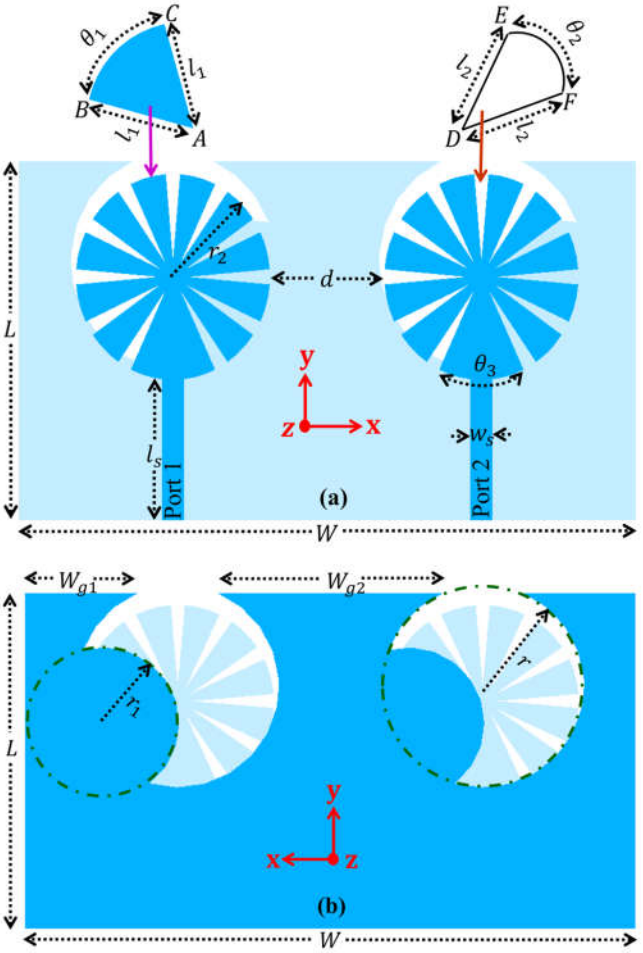


Fig. 1 Proposed MIMO antenna geometry (a) Front view (b) Bottom view.

Table II
Design Parameters Of The Proposed Antenna (in mm)

Symbol	Value	Symbol	Value	Symbol	Value
L	22.5	l_2	5.45	W_{g1}	5.07
W	36.0	d	7.63	W_{g2}	15.14
l_s	8.84	r	6.8	θ_1	21.17°
w_s	1.48	r_1	5.1	θ_2	8.83°
l_1	5.45	r_2	6.45	θ_3	38.83°

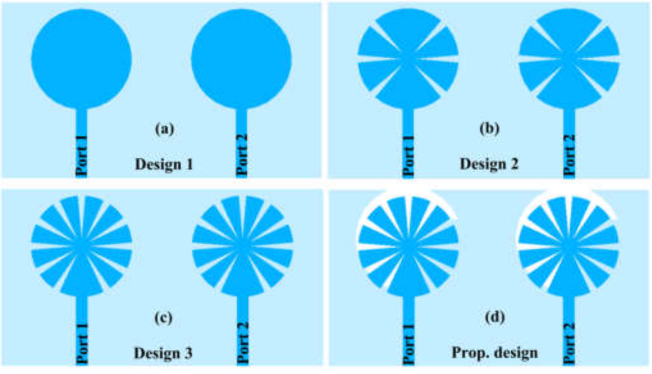
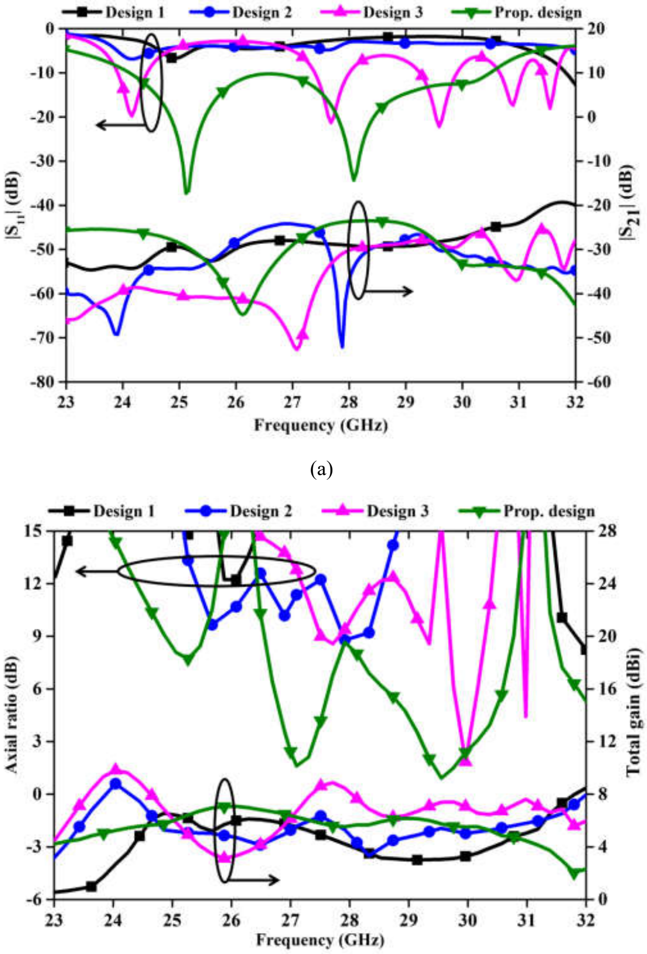


Fig. 2 Design iteration steps of the proposed *mm*-wave MIMO antenna.



(a)

(b)

Fig. 3 Simulated results of the proposed *mm*-wave MIMO antenna (a) *S*-parameters (b) A.R and gain.

$$f_r = \frac{k_{nm} v_0}{2\pi R_e \sqrt{\varepsilon_r}} \qquad (1)$$

$$R_e = r_2\left\{1+\frac{2h}{\pi\varepsilon_r r_2}\left[ln\left(\frac{\pi r_2}{2h}\right)+1.7726\right]\right\}^{1/2} \qquad (2)$$

Here, v_0 = velocity of light in vacuum, R_e = effective radius of the circular patch. Design-1 operates in the dual mode TM$_{21}$ (k_{nm}= 5.135) and TM$_{31}$ (k_{nm}= 6.110); where '*n*' refers to the number of circumferential variations and '*m*' refers to the radial variations. As shown in the current distribution (Fig. 5), at 25.7 GHz and 29 GHz, direction of the current varies twice and thrice in the radial direction, respectively; while travelling the circular boundary of the patch in the circumferential direction. However, a bit of asymmetry in the lower half of the surface is due to the feed microstrip line. However, at both the modes design-1 with r_2 = 6.45mm ($\approx \lambda_o/2$) is found to be non-resonating structure as |S$_{11}$| is > -10dB in the targetted *mm*-wave frequency band (24-31 GHz).

In order to obtain the impedance matching in the band of interest, a sectoral slot is embedded in the patch of the single element, and when rotated at the particular location on the patch makes the structure to resonate at different frequencies, as shown in Fig. 6.

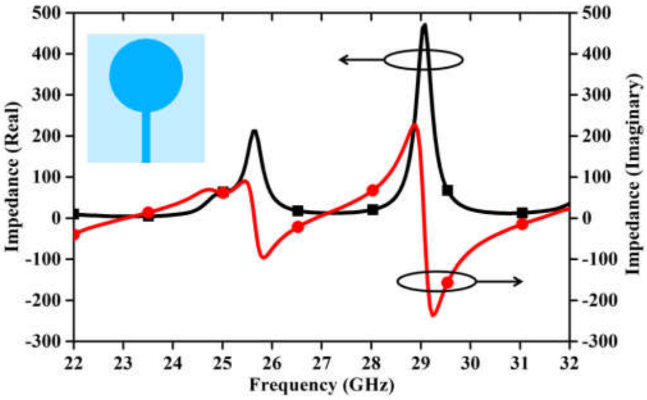


Fig. 4 Input impedance plot for the single element of design-1.

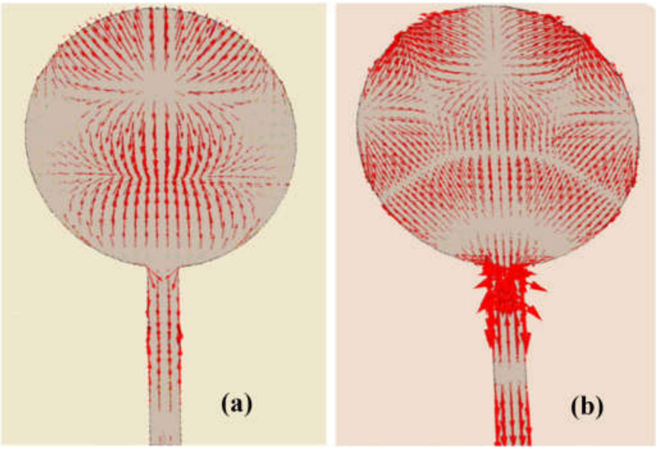
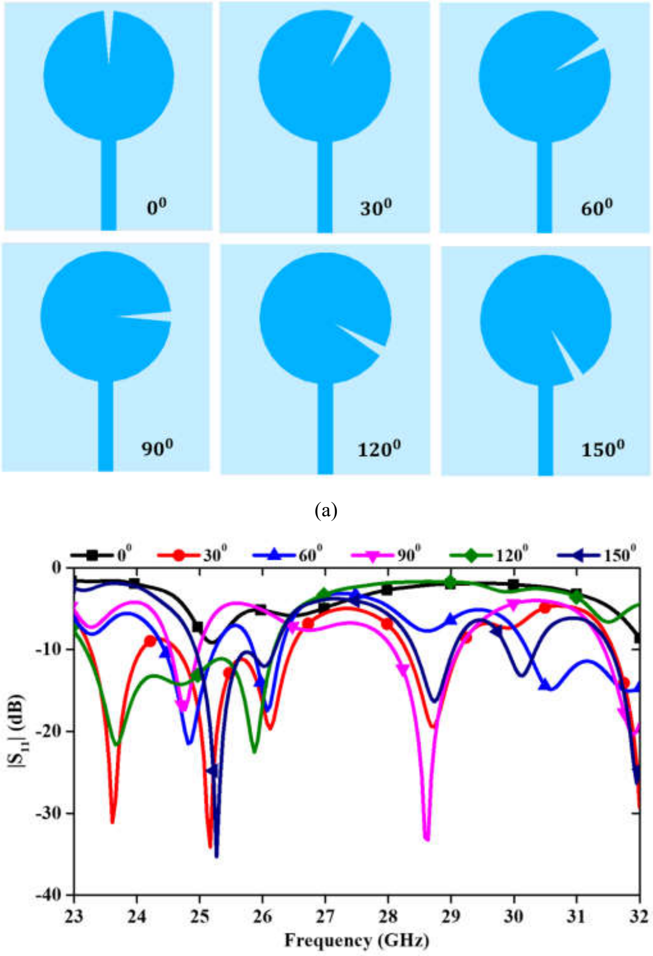


Fig. 5 Current distribution (0 deg. phase) on the surface of the single element of design-1 (a) 25.7 GHz (b) 29.0 GHz.

Therefore, six sectoral slots with the optimized dimensions (l_2 and θ_2) are incorporated in the two circular patches of symmetrical design-2. The design-2 iteration provides the two peaks at 24.12 GHz and 27.5 GHz, but still, the antenna structure stays non-resonant (see Fig. 3a). Subsequently, on increasing the optimized sectoral slots upto eleven in each element without losing the symmetricity, sectoral patch fins in the structure of design-3 resonated at five modes (24.16 GHz, 27.68 GHz, 29.59 GHz, 30.89 GHz, and 31.55 GHz) with fairly improved matching. Also, the -10 dB bandwidth is more than 500 MHz in most of the frequency bands. Fig. 7 shows the current distribution on the patch surface of design-3 at the first two resonant modes for brevity. At 24.16 GHz, on fins-1 to 6, the current with high intensity (red colour vectors) is going outside towards the periphery, whereas on fins-7 to 11, the current with high intensity going inside towards the centre. Note, fin-9 can be treated as a combination of two fins. From the surface current distribution, we can observe that in design-3 the combination of different patch fins makes a wave of different modes that are encircling the patch periphery. Therefore, different combinations of patch fins are providing different resonances (see Fig. 3a). Moreover, most of the radiation takes place only from the edges of the slot, and the current on the periphery of the patch fins is not contributing to the radiation significantly (because opposite current vectors cancel each other). In contrast, the current on the edges of the particular slot is in the same direction. Finally, to achieve the wide mm-wave frequency band of interest (24-31 GHz), all the resonating modes are merged by adding a moon-shaped slot (each with a diameter of length λ_o) beneath the patch in the ground of the proposed design-4.



(a)

(b)

Fig. 6 Simulated $|S_{11}|$ results of the single element of design-1 for different angular position of a sectoral slot on the circular patch.

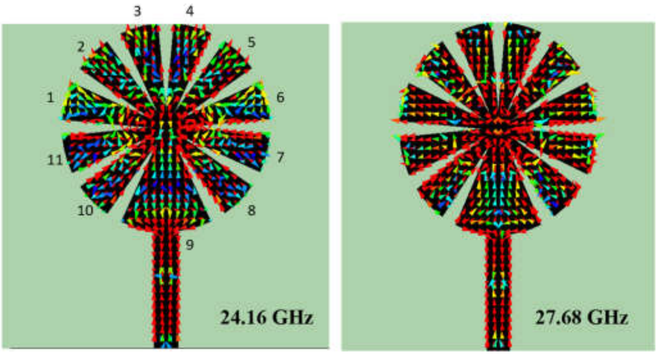


Fig. 7 Current distribution (0 deg. phase) on the patch surface of design-3 at the first two resonant modes only for brevity.

This design iteration provides the simulated impedance bandwidth from 24.14 to 30.48 GHz, as the moon shaped slots sustained the purely resistive impedance in the whole frequency range. Design-4 is the proposed MIMO antenna. Further, the $|S_{11}|$ value also improves to -37.29 dB and -34.33 dB at the two resonant frequencies, 25.11 GHz and 28.08 GHz, respectively. Additionally, the EM radiations from design-3 are circularly polarized (CP) at only 29.96 GHz with 3-dB bandwidth of 80 MHz, whereas the proposed design-4 is circularly polarized at 27.10 GHz and 29.53 GHz with wide impedance bandwidth due to the two moon-shaped slots in the ground plane. The simulated 3-dB AR bandwidth is 350 MHz and 940 MHz, respectively, with a total gain > 4dBi in the whole impedance bandwidth and a peak gain of 6.1 dBi at 26.08 GHz, shown in Fig. 3(b). Noted, when the sectoral slots are less than seven, design-1 and -2 do not exhibit CP characteristics. Hence, it verifies that CP characteristics are achieved due to the combination of moon-shaped slots in the ground plane and optimized eleven sectoral slots in the circular patch. Furthermore, the two antenna elements are well isolated from each other as simulated $|S_{12}|$ is < -25 dB in the entire frequency band. To have low mutual coupling, both the antenna elements are kept isolated from each other with the optimized edge-to-edge separation distance (d) of $\approx \lambda_o/2$ at the lower resonant frequency (f_r = 25.11 GHz) in the wideband range.

C. *Circular polarization:*

The proposed MIMO antenna (design-4) radiates CP waves at 27.10 GHz and 29.53 GHz. At both frequencies, the vertically polarized electric field E_θ along the slant length of the fins and horizontally polarized field E_ϕ in the radial direction of the fins can be elaborated by the surface currents on the sectoral patch and the slotted ground plane, as shown in Fig. 8(a-b). For brevity, the current distribution density is shown for port-1 only. For CP radiation, amplitudes of E_θ and E_ϕ must be equal, and their phase difference should be 90°. The 90° phase difference between the two radiations is obtained by having the slant length of each fin around half-wavelength $\lambda_0/2$ at the two frequencies of the proposed antenna. At 27.10 GHz, the combined current vectors in the radial directions of all the fins leading to the E_ϕ fields circulated in $+\phi$ direction (anticlockwise, i.e., RHCP) as time-phase (ωt) changes quarterly from 0° to 90°. In addition, the current on the periphery of the moon-shaped slot in the ground plane travels in a clockwise (LHCP) direction in a phase quadrature that also supports RHCP radiation in the broadside. At 29.53 GHz, the E_ϕ fields circulated in $-\phi$ direction (clockwise, i.e., LHCP) as ωt changes quarterly. Moreover, the current on the periphery of the ground plane's slot travels in an anticlockwise (RHCP) direction in a phase quadrature that supports LHCP radiation in the broadside. As surface currents due to each element of the proposed *mm*-wave MIMO antenna structure provides E_ϕ fields in two different directions ($+\phi$ and $-\phi$), while E_θ fields in the same direction (fin's slant length) at both the resonant modes; therefore, in broadside, the MIMO antenna radiates the RHCP waves at 27.10 GHz, and LHCP waves at 29.53 GHz. This working mechanism is also consistent with Alford's loop antenna [37].

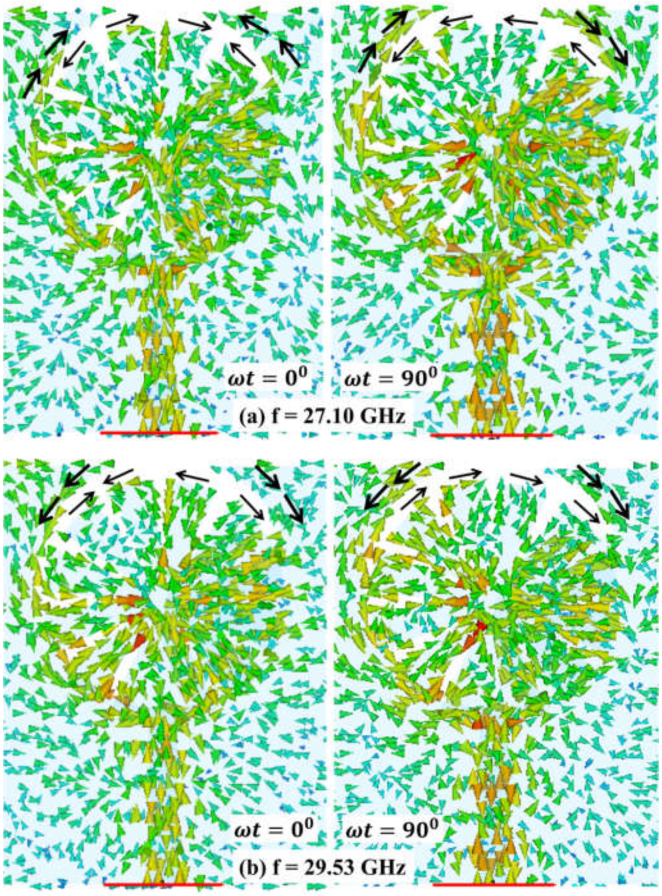


Fig. 8 Simulated surface current distribution on the patch (thin arrows) and ground plane (thick arrows) at $\omega t = 0°$ (left) and 90° (right) for port-1.

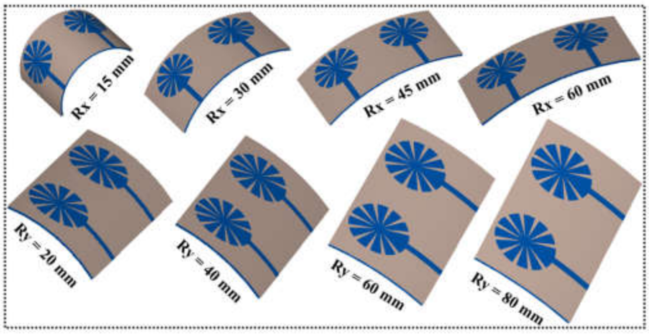


Fig. 9 The proposed MIMO antenna in bent form with radius Rx and Ry in x- and y-direction, respectively.

III. Conformability of Mimo Antenna

For wearable applications, this analysis is significant to carry out as the structure of the MIMO antenna, when placed on the body parts (chest, hand, and leg), can bend with the body curvature and detune the performance characteristics [18], [38], [39]. For such analysis, performance is characterized when the MIMO antenna structure is bent along x-and y-directions with radius R_x and R_y, respectively, as shown in Fig. 9. The radius values in both directions are chosen as per the typical body curvature [26].

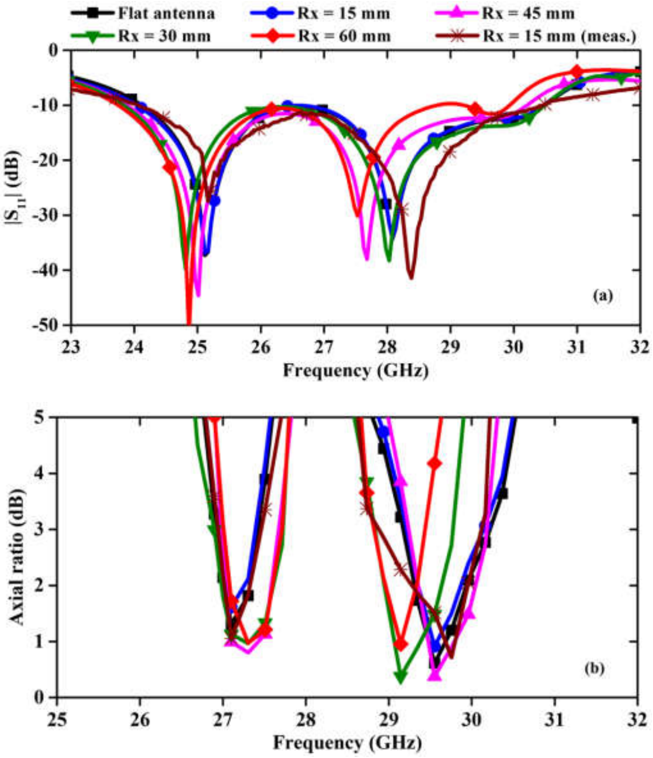


Fig. 10 Comparison of simulated and measured (worst case) results when the MIMO antenna is flat and in bent form with radius R_X (a) $|S_{11}|$ and (b) A.R.

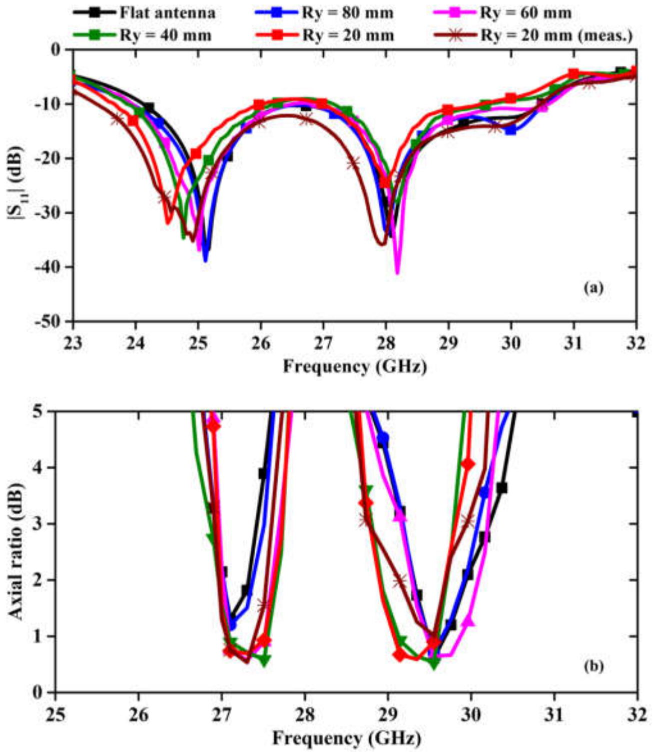


Fig. 11 Comparison of simulated and measured (worst case) results when the MIMO antenna is flat and in bent form with radius R_Y (a) $|S_{11}|$ and (b) A.R.

From Fig. 10(a-b), it is evident that the overall simulated -

10dB $|S_{11}|$ bandwidth, including the two resonant peaks and 3-dB AR bandwidth at both the resonant frequencies, shifts slightly to the lower frequency without any effective detuning in the performance when the bending gets more prominent in the x-direction. Moreover, a similar shift to the lower frequency with inconsequential detuning in the overall results is also observed when the bending gets notable in the *y*-direction, as shown in Fig. 11(a-b). The shift to the lower frequency relative to the flat version is expected as the current path on the radiating patch gets elongated due to bending in both directions. The measured results (measurement scenario for the bent antenna is shown further in Fig. 16(b)) for the worst bending scenario (R_x = 15mm, R_y = 20mm) are exceptionally following the corresponding simulated curves with noticeable variation in bandwidths and impedance matching. The simulated and measured gain and radiation efficiency of the proposed MIMO antenna is also analyzed for the worst bending scenario as shown in Fig. 12. The efficiency is more than 80% in the entire impedance bandwidth and gain improves to more than 5 dBi as unwanted coupling may have reduced between the elements due to bending. Notably, the measured performance for the worst bending cases follow the simulated performance convincingly, therefore, the proposed MIMO antenna is interestingly immune to the possible bending issues.

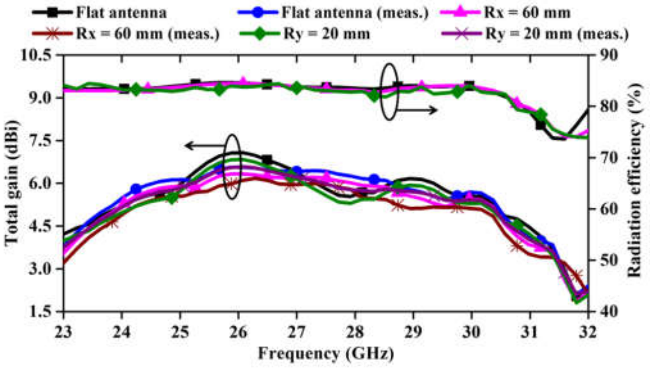


Fig. 12 Comparison of simulated and measured (worst case) gain and radiation efficiency when the MIMO antenna is flat and in bent form with radius Rx and Ry.

IV. On-Body Antenna Performance

In this section, the performance of the proposed MIMO antenna in terms of AR, gain, and S-parameters is analyzed and reported in the presence of a realistic anatomical human body model. The antenna structure is placed on different body parts (chest, hand, and leg) of the 3-D Gustav voxel model, as shown in Fig. 13. Noteworthy, the gap between the skin surface and the MIMO antenna is taken as 4 mm to consider the real-time framework where the whole antenna system and circuitry module are placed over the body-worn clothes. Generally, in normal cases, the thickness of the clothes can be 1-6 mm, and beyond 6 mm, the antenna performance can be expected to resemble the off-body case as sufficient free space cover the area beneath the ground plane. Noted, the frequency domain simulations are carried out in the EM simulator CST using a workstation (32 cores and 1-TB RAM). Each full wave simulation takes about 90 minutes to execute.

Fig. 14(a) compares simulated S-parameters for both the cases: on-body and off-body. It is observed that the impedance matching of the proposed MIMO antenna, particularly at the two resonant frequencies, gets worst hit by the body tissues. The cases of leg and hand are influenced the most as the curvature of these body parts may have restricted the current flow on the antenna, but for the chest, matching is flawless as the chest resembles a somewhat flat case. The higher frequency band shifts to the right by more than 1 GHz compared to the flat case, but the entire impedance bandwidth remains intact and acceptable. Further, the transmission coefficient ($|S_{12}|$) stays well below -30 dB in the entire bandwidth for the on-body case. From Fig. 14(b), it is also noticed that the AR bandwidth and gain also exposes to some insignificant variations compared to flat case, but the far-field attributes stay almost stable for different locations of the proposed antenna on the human body.

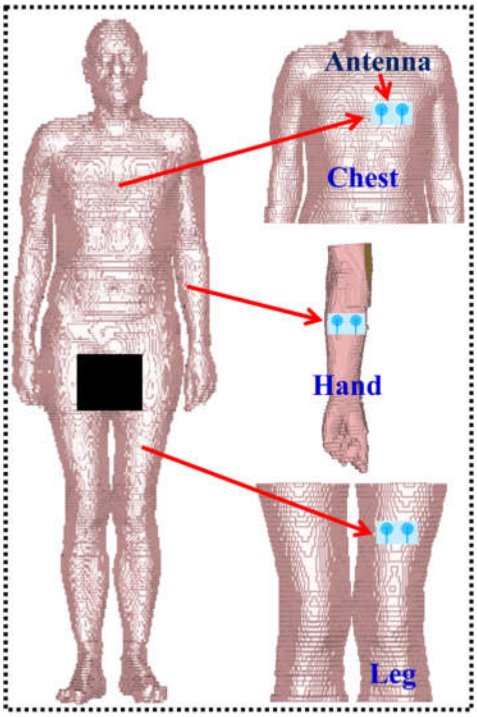


Fig. 13 The different locations of the proposed MIMO antenna on the surface of the anatomical voxel based Gustav human body model.

Fig. 15 shows the fabricated prototype of the proposed MIMO antenna structure where the feed line is connected to a connector. Also, the adopted an off-body scenario to measure S-parameters in flat and bent form is shown in Fig. 16 (a and b), respectively. For easy bending, the bent antenna is backed by the 20 mm thick foam ($\epsilon_r \approx 1$) to mimic free space beneath the ground plane and it is placed on the curved adhesive tape. Fig. 17 shows the on-body approach we adopted to measure the *S*-parameters of the MIMO antenna at 4 mm height (including cloth layer) over the body parts of the living human body. Noteworthy, for on-body measurement of AR, gain, and radiation patterns, the fabricated MIMO antenna is placed over the chest mimicking phantom in a plastic container of size 120 × 80 × 40 mm^3 inside an anechoic chamber. As the dielectric properties of the human body are anisotropic in the environment

[40], the gel-based phantom is developed using the recipe provided in [41], [42] to mimic the real chest tissues at 27.10 GHz and 29.53 GHz. Fig. 18(a) shows the comparison of measured S-parameters for both the cases: on-body and off-body. As expected after observing the simulated $|S_{11}|$, the measured $|S_{11}|$ shifts to the higher frequency, whereas impedance matching is improved without any loss in the entire bandwidth for both on-body and off-body scenario.

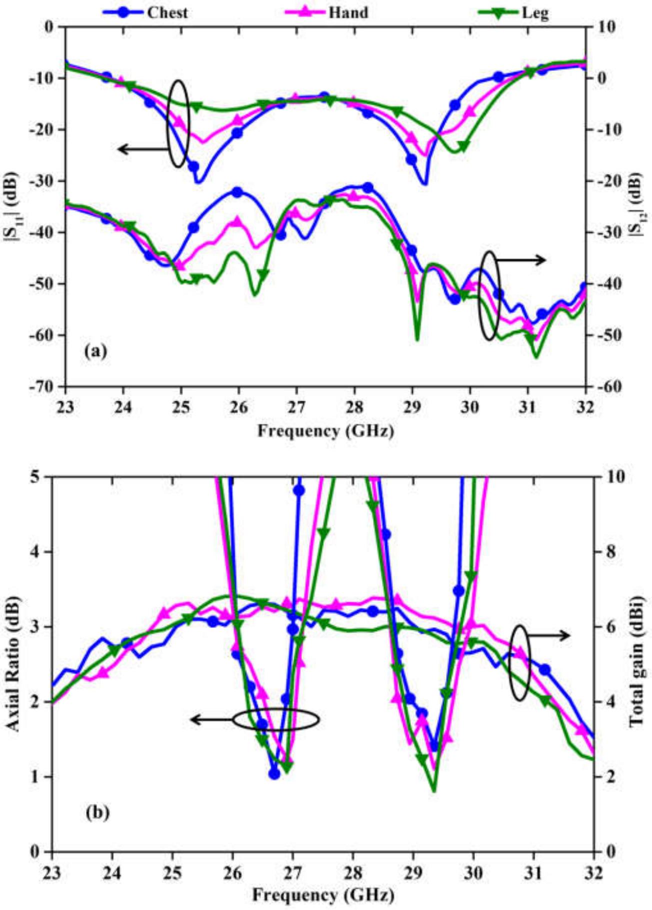


Fig. 14 Simulated results when the proposed MIMO antenna is placed on chest, hand, and leg of the realistic body model (a) S-parameters (b) A.R and gain.

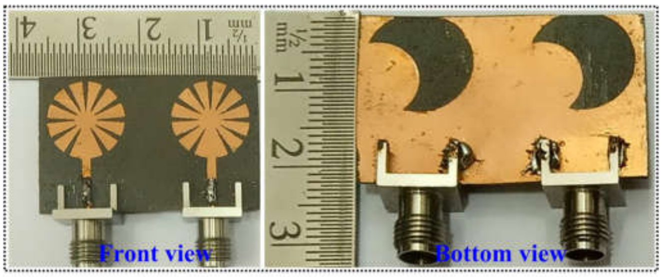


Fig. 15 Fabricated prototype of the proposed *mm*-wave MIMO antenna.

For the case of leg and hand, the shift is more with respect to the chest and flat case. The measured AR and gain in Fig. 18(b) are not harshly affected and follow the simulated results satisfactorily with improved 3-dB bandwidth for both on-body and off-body cases [43], [44]. The measured $|S_{12}| < -20$ dB in the entire measured impedance bandwidth of the proposed antenna varies from 23.96 to 31.41 GHz, and the measured AR bandwidth varies from 25.84 to 27.35 GHz and 28.57 to 29.85 GHz. It is worth to be mentioned that the measured performance of the proposed MIMO antenna is well tuned with the simulation, and the results are acceptable as evident from Fig. 18. In Table III, the simulated and measured impedance bandwidth and AR bandwidth are compared for different positions of the MIMO antenna located on the human body. From this comparison, it is inferred that the performance of the proposed MIMO antenna is strongly immune for different positions on the human body.

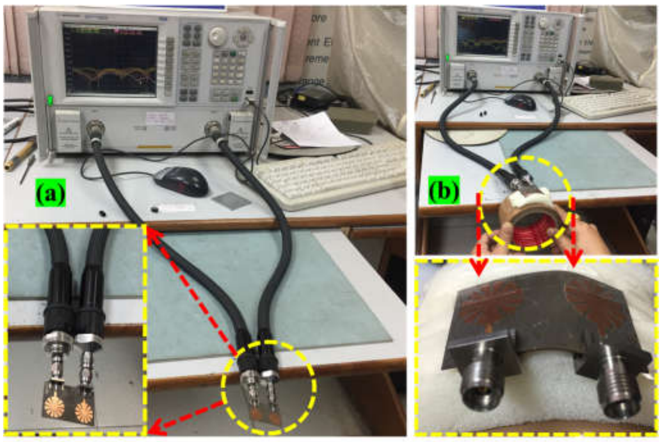


Fig. 16 Off body scenario to measure *S*-parameters of the fabricated MIMO antenna prototype (a) antenna in flat form (see inset) (b) antenna in bent form (see inset).

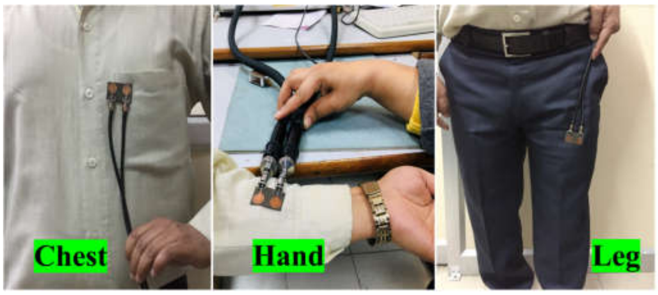
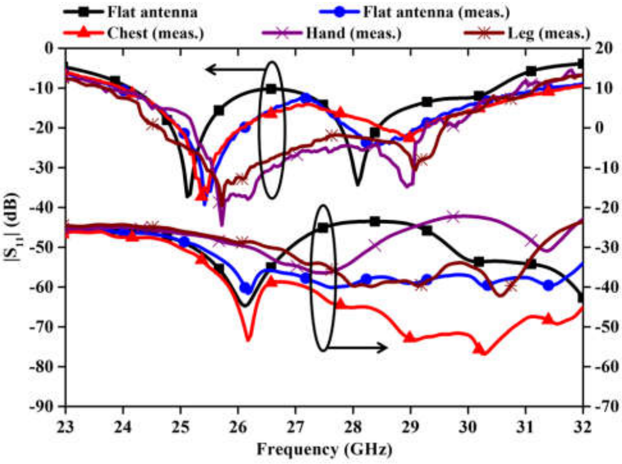
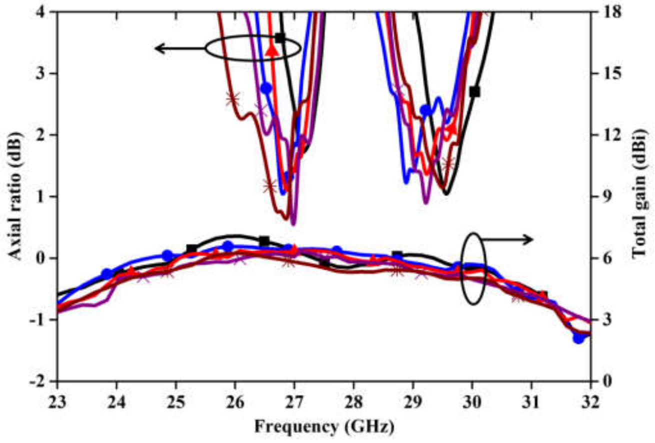


Fig. 17 On body scenario to measure *S*-parameters of the fabricated MIMO antenna prototype.



(a)



(b)

Fig. 18 Comparison of simulated and measured results for off body and on-body scenario (a) *S*-parameters and (b) AR and gain.

The radiation pattern performance is studied for the proposed MIMO antenna for off-body (flat case) and on-body (chest) scenario as demonstrated in Fig. 19(a-b), and the corresponding measurement setup is shown in Fig. 17. Note, for the on-body performance, the RHCP and LHCP radiation patterns are simulated over the chest of the anatomical voxel human model but measured over the custom-made chest phantom.

TABLE III
COMPARISON OF IMPEDANCE AND AR BANDWIDTH

MIMO Position	Sim. $\|S_{11}\|$ B.W (GHz)	Meas. $\|S_{11}\|$ B.W (GHz)	Sim. A.R B.W (GHz)	Meas. A.R B.W (GHz)
Flat (Air)	24.13−30.48	23.96−31.41	26.91−27.42 29.18−30.16	26.37−27.15 28.57−29.78
Chest	23.75−30.39	23.99−31.66	26.04−27.05 28.67−29.62	26.63−27.27 28.70−29.81
Hand	23.72−30.72	23.64−30.87	26.03−27.17 28.62−29.96	26.38−27.35 28.71−29.80
Leg	23.67−30.82	23.69−31.15	26.08−27.12 28.66−29.84	25.84−27.28 28.77−29.85

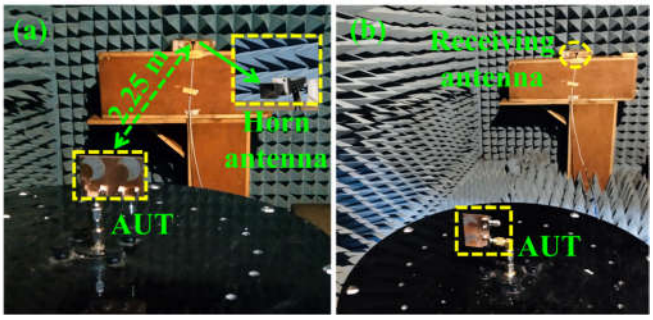


Fig. 19 Radiation pattern measurement setup in anechoic chamber (a-b) off-body.

Although the radiations patterns are measured in free space also for the flat case, for brevity, only the measured radiation patterns (received power distribution) for the on-body case are presented in Fig. 20; which shows 2D normalized LHCP and RHCP radiation patterns at 27.10 and 29.53 GHz in the principal elevation planes E-plane (ϕ = 0^0) and H-plane (ϕ = 90^0). The pattern plots show that at 27.10 GHz, measured RHCP surpassed LHCP by more than 20 dB and 9.34 dB in the zenith position ($\theta = 0^\circ$) and in the azimuthal position ($\theta = 90^\circ$), respectively. Whereas, at 29.53 GHz, LHCP surpassed its counterpart radiation by more than 25 dB and 12 dB in the two positions, respectively. Overall, the patterns resemble the omnidirectional property at both frequencies by having good radiation coverage at all the upper and lower hemisphere angles in both the principal planes. The stable measured far-field performance in the vicinity of the chest phantom demonstrates the proposed MIMO antenna as an excellent candidate for biomedical applications in the *mm*-wave range.

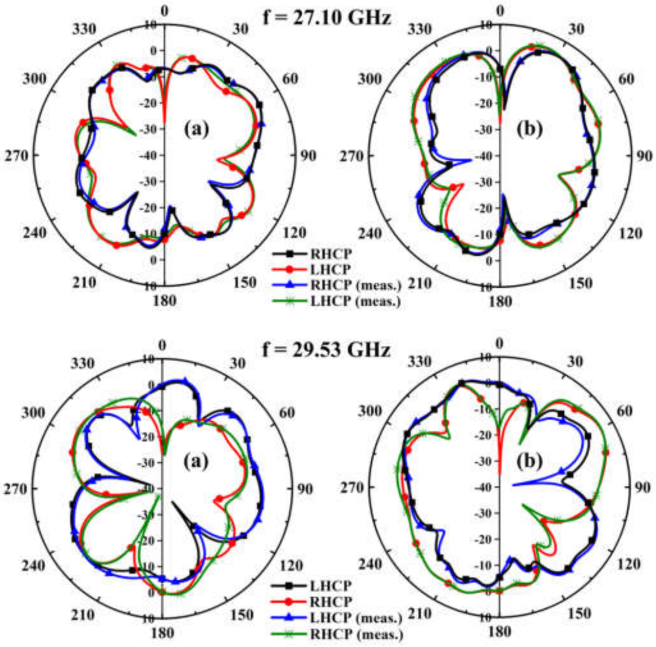


Fig. 20 Comparison of simulated and measured RHCP and LHCP patterns in on body scenario (chest) (a) ϕ = 0^0 plane (b) ϕ = 90^0 plane

V. DIVERSITY PERFORMANCE OF THE MIMO ANTENNA

This section presents the study of diversity parameters of the proposed mm-wave MIMO antenna when placed on the human chest to ensure the on-body performance quality in the MIMO configuration. The results are derived only for the on-body case (chest) for brevity. The simulated diversity parameters are further validated through comparison with the measured results.

A. Envelop Correlation Coefficient (ECC) and Diversity Gain (DG)

The ECC parameter signifies the difference in the radiation performance of an individual antenna element from the nearby elements. Ideally, the ECC should be zero for uncorrelated MIMO antenna elements, and practically, ECC < 0.5 is acceptable. The expression for ECC is given by (5) [45], [46]. In this 2-port MIMO system, ECC is evaluated using far-field radiation patterns. Here, the cross-polarization rate (XPR) is defined as [46]:

$$XPR = \frac{P_v}{P_h} = 1$$

$$\zeta_{ecc(kl)} = \frac{\left|\int_0^{2\pi}\int_0^{\pi}\{XPR\cdot E_{\theta k}E_{\theta l}^{*}P_\theta + E_{\phi k}E_{\phi l}^{*}P_\phi\}d\Omega\right|^2}{\int_0^{2\pi}\int_0^{\pi}\{XPR\cdot E_{\theta k}E_{\theta k}^{*}P_\theta + E_{\phi k}E_{\phi k}^{*}P_\phi\}d\Omega \times \int_0^{2\pi}\int_0^{\pi}\{XPR\cdot E_{\theta l}E_{\theta l}^{*}P_\theta + E_{\phi l}E_{\phi l}^{*}P_\phi\}d\Omega} \tag{5}$$

Further, the impact on the transmitted power due to diversity performance can be observed from diversity gain (χ_d) of the MIMO antenna, which can be expressed as (6):

$$\chi_d = 10\sqrt{1-|\zeta_{ecc}|^2} \tag{6}$$

In Fig. 21, the simulated and measured values of ζ_{ecc} are consistently < 0.003 in the entire operating mm-wave band (24-31 GHz). Since, the ECC depends upon the x-polarization level, and x-polarization value is low for the proposed MIMO antenna design, therefore, the ECC values are low. The lesser value of ζ_{ecc} indicates the good decoupling characteristics of the proposed MIMO antenna for the on-body scenario. Moreover, both simulated and measured diversity gain (χ_d) of the designed antenna stays above 9.9 dB across the entire impedance bandwidth.

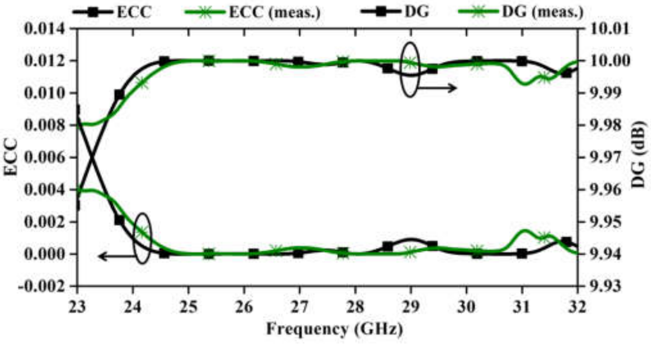


Fig. 21 Simulated and measured ECC and diversity gain of the proposed MIMO antenna.

B. Mean Effective Gain (MEG)

The gain performance of the proposed MIMO antenna can be analyzed by the MEG parameter (ξ), and its expression is given by (7) [34]:

$$\xi = \frac{P_r}{P_{in}} = \int_0^{2\pi}\int_0^{\pi}\left[\frac{XPR}{1+XPR}\times F_\theta(\theta,\phi)P_\theta(\theta,\phi) + \frac{1}{1+XPR}\times F_\phi(\theta,\phi)P_\phi(\theta,\phi)\right]sin\theta d\theta d\phi \tag{7}$$

Here, F_θ and F_ϕ represent the power gain patterns of the designed antenna. Noted, the difference between the MEG of the two ports should be < -3 dB in the case of a two-port MIMO configuration. The comparative results of MEG for port-1 and -2 are demonstrated in Fig. 22, where the MEG values for both the ports are not exceeding -6 dB across the operating frequency band.

C. Total Active Reflection Coefficient (TARC)

The TARC values signify the variation in reflection coefficient with a change in the phase angle of the signal feeding to the port. For the proposed two-port MIMO antenna elements, the TARC (χ) is given by (8) [47]:

$$\chi = \sqrt{\frac{|(S_{11}+S_{12}e^{j\theta})|^2 + |S_{21}+S_{22}e^{j\theta}|^2}{2}} \tag{8}$$

here, θ = signal phase angle, S_{11}/S_{22} = reflection coefficients of port 1/port 2, and S_{12}/S_{21} = isolation between port-1 and -2. The calculated value of χ for the designed antenna is shown in Fig. 23. The phase of the input signal is varied from 0 to 180^0 with regular intervals of 45^0. The value of χ for different phase angles of the exciting signal is stable across the entire frequency band, which indicates the low mutual coupling between the two antenna elements.

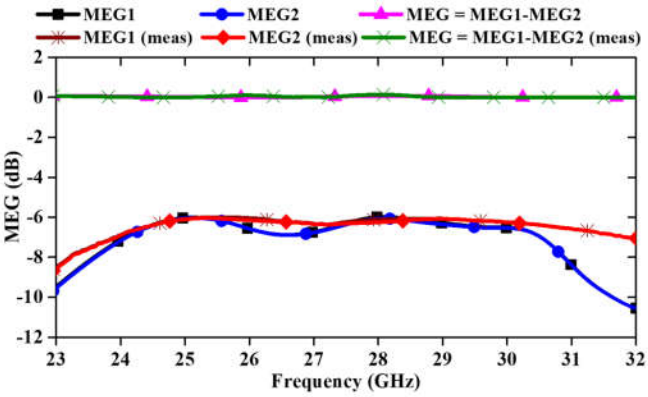


Fig. 22 Simulated and measured MEG for port-1 and -2 of the proposed MIMO antenna.

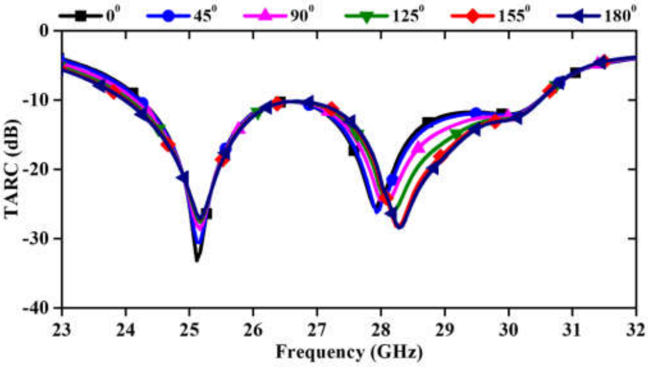


Fig. 23 Calculated TARC at different angles of the proposed MIMO antenna.

D. Channel Capacity Loss (CCL)

The CCL $(C_{\mathbb{C}})$ for MIMO antenna is defined as the flow of EM signal with maximum achievable data rate and with least distortion. The value of $C_{\mathbb{C}}$ can be calculated as (9) [48]:

$$C_{\mathbb{C}} = -\log_2 det(\delta) \tag{9}$$

and,

$$\delta = \begin{bmatrix} \cap_{11} & \cap_{12} \\ \cap_{21} & \cap_{22} \end{bmatrix}$$

where,

$$\cap_{11} = 1 - [|S_{11}|^2 + |S_{12}|^2]$$
$$\cap_{12} = -[S_{11}^* S_{12} + S_{21}^* S_{12}]$$
$$\cap_{21} = -[S_{22}^* S_{21} + S_{12}^* S_{21}]$$
$$\cap_{22} = 1 - [|S_{22}|^2 + |S_{21}|^2]$$

The comparison of $C_{\mathbb{C}}$ between its simulated and measured values are depicted in Fig. 24. The measured value of $C_{\mathbb{C}}$ is < 0.26 bits/s/Hz which is quite acceptable over the desired millimeter frequency band.

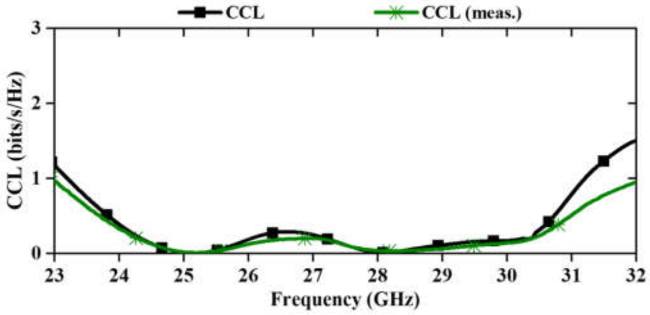


Fig. 24 Simulated and measured CCL of the proposed MIMO antenna.

E. *Multiplexing Efficiency (ME)*

Another important parameter of the MIMO antenna is multiplexing efficiency (η_{muxeff}) and it is given by (10) [49]:

$$\eta_{muxeff} = \sqrt{\eta_1 \eta_2 [1 - |\delta|^2]} \qquad (10)$$

here, η_1 , η_2 are the radiation efficiencies of two radiating structures, and δ = complex correlation coefficient between these two radiating elements ($i.e. \delta = |\zeta_{ecc}|^2$). The maximum multiplexing efficiency of the proposed MIMO antenna on the chest is found to be -0.43 dB throughout the operating frequency band.

VI. Communication Link

A. *Specific absorption rate:*

The absorption of EM radiations emitted from the antenna operating in close contact with the human body ultimately raises the temperature of the surrounding tissues. The excess tissue heating beyond the permissible limit can be life-threatening. At present, the exposure to the EM radiations is estimated by the dosimetric quantity, specific absorption rate (SAR). In our work, the SAR is evaluated numerically when the proposed MIMO antenna is placed 4 mm above the chest of the Gustav voxel model in the simulator. The evaluation is done considering the strictest of all, IEEE C95.1-1999 guidelines, which restrict the average SAR to 1.6 W/Kg for 1g of cubic tissue [50]. As the typical maximum value for wearable devices is 50 mW (17 dBm), therefore, when the single antenna element is given 50 mW of input power, the surrounding tissues in the neighbourhood of the chest provide the 1-g peak SAR value of 2.18 W/Kg and 1.64 W/Kg at 27.10 GHz and 29.53 GHz, respectively. It is noticeable from Fig. 25 that the maximum absorptions are just below the slot in the ground, and the remaining intact ground plane is acting as a shield and preventing the exposure to harmful radiations. Conforming to the IEEE guidelines, the SAR value of the single element of the proposed MIMO configuration can lead to 1.6 W/Kg, only when excited with maximum input power of 36.70 mW (15.65 dBm) and 48.78 mW (16.88 dBm), respectively. As the transmitter power values of the wearable devices are in the order of 0.1 mW (-10 dBm) [22], therefore, the operation of the MIMO antenna can be satisfactorily adopted for practical use in the mm-wave frequency band.

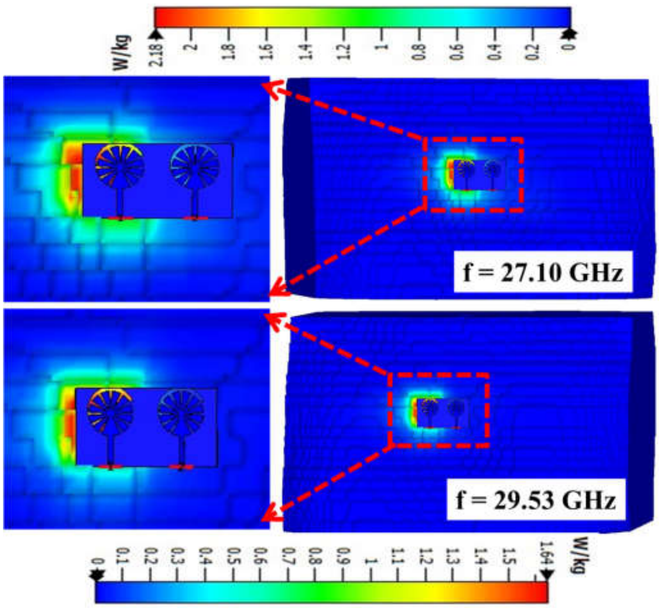


Fig. 25 Peak SAR distribution for the proposed MIMO antenna on chest of the realistic Gustav human body model.

B. *Link margin:*

To investigate the communication capacity of the proposed wearable mm-wave MIMO antenna, the link calculations are performed in two different layouts at resonant frequencies 27.10 GHz and 29.53 GHz. For brevity, polarization losses are neglected, and the antennas are assumed to be perfectly matched with 50Ω impedance at both the *mm*-wave frequencies.

Layout-1(uplink): In this scenario, the MIMO antenna is considered to be located on the body surface (chest) and acts as a transmitter (*T_x*); whereas the ideal $\lambda_o/2$ long dipole antenna is assumed to be a receiver (*R_x*) installed in free space at a distance (*d*) from the *T_x*. Note that the transmitted power (*P_t*) is fixed at 10 dBm and 27 dBm, and *R_x* gain (*G_r*) is taken as 2 dBi to simplify the demonstration of uplink communication. Thereafter, the received power (*P_r*) is calculated by (11), where the *T_x* gain is G_t and PL signifies the inevitable free-space path loss. The path loss in (12) can be controlled by the *T_x*−*R_x* distance [27]. Here, n is the free-space path loss exponent. In a multipath propagation environment, n becomes 1.5 for a line of sight (LOS) and 3.0 for non-line of sight (NLOS) communication, respectively. The Gaussian distribution with standard deviation σ gives the shadowing factor X_σ(taken as 0 dBm in this article) and the reference distance (d_0) is taken as 1.0 m [40].

$$P_r(dBm) = P_t(dBm) + G_t(dB) + G_r(dB) - PL_{dB}(dB) \quad (11)$$

$$PL_{dB}(d) = 10n\log_{10}\left(\frac{d}{d_0}\right) + 20\log_{10}\left(\frac{4\pi d}{\lambda_0}\right) + X_\sigma \quad (12)$$

From Fig. 26, it is observed that the MIMO antenna with high transmit power (27 dBm) can enable the dipole antenna with 2 dBi gain to receive more than -50 dBm and -65 dBm power up to 10 m distance during LOS and NLOS link, respectively. Further, when the transmitted power is reduced to 10 dBm, the dipole antenna can still sufficiently receive more than -60 dBm and -80 dBm power up to 5 m distance during LOS and NLOS link, respectively. Moreover, the received powers at both the frequencies are comparable because at higher frequencies, the high radiation efficiency and path loss both get balanced. As per the SAR analysis, the maximum transmit power of 16 dBm can be delivered to the MIMO antenna within the safe limits; therefore, our investigation concludes that the proposed antenna can conduct reliable uplink communication with the highly sensitive *mm*-wave base-station transceivers up to a long and sufficient distance of 10 m.

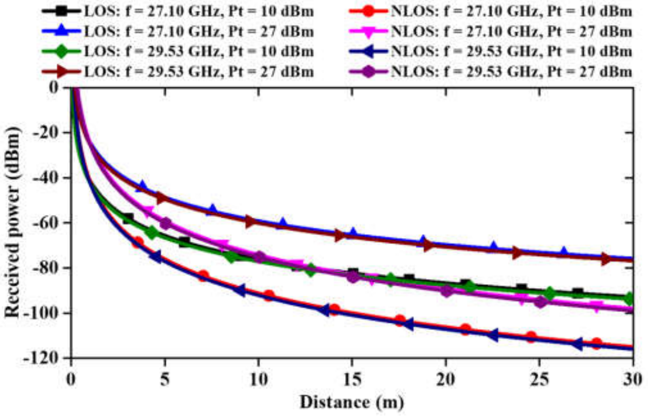


Fig. 26 Received power by the dipole antenna as a function of *Tx*−*Rx* distance in LOS and NLOS uplink communication.

Layout-2(downlink): In this scenario, the wearable MIMO antenna on the body surface (chest) acts as a receiver (R_x); whereas an implantable capsule antenna for endoscopy applications is assumed to be a transmitter (T_x) with $P_t = -4$ dBm located inside the stomach at a distance (d) from the R_x. The link margin (LM) is calculated using (13-18) to simplify the demonstration of downlink communication, and the link budget parameters are shown in Table IV. The available power from the T_x antenna is derived from (14), and the required power at the R_x antenna is derived from (15) for different bit rates such as 25 Mbps, 50 Mbps, 75 Mbps, and 100 Mbps.

$$L.M\ (dB) = Link\left(\frac{c}{N_0}\right) - Required\left(\frac{c}{N_0}\right) \quad (13)$$

$$Link\left(\frac{c}{N_0}\right) = EIRP - L_f + G_r - N_0 \quad [dB/Hz] \quad (14)$$

$$Required\left(\frac{c}{N_0}\right) = \frac{E_b}{N_0} + 10log_{10}(B_r) - G_c + G_d \quad [dB/Hz] \quad (15)$$

$$L_f = 20\log_{10}\left(\frac{4\pi d}{\lambda}\right) \quad (dB) \quad (16)$$

$$N_0 = 10log_{10}(k) + 10log_{10}(T_i) \quad [dB/Hz] \quad (17)$$

$$T_i = T_0(NF - 1) \quad [K] \quad (18)$$

TABLE IV
LINK BUDGET PARAMETERS FOR DOWNLINK COMMUNICATION

Transmitter (Implant Antenna)	
Frequency, f (GHz)	27.10/29.53
Transmit power, P_t (dBm)	-4
Transmit antenna gain, G_t (dBi)	-27
EIRP (P_t + G_t) (dBm)	-31
Propagation	
Distance, d (m)	0-5
Free space loss, L_f (dB)	Adaptive (distance)
Receiver (Wearable Proposed MIMO Antenna)	
Receive antenna gain, G_r (dBi)	4.1
Ambient temperature T_0 (K)	293
Boltzmann constant, k	1.38×10^{-23}
Noise power density, N_0 (dB/Hz)	-203.9
Signal	
Bit rate, B_r (Mb/s)	25, 50, 75, 100
Bit error rate	1×10^{-5}
E_b/N_0 (ideal PSK), (dB)	9.6
Coding gain, G_c (dB)	0
Fixing deterioration, G_d (dB)	2.5

To have a satisfactory communication link, 10-20 dB LM is chosen as a reference in this paper. From Fig. 27, it can be seen that the MIMO antenna can communicate effectively even for a high data rate of 100 Mbps up to a distance of 5 m with 10 dB LM. Note that the LM at a higher frequency is lower relative to a lower frequency for the same data-rate. This behavior can be expected due to the fact that the path loss increases as the frequency increases. Further, when the separation distance reduces to less than 1 m, as is typically the case between the wearable and implantable capsule antenna, the LM improves to 40 dB. This concludes the strong candidature of the proposed MIMO antenna to handle high data rates at *mm*-wave frequencies.

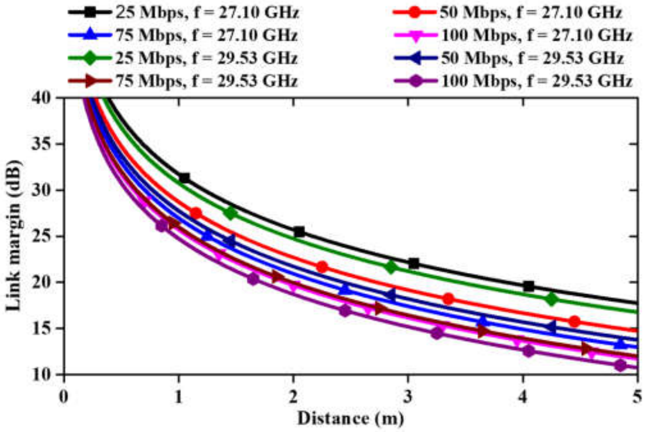


Fig. 27 Calculated link margin with varying data rates between reference implantable antenna and the proposed MIMO antenna.

VII. CONCLUSION

This article, for the first time, reported a semi-flexible circularly polarized two port MIMO antenna for wearable application. The antenna operates in an mm-wave broad frequency range varying from 24 to 31 GHz, with RHCP radiations at 27.10 GHz and LHCP at 29.53 GHz. The 3-dB AR bandwidth varies from 25.84 - 27.35 GHz and 28.57 - 29.85 GHz, respectively. The antenna's performance in bending profiles along *x* and *y*-direction proved to be robust with a marginal shifting of the results to the lower frequency relative to the flat profile as the current on radiating patch gets elongated due to bending. In the on-body study, the MIMO antenna is located 4 mm above the surface of the chest, hand, and leg of the realistic voxel based Gustav human model. The S-parameters for the case of hand and leg are worst affected with degraded impedance matching as the tissue surface is curved. But, in the case of the chest, the frequency response is least impacted due to the flat tissue surface. Nevertheless, the overall far-field performance stayed intact for on-body, as compared to off-body performance. Also, the diversity parameters of MIMO design are reported to demonstrate the quality of the proposed antenna in terms of ECC, DG, MEG, TARC, CCL, and ME. Further, the measured performance both in free space and on the surface of a custom-made in-vitro phantom is validated through comparison with the simulation. Finally, for safety concerns, the maximum input power of 40 mW and 47 mW is designated for the proposed antenna to avoid excess heating, followed by an on-body and off-body link study. In conclusion, the proposed mm-wave MIMO antenna is an excellent candidate for wearable biotechnologies.

References

[1] D. Anzai *et al.*, "Experimental evaluation of implant UWB-IR transmission with living animal for body area networks," *IEEE Trans. Microw. Theory Tech.*, vol. 62, no. 1, pp. 183–192, 2014, doi: 10.1109/TMTT.2013.2291542.

[2] L. W. Y. Liu, A. Kandwal, Q. Cheng, H. Shi, I. Tobore, and Z. Nie, "Non-invasive blood glucose monitoring using a curved goubau line," *Electron.*, vol. 8, no. 6, pp. 1–12, 2019, doi: 10.3390/electronics8060662.

[3] T. Karacolak, A. Z. Hood, and E. Topsakal, "Design of a dual-band implantable antenna and development of skin mimicking gels for continuous glucose monitoring," *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 4, pp. 1001–1008, 2008, doi: 10.1109/TMTT.2008.919373.

[4] A. Kiourti, "RFID Antennas for Body-Area Applications: From Wearables to Implants," *IEEE Antennas Propag. Mag.*, vol. 60, no. 5, pp. 14–25, 2018, doi: 10.1109/MAP.2018.2859167.

[5] N. Ganeshwaran, J. K. Jeyaprakash, M. G. N. Alsath, and V. Sathyanarayanan, "Design of a Dual-Band Circular Implantable Antenna for Biomedical Applications," *IEEE Antennas Wirel. Propag. Lett.*, vol. 19, no. 1, pp. 119–123, 2020, doi: 10.1109/LAWP.2019.2955140.

[6] M. Ur-Rehman, N. A. Malik, X. Yang, Q. H. Abbasi, Z. Zhang, and N. Zhao, "A Low Profile Antenna for Millimeter-Wave Body-Centric Applications," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6329–6337, 2017, doi: 10.1109/TAP.2017.2700897.

[7] P. Soontornpipit, C. M. Furse, and Y. C. Chung, "Design of implantable microstrip antenna for communication with medical implants," *IEEE Trans. Microw. Theory Tech.*, 2004, doi: 10.1109/TMTT.2004.831976.

[8] M. Joler and M. Boljkovac, "A Sleeve-Badge Circularly Polarized Textile Antenna," *IEEE Trans. Antennas Propag.*, vol. 66, no. 3, pp. 1576–1579, 2018, doi: 10.1109/TAP.2018.2794420.

[9] Z. H. Jiang, Z. Cui, T. Yue, Y. Zhu, and D. H. Werner, "Compact, Highly Efficient, and Fully Flexible Circularly Polarized Antenna Enabled by Silver Nanowires for Wireless Body-Area Networks," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 4, pp. 920–932, 2017, doi: 10.1109/TBCAS.2017.2671841.

[10] R. B. V. B. Simorangkir, Y. Yang, L. Matekovits, and K. P. Esselle, "Dual-Band Dual-Mode Textile Antenna on PDMS Substrate for Body-Centric Communications," *IEEE Antennas Wirel. Propag. Lett.*, vol. 16, no. c, pp. 677–680, 2017, doi: 10.1109/LAWP.2016.2598729.

[11] R. B. V. B. Simorangkir, Y. Yang, K. P. Esselle, and B. A. Zeb, "A Method to Realize Robust Flexible Electronically Tunable Antennas Using Polymer-Embedded Conductive Fabric," *IEEE Trans. Antennas Propag.*, vol. 66, no. 1, pp. 50–58, 2018, doi: 10.1109/TAP.2017.2772036.

[12] R. C. Webb *et al.*, "Ultrathin conformal devices for precise and continuous thermal characterization of human skin," *Nat. Mater.*, vol. 12, no. 10, pp. 938–944, 2013, doi: 10.1038/nmat3755.

[13] F. Faisal, Y. Amin, Y. Cho, and H. Yoo, "Compact and Flexible Novel Wideband Flower-Shaped CPW-Fed Antennas for High Data Wireless Applications," *IEEE Trans. Antennas Propag.*, vol. 67, no. 6, pp. 4184–4188, 2019, doi: 10.1109/TAP.2019.2911195.

[14] A. Y. I. Ashyap *et al.*, "Inverted e-shaped wearable textile antenna for medical applications," *IEEE Access*, vol. 6, no. c, pp. 35214–35222, 2018, doi: 10.1109/ACCESS.2018.2847280.

[15] T. K. Roshna, U. Deepak, V. R. Sajitha, K. Vasudevan, and P. Mohanan, "A compact UWB MIMO antenna with reflector to enhance isolation," *IEEE Trans. Antennas Propag.*, vol. 63, no. 4, pp. 1873–1877, 2015, doi: 10.1109/TAP.2015.2398455.

[16] J. Deng, J. Li, L. Zhao, and L. Guo, "A Dual-Band Inverted-F MIMO Antenna with Enhanced Isolation for WLAN Applications," *IEEE Antennas Wirel. Propag. Lett.*, vol. 16, no. c, pp. 2270–2273, 2017, doi: 10.1109/LAWP.2017.2713986.

[17] L. Wang *et al.*, "Compact UWB MIMO Antenna with High Isolation Using Fence-Type Decoupling Structure," *IEEE Antennas Wirel. Propag. Lett.*, vol. 18, no. 8, pp. 1641–1645, 2019, doi: 10.1109/LAWP.2019.2925857.

[18] C. X. Mao, Y. Zhou, Y. Wu, H. Soewardiman, D. H. Werner, and J. S. Jur, "Low-Profile Strip-Loaded Textile Antenna with Enhanced Bandwidth and Isolation for Full-Duplex Wearable Applications," *IEEE Trans. Antennas Propag.*, vol. 68, no. 9, pp. 6527–6537, 2020, doi: 10.1109/TAP.2020.2989862.

[19] M. Mirmozafari, G. Zhang, C. Fulton, and R. J. Doviak, "Dual-Polarization Antennas with High Isolation and Polarization Purity: A Review and Comparison of Cross-Coupling Mechanisms," *IEEE Antennas Propag. Mag.*, vol. 61, no. 1, pp. 50–63, 2019, doi: 10.1109/MAP.2018.2883032.

[20] R. N. Tiwari, P. Singh, B. K. Kanaujia, and P. Kumar, "Compact circularly polarized MIMO printed antenna with novel ground structure for wideband applications," *Int. J. RF Microw. Comput. Eng.*, vol. 31, no. 8, Aug. 2021, doi: 10.1002/mmce.22737.

[21] Z. Ren, A. Zhao, and S. Wu, "MIMO Antenna With Compact Decoupled Antenna Pairs for 5G Mobile Terminals," *IEEE Antennas Wirel. Propag. Lett.*, vol. 18, no. 7, pp. 1367–1371, 2019, doi: 10.1109/LAWP.2019.2916738.

[22] A. Iqbal, A. Smida, A. J. Alazemi, M. I. Waly, N. K. Mallat, and S. Kim, "Wideband circularly polarized MIMO antenna for high data wearable biotelemetric devices," *IEEE Access*, vol. 8, no. January, pp. 17935–17944, 2020, doi: 10.1109/ACCESS.2020.2967397.

[23] H. Li, S. Sun, B. Wang, and F. Wu, "Design of Compact Single-Layer Textile MIMO Antenna for Wearable Applications," *IEEE Trans. Antennas Propag.*, vol. 66, no. 6, pp. 3136–3141, 2018, doi: 10.1109/TAP.2018.2811844.

[24] A. K. Biswas and U. Chakraborty, "Compact wearable MIMO antenna with improved port isolation for ultra-wideband applications," *IET Microwaves, Antennas Propag.*, vol. 13, no. 4, pp. 498–504, 2019, doi: 10.1049/iet-map.2018.5599.

[25] S. F. Jilani, M. O. Munoz, Q. H. Abbasi, and A. Alomainy, "Millimeter-Wave Liquid Crystal Polymer Based Conformal Antenna Array for 5G Applications," *IEEE Antennas Wirel. Propag. Lett.,* vol. 18, no. 1, pp. 84–88, 2019, doi: 10.1109/LAWP.2018.2881303.

[26] A. Iqbal *et al.*, "Electromagnetic bandgap backed millimeter-wave MIMO antenna for wearable applications," *IEEE Access*, vol. 7, pp. 111135–111144, 2019, doi: 10.1109/ACCESS.2019.2933913.

[27] M. Faridani, M.C.E. Yagoub and R.E. Amaya , "Novel Body Matched Millimeter-Wave Sandwiched Antenna for Advanced Medical Communications," *2020 International Conference on Microwave and Millimeter Wave Technology (ICMMT)*, pp. 5–7, 2020. doi:10.1109/ICMMT49418.2020.9387026.

[28] P. Njogu, B. Sanz-Izquierdo, A. Elibiary, S. Y. Jun, Z. Chen, and D.

Bird, “3D Printed Fingernail Antennas for 5G Applications,” *IEEE Access*, vol. 8, pp. 228711–228719, 2020, doi: 10.1109/ACCESS.2020.3043045.

[29] D. Wen, Y. Hao, M. O. Munoz, H. Wang, and H. Zhou, “A Compact and Low-Profile MIMO Antenna Using a Miniature Circular High-Impedance Surface for Wearable Applications,” *IEEE Trans. Antennas Propag.*, vol. 66, no. 1, pp. 96–104, 2018, doi: 10.1109/TAP.2017.2773465.

[30] L. Qu, H. Piao, Y. Qu, H. H. Kim, and H. Kim, “Circularly polarised MIMO ground radiation antennas for wearable devices,” *Electron. Lett.*, vol. 54, no. 4, pp. 189–190, 2018, doi: 10.1049/el.2017.4348.

[31] Z. H. Jiang, M. D. Gregory, and D. H. Werner, “Design and Experimental Investigation of a Compact Circularly Polarized Integrated Filtering Antenna for Wearable Biotelemetric Devices,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 2, pp. 328–338, 2016, doi: 10.1109/TBCAS.2015.2438551.

[32] Z. H. Jiang and D. H. Werner, “A Compact, Wideband Circularly Polarized Co-designed Filtering Antenna and Its Application for Wearable Devices with Low SAR,” *IEEE Trans. Antennas Propag.*, vol. 63, no. 9, pp. 3808–3818, 2015, doi: 10.1109/TAP.2015.2452942.

[33] I. S. C. C. 28, IEEE Standards Coordinating Committee, I. S. C. C. 28, and IEEE Standards Coordinating Committee, *IEEE Recommended Practice for Measurements and Computations of Radio Frequency Electromagnetic Fields with Respect to Human Exposure to Such Fields*, vol. 2002. 2003.

[34] S. Velan *et al.*, “Dual-band EBG integrated monopole antenna deploying fractal geometry for wearable applications,” *IEEE Antennas Wirel. Propag. Lett.*, vol. 14, no. c, pp. 249–252, 2015, doi: 10.1109/LAWP.2014.2360710.

[35] Y. M. Cai, S. Gao, Y. Yin, W. Li, and Q. Luo, “Compact-Size Low-Profile Wideband Circularly Polarized Omnidirectional Patch Antenna With Reconfigurable Polarizations,” *IEEE Trans. Antennas Propag.*, vol. 64, no. 5, pp. 2016–2021, 2016, doi: 10.1109/TAP.2016.2535502.

[36] R. Garg, P. Bhartia, I. Bahl, and A. Ittipiboon, “Microstrip antenna design handbook,” *Artech House*: London, 2001.

[37] D. Yu, S. X. Gong, Y. T. Wan, and W. F. Chen, “Omnidirectional dual-band dual circularly polarized microstrip antenna Using TM01 and TM02 modes,” *IEEE Antennas Wirel. Propag. Lett.*, vol. 13, pp. 1104–1107, 2014, doi: 10.1109/LAWP.2014.2328020.

[38] H. F. Abutarboush, W. Li, and A. Shamim, “Flexible-Screen-Printed Antenna with Enhanced Bandwidth by Employing Defected Ground Structure,” *IEEE Antennas Wirel. Propag. Lett.*, vol. 19, no. 10, pp. 1803–1807, 2020, doi: 10.1109/LAWP.2020.3019462.

[39] K. Sreelakshmi, G. S. Rao, and M. N. V. S. S. Kumar, “A compact grounded asymmetric coplanar strip-fed flexible multiband reconfigurable antenna for wireless applications,” *IEEE Access*, vol. 8, pp. 194497–194507, 2020, doi: 10.1109/ACCESS.2020.3033502.

[40] T. S. P. See, T. M. Chiam, M. C. K. Ho, and M. R. Yuce, “Experimental study on the dependence of antenna type and polarization on the link reliability in on-body UWB systems,” *IEEE Trans. Antennas Propag.*, vol. 60, no. 11, pp. 5373–5380, 2012, doi: 10.1109/TAP.2012.2208611.

[41] A. T. Mobashsher, “Artificial Human Phantoms,” *IEEE Microw. Mag.*, no. July, pp. 42–62, 2015.

[42] V. Kaim, B. K. Kanaujia, and K. Rambabu, “Quadrilateral Spatial Diversity Circularly Polarized MIMO Cubic Implantable Antenna System for Biotelemetry,” *IEEE Trans. Antennas Propag.*, vol. 69, no. 3, pp. 1260–1272, 2021, doi: 10.1109/TAP.2020.3016483.

[43] U. Ullah, M. Al-Hasan, S. Koziel, and I. Ben Mabrouk, “A Series Inclined Slot-Fed Circularly Polarized Antenna for 5G 28 GHz Applications,” *IEEE Antennas Wirel. Propag. Lett.*, vol. 20, no. 3, pp. 351–355, 2021, doi: 10.1109/LAWP.2021.3049901.

[44] Z. Xia *et al.*, “A Wideband Circularly Polarized Implantable Patch Antenna for ISM Band Biomedical Applications,” *IEEE Trans. Antennas Propag.*, vol. 68, no. 3, pp. 2399–2404, 2020, doi: 10.1109/TAP.2019.2944538.

[45] A. Ramachandran, S. Mathew, V. P. Viswanathan, M. Pezholil, and V. Kesavath, “Diversity-based four-port multiple input multiple output antenna loaded with interdigital structure for high isolation,” *IET Microwaves, Antennas Propag.*, vol. 10, no. 15, pp. 1633–1642, 2016, doi: 10.1049/iet-map.2015.0828.

[46] R. N. Tiwari, P. Singh, B. K. Kanaujia, and K. Srivastava, “Neutralization technique based two and four port high isolation MIMO antennas for UWB communication,” *AEU - Int. J. Electron. Commun.*, vol. 110, p. 152828, 2019, doi: 10.1016/j.aeue.2019.152828.

[47] M. Manteghi and Y. Rahmat-Samii, “Multiport characteristics of a wide-band cavity backed annular patch antenna for multipolarization operations,” *IEEE Trans. Antennas Propag.*, vol. 53, no. 1 II, pp. 466–474, 2005, doi: 10.1109/TAP.2004.838794.

[48] Y. K. Choukiker, S. K. Sharma, and S. K. Behera, “Hybrid fractal shape planar monopole antenna covering multiband wireless communications with MIMO implementation for handheld mobile devices,” *IEEE Trans. Antennas Propag.*, vol. 62, no. 3, pp. 1483–1488, 2014, doi: 10.1109/TAP.2013.2295213.

[49] R. Tian, B. K. Lau, and Z. Ying, “Multiplexing efficiency of MIMO antennas in arbitrary propagation scenarios,” *Proc. 6th Eur. Conf. Antennas Propagation, EuCAP 2012*, vol. 10, pp. 373–377, 2012, doi: 10.1109/EuCAP.2012.6205897.

[50] A. Kiourti, J. R. Costa, C. A. Fernandes, and K. S. Nikita, “A broadband implantable and a dual-band on-body repeater antenna: Design and transmission performance,” *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 2899–2908, 2014, doi: 10.1109/TAP.2014.2310749.

Rakesh Nath Tiwari completed B.Sc. and M.Sc. (Electronics) degree from University of Allahabad and Deen Dayal Upadhyaya Gorakhpur University, Gorakhpur, India in 2002 and 2004 respectively. He received M.Tech. degree in Optical & Wireless Communication Technology with Gold Medal from Jaypee University of Information Technology, Waknaghat, Solan, India in 2008. He has completed Ph.D from Uttarakhand Technical University, Department of Electronics and Communication Engineering, Dehradun, India, in August 2020. Presently he is working as Assistant Professor, at Madanapalle Institute of Technology & Science, Madanapalle, Andhra Pradesh, India. He has published more than 30 papers in peer reviewed International/National journals and conferences with more than 350 citations with h-index of 10. He is life member of Materials Research Society of India (MRSI), and International Association of Engineers (IAENG). He is reviewer of many reputed journals and his research interest includes design and modelling of slot patch antennas, UWB antennas, circularly polarized microstrip antennas, MIMO antenna, wearable antennas, capsule endoscopy and microwave/millimeter wave integrated circuits & devices.

Vikrant Kaim (Member, IEEE) was born in New Delhi, India, in 1990. He received the B.Tech. degree in electronics and communication engineering from the Bharati Vidyapeeth’s College of Engineering, New Delhi, in 2014, and the M.Tech. degree in microwave electronics from the Department of Electronic Science, University of Delhi South Campus, New Delhi, in 2016. He received Ph.D. degree in computational biology and bioinformatics with Jawaharlal Nehru University, New Delhi in 2022. He has published five journal articles and two conference papers. His research interests include electromagnetic theory, implantable antennas and devices, wearable antennas, printed antennas, flexible antennas, multi-input multi-output (MIMO) antennas, wireless power transfer, pacemaker systems, and capsule endoscopy. He has been a recipient of the CSIR Senior Research Fellowship since 2019.

Prabhakar Singh was born in village Semara, Chandauli (U.P), India in 1984. He completed his B.Sc. and M.Sc. degree in Physics from V. B. S. Purvanchal University in 2004 and 2006, respectively. He received his Ph.D. degree from J. K. Institute of Applied Physics, Department of Electronics and Communication, University of Allahabad, India in 2010. He has worked as Lecturer at Delhi Technological University, Delhi, India for one year from 2009 to 2010. He has worked as Assistant Professor at Bahra University, Shimla Hills, Himanchal Pradesh, India from 2010 to 2011. Dr. Singh has joined Galgotias University in 2011 and presently he is working as Professor and division chair of Physics, Galgotias University, Greater Noida, India. He has published more than 65 research papers in peer reviewed International/National journals and conference proceedings with more than 900 citations and h-index of 18. He has published two patents and also the reviewer of many International and National journals. He had supervised more than 7 M.Tech./M.Sc. students and currently supervising 2 Ph.D. research scholars in the field of electromagnetics. He is presently working on antennas for biotechnology applications, size miniaturization techniques in patch antenna, UWB antenna, MIMO designs, photonic band gap antennas and RF harvesting.

Taimoor Khan (Senior Member, IEEE) received his Ph.D. in Electronics and Communication Engineering, National Institute of Technology Patna. Dr. Khan joined the Department of Electronics and Communication Engineering at the NIT Silchar in November 2014 and presently he is working as an Associate Professor. Before that, he has served at different organizations in different capacities like Lab Instructor, Lecturer, Assistant Professor and Associate Professor for more than 14 years. Dr. Khan has also worked as Visiting Researcher at Queen's University Canada and as a Visiting Assistant Professor at Asian Institute of Technology, Bangkok, Thailand. His active research interest includes Ambient RF Energy Harvesting, Microwave Power Transfer, Ultra-Wideband Technologies, EBG Structures and Microwave Components. Dr. Khan has published more than 70 research articles and 51 papers in renowned Journals/ proceedings. He has completed one major project and two minor projects funded by SERB, MHRD and AICTE, respectively. At present, he has collaborative research projects with Queen's University Canada and California State University, Northridge USA, financially sponsored by IEEE, MHRD and SERB, respectively. He is a Fellow of Institution of Engineers (India), Fellow of Institution of Electronics and Telecommunication Engineers India, Senior Member of IEEE USA, IEEE AP Society USA, IEEE MTT Society, USA and URSI. Dr. Khan et al. has edited seven books including two research books. He was recipient of a prestigious IETE-Prof SVC Aiya Memorial Award for the year 2020. He is actively involved in IEEE activities and most recently, he has been nominated as Co-Chair, IEEE APS Best Paper Award Committee 2023.

Binod Kumar Kanaujia (Senior Member, IEEE) is working as Director (deputation), Dr. B.R Ambedkar National Institute of Technology Jalandhar, Punjab since Feb. 2022. He is also a Professor in School of Computational and Integrative Sciences (SC & IS), Jawaharlal Nehru University (JNU), New Delhi since August, 2016. He held the position of Dean, SC & IS and chief advisor of the Equal Opportunity Officer, JNU. Before joining Jawaharlal Nehru University, he had been in the Department of Electronics & Communication Engineering in Ambedkar Institute of Advanced Communication Technologies & Research (formerly, Ambedkar Institute of Technology), Delhi as a Professor since Feb. 2011 & Associate Professor (2008-2011). Earlier, he held various positions at M.J.P. Rohilkhand University, Bareilly, India. Prior to his career in academics, he had worked as Executive Engineer in the R&D division of M/s UPTRON India Ltd. He had completed his B.Tech. in Electronics Engineering from KNIT Sultanpur, India in 1994. He did his M.Tech. and Ph.D. in 1998 and 2004; respectively from Department of Electronics Engineering, Indian Institute of Technology, BHU, Varanasi, India. He has been awarded JRF by UGC Delhi in the year 2001-02 for his outstanding work in the electronics field. His research interest includes in design and modelling of microstrip antenna, DRA, UWB antennas, reconfigurable and circular polarized antennas for wireless communication. He has been credited to publish more than 430 research papers, 3 books, 11 book chapters and 11 patents with more than 5700 citations with h-index of 34 in several peer-reviewed journals and conferences. He had supervised 50 M.Tech. and 35 Ph.D. research scholars in the field of microwave engineering. He is a reviewer for several journals of international repute. Dr. Kanaujia had successfully executed 06 research projects with 03 on-going projects sponsored by several agencies of Government of India i.e. DRDO, DST, DBT, SERB, AICTE and ISRO. He is also a senior member of IEEE, Fellow of Institute of Engineers (India), also member of several academic and professional bodies. He is Associate Editor in IEEE ACCESS, Editor of AEU-International Journal of Electronics and Communication, and IETE Technical Review.

Settlement in Geosynthetic Reinforced Square Footing over Plastic Soil

Ankur Mudgal[1], Bibek Jha[2], Raju Sarkar[3], Amit Kumar Srivastava[4], Akshit Mittal[5] and Nehal Jain[6]

[1]Senior Manager, CEG Test House & research Center PVT. LTD, Jaipur, India
E-mail: ankur.mudgal@cegtesthouse.com (PhD, DTU, Delhi)

[2]Senior Vice President, CEG Test House & research Center PVT. LTD, Jaipur, India
E-mail: bibek.jha@cegtesthouse.com (M. Tech, IIT, Bombay)
[3] Professor, Department of Civil Engineering, Delhi Technological University, Delhi, India (PhD, Jamia Millia Islamia, Delhi)
[4] Professor, Department of Civil Engineering, Delhi Technological University, Delhi, India (PhD, IIT, Delhi)

[3]UG Student, Department of Civil Engineering, Delhi Technological University, Delhi, India
E-mail: akshit97mittal@gmail.com
[4]General Manager, CEG Test House & research Center PVT. LTD, Jaipur, India

sarkar_raju@yahoo.co.in, aksrivastava@dce.ac.in,
nehal.jain@cegtesthouse.com

Abstract. Construction over Plastic soil can cause adverse effects on the performance of the earth structures, due to the low load carrying capacities of such soils. Many civil engineering structures like buildings, Major and Minor bridges, Under passes, and Flyovers collapse and undergo crack formation in areas where Plastic soil with poor load carrying capacity is present. Geosynthetic reinforcements have successfully been used in recent times as a low cost method for reinforcing such soils to improve their stability and bearing capacity. However, to recover significant benefit from the geosynthetics, the materials need to be placed at optimum locations within the foundation. Hence, in this paper, small scale laboratory footing tests have been performed to study the effect of depth of the first layer of reinforcement (u), Number of reinforcement layers (N), width of reinforcement (b) and the vertical spacing between reinforcements (h). The results obtained demonstrated that the placement and the loading condition of the geosynthetics greatly influences the bearing capacity of the foundation. The results obtained from the experimental analysis were used for the computation of a regression model in R Studio, for the determining the load carrying capacity of reinforced Soil foundation. The model presented obtained a confidence level of more than 95%, when parameters significant for the computation of load carrying capacity of square footing were included, thus showing great convergence with the experimental results

TH-03-068

Keywords: Plastic soil; Square footing; Geogrid; Geotextile.

1. Introduction

In Geotechnical engineering, failure of a foundation constructed over Plastic soil is a very prevalent issue. The foundation may lose its stability due to poor load carrying capacity of the surrounding soil, which results in higher settlements than the acceptable values and deterioration in the load carrying capacity of the soil. Many researchers have proffered many solutions to mitigate these problems, through different soil improvement techniques. The first comprehensive study on soil reinforcement was conducted by [1], wherein aluminium strips were used to reinforce the sand beds. [2] examined the performance of strip footings on geogrid reinforced sand bed over a soft clay slope. The study showed that the use of geogrid layers in the replaced sand not only significantly improves the performance of footing but also leads to high reduction in the depth of reinforced sand layer required to achieve the allowable settlement. [3] carried out the laboratory model footing tests on reinforced soil bed and reported their results as a comparative study of effectiveness of geogrid and geotextile as soil reinforcement. The results showed that the geogrid is more efficient than the geotextile in respect of bearing capacity of foundation on reinforced sand. In in this paper, small scale laboratory tests have been performed and obtained results are presented.

2. Material used:

2.1 Soil: The geotechnical properties of the soil were determined as per Indian standards listed in Table 1. The soil was classified as a low compressible soil.

2.2 Geosynthetics: A single type of geogrid, and geotextile were used in this study. The geogrid used was bi-oriented and was made of polypropylene thermoplastic, whereas, the Geotextile used was woven type also made of polypropylene material. The material testing certificate of geosynthetics as provided by the manufacturer is tabulated in Table 2. The geogrid and geotextile in the study have been represented by GGR and GTX respectively, throughout the study.

Table 1. Geotechnical properties of soil

Properties	Values	Protocols/Standards
Specific Gravity	2.67	IS 2720 (Part III)
Liquid limit (%)	29	IS 2720 (Part V)
Plastic limit (%)	20	IS 2720 (PartV)
Plasticity index (%)	9	IS 2720 (Part V)
Maximum dry Unit Weight (kN/m^3)	17.6	IS 2720(Part VII)

Optimum Moisture Content (%)	15	IS 2720(Part VII)
Angle of Friction	22˚	IS 2720 (Part XI)

Table 2. Technical characteristics of geosynthetics (Courtesy: Supplier's Data)

Geosynthetic	Property	Data	Unit	Test Method
Geogrid	Mesh Type	Quadrangular apertures	-	-
	Polymer Type	Polypropylene	-	-
	Aperture Size	30 × 30 (MD × CMD)	mm	-
	Stiffness at 0.5% Strain	550 × 350 (MD × CMD)	kN/m	ISO-10319
Geotextile	Tensile Strength	475 × 384 (MD × CMD)	kN/m	IS 1969
	Opening Size	0.075	mm	ASTM D4751
	Weight of fabric	200	g/m^2	ASTM D5261
	Elongation at break	30 × 28 (MD × CMD)	(%)	IS 1969

3. Experimental Setup

3.1 Preparation of test bed

Test bed was prepared with dimensions of 750 mm × 450 mm × 600 mm (L × B × H). Initially the soil was air dried and pulverized and then it was compacted at its maximum dry unit weight of 17.6 kN/m^3. Predetermined water content was thoroughly mixed to soil to achieve the optimum moisture content, i.e. 15%. In case of unreinforced condition, soil was compacted in three lifts, whereas for the reinforced case, the thickness of each lift was decided according to the spacing between the reinforcements. The soil was then poured into the tank and compacted using a rammer with a base of 150 mm diameter up to a marked height. During the test, the height of fall of the rammer, number of blows to be given, and the required amount of soil sample was determined to maintain the condition of uniformity of soil sample in the tank. At the end of compaction, a spirit level was used to check the alignment of the horizontal surface of prepared test bed.

3.2 Layout of Geosynthetics

The geosynthetic configurations were decided according to testing procedure described in the testing programme i.e. at the respective u/B, b/B, h/B, ratios, and number of reinforcement layers N. All the Five different dimensionless parameters (i.e. u/B, b/B, d/B, N and h/B) were varied to ascertain the optimums in geosynthetic placements. Geometry of geosynthetics reinforced bed is shown in Fig 1.

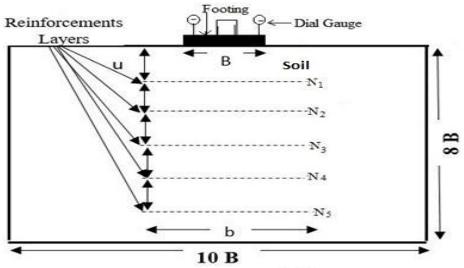
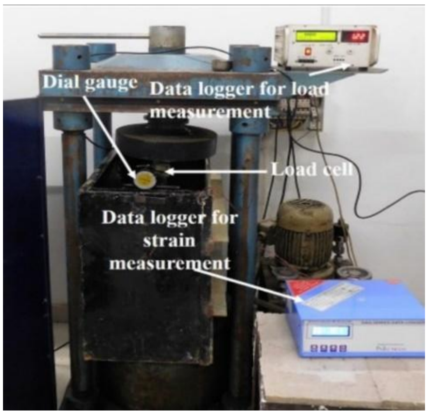


Fig. 1 Geometry of geosynthetics reinforced bed

3.3 Test Setup

A number of laboratory footing tests were conducted on unreinforced and reinforced soil by using testing facilities developed at laboratory. The dimensions of the test set up are as follows: Length = 750 mm, width = 450 mm and height = 600 mm. The back and sides of the tank was fabricated from 20 mm thick steel sheet braced with structural steel member whereas front side of the tank consisted of acrylic sheet of 20mm thickness for visual observation. An angle was fixed on the face of acrylic sheet to prevent its buckling during the compaction and loading. Inner surface of tank was greased to prevent the adverse effect of friction on the test results. A steel plate of size 75 mm × 75 mm was used as a model footing and the dimensioning of the tank were done in accordance with the footing width, so as to avert the boundary effect. Hence, dimensions of 10B, 6B, and 8B were chosen as the length, width and height of the tank respectively, where B is the model footing width, i.e. 75 mm. The base of the footing was kept rough by gluing the sand with epoxy glue. The tank was tested in loading frame consisting of two rigid and heavy steel plate columns of thickness 150mm attached to top head of the loading frame. A load cell of 25kN capacity was placed at centre between the footing plate and upper platen to avoid the eccentric loading. The output of the load cell was logged using a data logger in the form of pressure. Two dial gauges with accuracy of 0.001mm were used at points diametrically opposite to the footing. Average reading obtained by both the dial gauges was considered for settlement analysis. The test bed was tested as per the provision of (ASTM 1997) up to a settlement of 20 mm. where the load increments were applied and maintained at the obtained value until the rate of settlement was less than 0.03 mm/min over three consecutive minutes. Sitting load was applied initially over the footing to fix the footing over the soil base, so as to obtain planar strain conditions.



Dial gauge
Data logger for load measurement
Load cell
Data logger for strain measurement

Fig. 2 General arrangements of testing setup

3.4 Testing Procedure

The aim of the study was to investigate the effect of reinforcements in bearing capacity of reinforced soil foundation. The model tests were conducted with both the reinforcements i.e. GGR, and GTX. The detailed experimental programme is shown in Table 3. Test series"A" was carried out on unreinforced soil footing bed to compare the results obtained by tests conducted on reinforced soil bed. Two series of tests (B and C) were conducted for both the reinforcements i.e. GGR and GTX. Initially, at reinforcement width equal to 5B, the u/B values were varied form 0.17, 0.34, 0.51, 0.68 and 0.85 respectively. Effect of number of reinforcement layers was estimated by fixing the top layer at maxima obtained from the previous test and varying the number of reinforcements until the effect of the reinforcement becomes diminished, or becomes considerably insignificant for any further extensions in number of reinforcement layers. The vertical spacing between the reinforcement layers was also fixed at optimum u/B value. Similar testing procedure were adopted for different reinforcement widths i.e. 4B and 6B. The effect of spacing between two corresponding reinforcements was analysed by varying the distance between two reinforcements by a factor of 0.08 B, 0.16 B, 0.24 B, 0.32 B and 0.4 B respectively and fixing the top layer at the optimum obtained from the previous tests.

Table 3 Experimental programme

Test Series	**Reinforcement**	***N***	***u/B***	***b/B***	**h/B**	**No. of test**	**Remarks**
A	Unreinforced soil	-	-	-	-	3	To estimate the improvements due to reinforcement
B	GGR	1	0.17,	5B	-	5	To find out the

Test Series	**Reinforcement**	***N***	***u/B***	***b/B***	**h/B**	**No. of test**	**Remarks**
			0.34,0.51 ,0.68,0.85				optimum *u/B* value
		1,2,3,4, 5	Optimum depth	5B	0.34B	5	To find the effect of number of layers of geogrid
		Optimum depth	0.34	4B, 5B, 6B		15	To check the optimum values of geogrid width and number of geogrid layers.
		2	0.34	5B	0.08 B, 0.16 B, 0.24B, 0.32 B and 0.4 B	5	To check the effect of vertical spacing between reinforcements
C	GTX	1	0.17, 0.34, 0.51, 0.68,0.85	5B		5	To find out the optimum *u/B* value
		1,2,3,4, 5	Optimum depth	5B	0.34B	5	To find the effect of number of layers of geogrid
		Optimum depth	0.34	4B, 5B, 6B		15	To check the optimum values of geotextile width and number of geotextile layers.
		2	0.34	5B	0.08 B,0.16 B, 0.24 B, 0.32 B and 0.4 B	5	To check the effect of vertical spacing between reinforcements

4. Results and Discussion

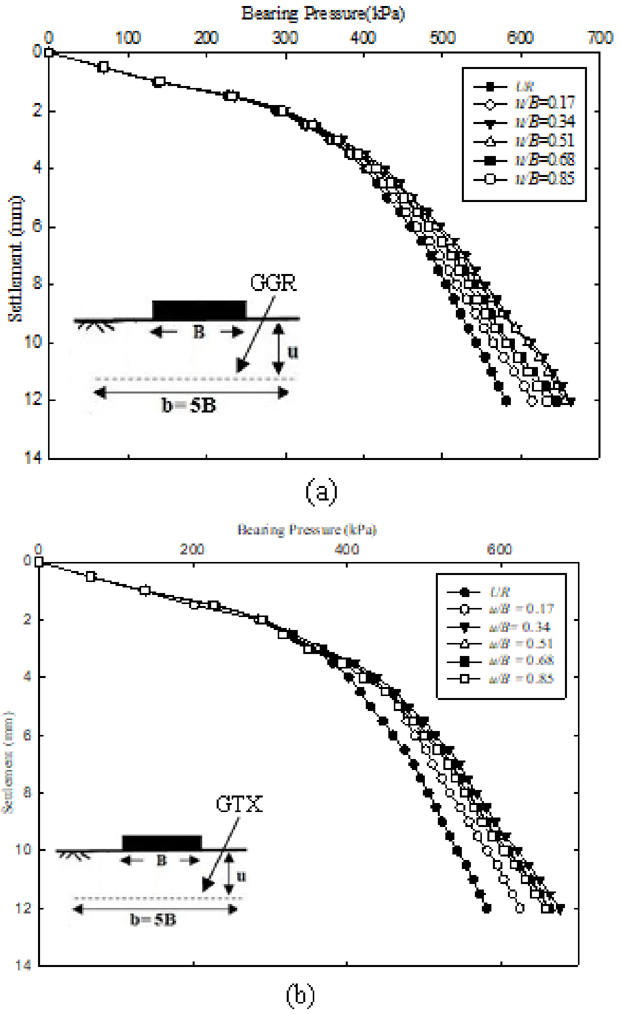
4.1 Determination of optimum depth of first layer of reinforcement

The settlement measured from the dial gauges is considered as footing settlement and denoted as (s). The Ratio of footing settlement (s) to the width of footing (B) is defined as settlement ratio (SR), indicated in percentage. A non-dimensional parameter bearing capacity ratio (BCR) was calculated at each settlement ratio to give an incisive account of the improvement in load carrying capacity due to inclusion of reinforcement in the Plastic soil bed. The BCR is defined as the ratio of the bearing pressure of a reinforced soil to that of an unreinforced soil, when evaluated at the same settlement ratio. In this study, BCR was calculated at four different settlement ratios i.e., UBC, 4%, 8%, 12% and 16%. It should be noted that the BCR is similar to an improvement factor used by many researchers in their studies [4-5]. Another parameter Settlement reduction factor (SRF) being also used in the study, SRF can be calculated as

$$\text{SRF} = 1 - \frac{(S)r}{(S)ur} \times 100 \quad \text{for s/B= UBC, 4\%, 8\%, 12\% and 16\%}$$

Where $(S)_{ur}$ = Settlement of unreinforced soil and $(S)_{r}$ = the settlement of reinforced soil bed at bearing pressure with respect to $(S)_{ur}$. This Settlement reduction factor (SRF) is similar to percentage reduction settlement (PRS) used by [6-7] in their studies to quantify the performance improvement in settlement in terms of percentage. Many researchers have also considered settlement ratio (SR) as a parameter to compare the settlement reduction with the application of geosynthetics. Foundations are designed in accordance to the allowable bearing pressure of the soil. Thus, computation of ultimate bearing capacity (UBC) becomes substantial to correctly assess the increase in construction viability of the soil with applications of geosynthetics. To determine the optimum value of first depth of reinforcement layer, initially five tests were conducted on square model footing, supported by single layer of each geosynthetic. Fig. 3 (a-b) shows the pressure settlement curves for GGR and GTX respectively. As can be deciphered from the curves that bearing pressure of soil increases as the ratio of u/B increases. However the rate of increases in the bearing pressure is significant until a value of u/B = 0.34 after which bearing pressure rapidly decreases with increasing the u/B value. Fig 4 (a-b) depict the improvement factor versus u/B for GGR and GTX respectively. It can be observed from the graph that the BCR gradually increases as u/B value increase from 0.17 to 0.34, afterwards a decrease in BCR can be observed with increases in u/B. [8] reported somewhat similar findings that the bearing capacity of a square footing on a geogrid reinforced Soil bed improved significantly to a depth of placement of u/B =0.33. The probable reason of these optimum values of u/B is that when u/B<

0.34, the surcharge pressure was not sufficient to generate the friction at soil - reinforcement interface.



(a)

(b)

Fig.3 Pressure- settlement curves for (a) Geogrid (b) Geotextile at different u/B ratios

GGR

(a)

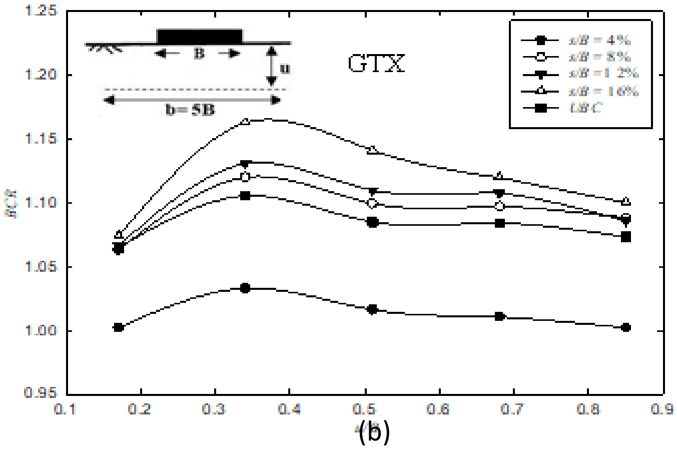


Fig. 4 Bearing Capacity Ratio (BCR) for (a) Geogrid (b) Geotextile at different u/B ratio

4.2 Effect of number of reinforcement layers and effective depth:

For ascertaining the effect of increasing number of reinforcements on the bearing capacity of the soil, the top layer of reinforcement was fixed at 25.5 mm (i.e. u/B =0.34) and then reinforcement layer was varied by fixing the vertical spacing of 25.5 mm till the effect of reinforcement becomes insignificant. The reinforcement ratio (d), which is defined as the ratio of the total depth of the reinforcement and the width of the footing was ascertained for each reinforcement case. The reinforcement depth below the base of the footing can be expressed as

$$d = u + (N - 1) \times h \tag{1}$$

Where u= first reinforcement depth below the base of footing, h = vertical spacing between two consecutive layers of reinforcement, N= Number of reinforcement layers. The pressure-settlement curves were plotted for each number of reinforcement layer to compare with unreinforced one. Fig.5 (a-b) show the pressure settlement for geogrid and geotextile respectively. As expected, the value of bearing pressure increases with increment in number of reinforcement layers. However, improvement in bearing capacity becomes almost insignificant after the addition of fourth layer and third layer which are located at a depths of 1.36B and 1.02B for geogrid and geotextile respectively. The probable reason of bearing capacity improvement is that the increase in friction at the soil reinforcement interface, which increases with the increase in reinforcement layers. Also, better interlocking between the soil and geogrids and passive earth resistance can be attributed as the reason for the development of the bearing capacity. Better interlocking between the geogrid and the soil prevented the lateral deformation of the soil. As a result of applied load, tension is mobilized in the geosynthetics which resist the shear stresses developed in the soil below the loading area and transfer them to the stable soil, thus eventually increasing the depth of the failure zone thus, results in higher bearing capacity and settlement reduction. Similar findings were also observed by [8], they reported that the inclusion of geogrid reinforcements became insignificant after a 1.33B depth of reinforcement. On the basis of their findings, they reported a maximum of four geogrid layers as optimum reinforcement in case of a strip footing.

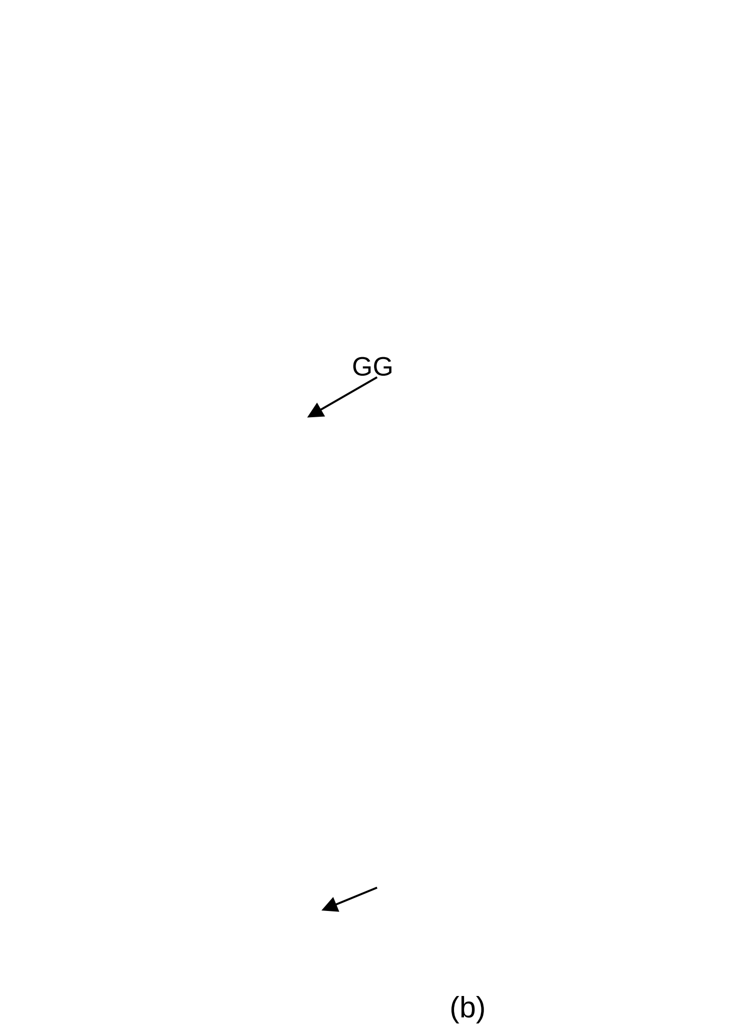
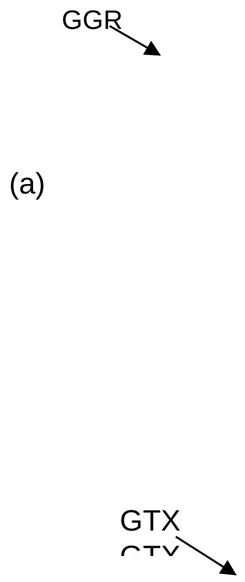


Fig. 5 **Pressure- settlement curves for (a) Geogrid (b) Geotextile at different reinforcement depth**

4.3 Effect of type of reinforcement

Geogrid and geotextile of different stiffness are used in the present study. Technical characteristics of geosynthetics are presented in Table. 3. Reinforcements used in the study were made of same material but the tensile properties of geotextile is higher than the geogrid. In order to investigate the effect of tensile properties of reinforcement in reduction in the development of bearing capacity of soil foundation bed, the values of BCR were estimated at different settlement ratios i.e. (4, 8, 12, and 16% and UBC). Fig. 6 (a-b) depict the variation of BCR with reinforcement depth i.e. d/B for geogrid and geotextile respectively. The nature of the curve may be classified in to two groups; one for settlement level s/B< 4% (lower settlement level) in which ultimate bearing capacity lies and other for s/B>4% (higher settlement level). For the first group, geogrid impart much substantial improvement in bearing capacity than geotextile. The reason for the same can be explained that at lower settlement level, geogrid efficiently mobilized the lateral stress resistance capacity due to the confinement effect which plays a vital role in reinforcement mechanism. However for higher settlement (s/B> 4%), performance of geotextile is much better and gives more improvement than the geogrid. The reason behind this can be explained as, at certain settlement, geotextile requires higher deformation to perform on its full capacity due to its higher tensile strength. Generally, foundation constructions require to be constructed for the ultimate bearing capacity of the soil. Geogrid is better performing material than geotextile for limited settlement requirements or settlement upto ultimate bearing capacity. However, geotextiles primary function is to act as a filter or in drainage system behind retaining walls, adjacent to roads, and within slopes etc. thus can be considered as reinforcement material where small tensile strength is required. [9] presented the similar findings when they compared the geogrids with geotextile at a certain settlement level.



(a)

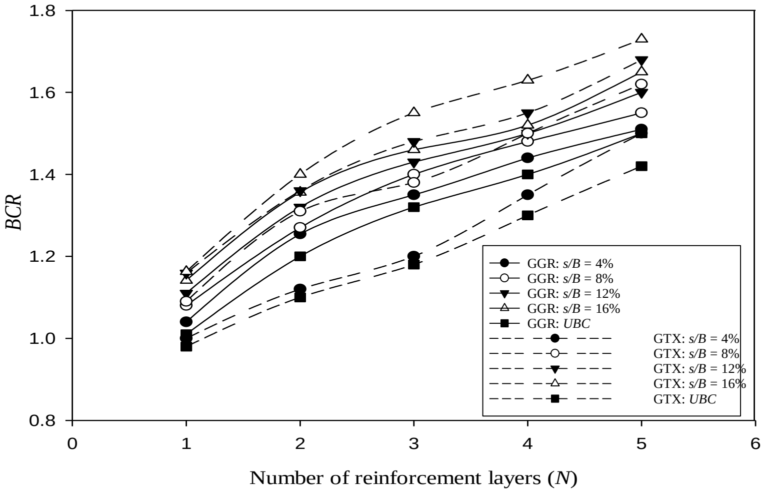
(b)

Fig. 6 Variation of BCR vs reinforcement depth (d) for (a) Geogrid (b) Geotextile

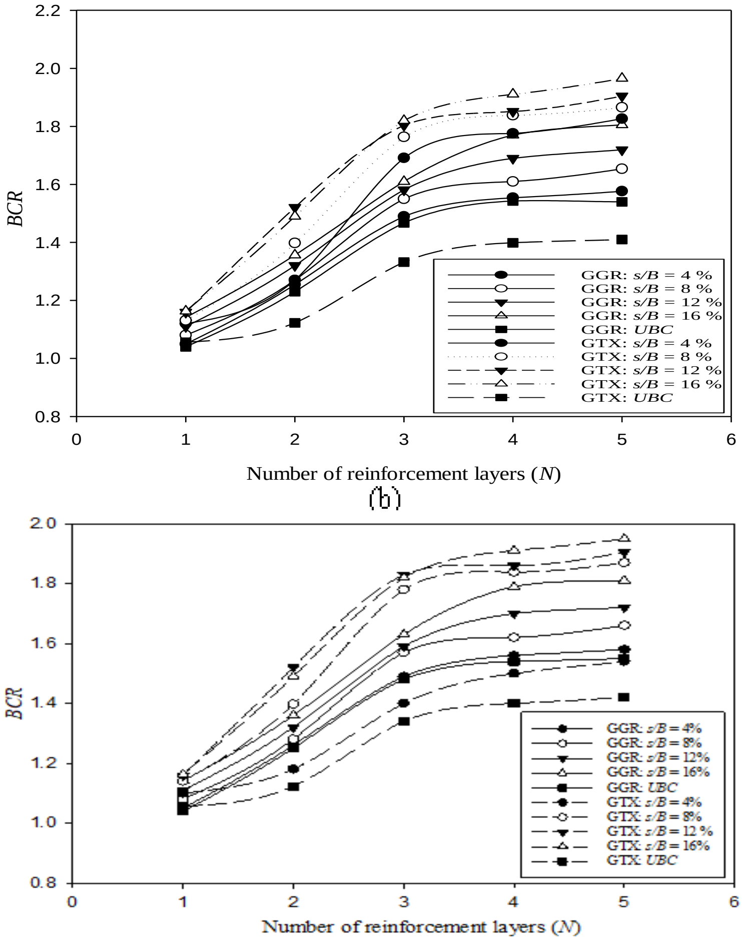
4.4 Influence of reinforcement width:

The effect of reinforcement width was analysed by the variation of reinforcement width as 4B, 5B, and 6B. When comparing the values of BCR from Fig.7 (a-c) for the same number of reinforcement layers at a different reinforcement width, it is clear that the footing performance in terms of bearing capacity improvement for both the reinforcement geogrid and geotextile are significantly improved with an increase in reinforcement width. This significant improvement continues at around five times of width of footing for both the reinforcements. Consider, for example for the geogrid case

with N= 4 at UBC. The BCR increased from 1.44 to 1.56 when b/B ratio varied from 4 to 5. Further increment in b/B ratio from 5 to 6, the BCR was increased from 1.56 to 1.58. It is observed that a maximum rise in BCR values was observed for b/B ratio between 4 to 5 than those between 5 to 6. It is clear that satisfactory results may not be expected with increment in reinforcement width beyond 6B. Similar to this finding, [5] reported that the optimum width of geogrid to reinforce the square footing resting over the sand observed at b/B = 5 - 5.93. They explained that the concept of optimum width of reinforcement comes from the fact that only those portion of reinforcement is mobilized the tensile strength effectively, which lies in the shear zone below the footing. Beyond the shear zone, some more length is required as anchorage to impart pull-out resistance to the reinforcement thus, the optimum width of reinforcement is sum of length of reinforcement in anchorage zone and shear zone on the both sides. Further increment in reinforcement width beyond the optimum value will not be effective and satisfactory results cannot be expected. [8] considers the similar findings, they suggested the optimum reinforcement width equal to five times of the footing when they used the different planer geosynthetics in reinforced soil foundation. The similar results can be observed in case of geotextile.



(a)



(b)

(c)

Fig. 7 Bearing capacity ratio (BCR) for (a) width of reinforcement = 4B (b) width of reinforcement = 5B (c) width of reinforcement = 6B

4.5 Effect of vertical spacing between reinforcements

Fig.8 shows the pressure settlement curves with variations of vertical spacing between two reinforcements. It can be seen that 0.16B is the optimum vertical spacing for all reinforcements. Considering, for example, GGR case at settlement ratio of s/B = 4%, the bearing capacity increases to 25.0% and then the improvement decreases to 24%, 23.0, 21.4 and 19.8 % (Bearing pressure = 445.3, 441.7, 438.2, 432.5 and 426.8kPa) for h/B ratios 0.08, 0.16, 0.0.24, 0.32 and 0.4 respectively. The optimum value of vertical spacing was obtained at 0.16B for both the geosynthetics under central loading. Similar results were suggested by [8]&[10].

(a)

Fig 8. Variation of Improvement Factor versus h/B for (a) GGR (b) GTX

5. Regression Analysis

Regression analysis was carried out on the results obtained from the experimental analysis. The analysis was carried out on a R Integrated Development Environment (IDE) RStudio. The analysis was carried out by considering five dimensionless parameters, viz. via. u/B, h/B, N, Normalized stiffness and Normalized tensile strength and estimating the degree of significance of each parameter with the improvement factor, computed at ultimate bearing capacity of soil.

For the purpose of analysis, data set was created using the results obtained from the experimental analysis, and imported into the RStudio framework. Bearing Capacity Ratio was kept as the Y intercept for the purpose of analysis and functioned as the dependent variable, and all the other parameters were analyzed for their influence on the improvement factor at the ultimate bearing capacity for all the reinforcements and their corresponding configurations. The relative importance of each independent parameter for computation of ultimate bearing capacity of a reinforced foundation was assessed by computing the individual t values for each of the variable. The higher the value of |t|, the greater is the variable significance. Table 4. shows the fittings obtained with different parameters.

From Table 4. it can be observed that linear model including all dimensionless parameters is the best fitted for computation. Conducting linear regression including all parameters yields the following results as reported in Table 5.

From the analysis, it can be observed that N is the most significant factor when computing the ultimate bearing capacity of reinforced foundation, with an overall t value of 23.911. Also, from the analysis of the obtained results, it can be concluded that spacing between reinforcements is more significant that the tensile modulus of the reinforcement for a range of

$$0.08 \leq h/B \leq 0.4$$

which is a critical observation, as the project cost is usually associated with the spacing between reinforcements. From Table 4. it can be observed that the highest adjusted R^2 value is obtained when u/B, N and Norm. Tensile Strength are taken as the parameters for computation of the ultimate bearing capacity, even though the multiple R^2 value reduces (confidence level still greater than 95%).

Table 4. Various possible linear models and their fittings

Multiple R^2	R^2 adjusted	Parameters
0.9685	0.9524	*u/B, h/B, N,* N_s and N_t
0.9657	0.9516	*u/B, h/B, N* and N_t
0.9644	0.9537	*u/B, N* and N_t
0.9599	0.9508	*u/B, h/B, N* and N_s
0.9547	0.9499	*u/B, N* and N_s
0.9511	0.9433	*h/B, N,* N_s and N_t
0.5197	0.4416	*u/B, h/B,* N_s, N_t
0.4785	0.4315	N
0.2996	0.2277	*h/B,* N_s and N_t
0.2768	0.2226	*u/B* and *h/B*
0.2323	0.2141	*h/B*
0.1884	0.1448	N_s and N_t
0.1871	0.1678	N_t
0.1796	0.1601	*u/B*
0.1117	0.1107	N_s

Table 5. Linear regression computations with all dependent variables

Parameters	Coefficients	t Value
u/B	*-0.1187*	*-3.009*
N	*0.1948*	*23.911*
h/B	*0.0159*	*2.175*
Ns	*0.01956*	*0.678*
Nt	*-0.4798*	*-1.51*

Conclusions

On the basis of obtained results following conclusions were drawn.

1. The optimum depth of top most layer was found to be 0.34B times the width of square footing for both the reinforcements i.e. geogrid and geotextile whereas optimum depth of reinforcement (d) was obtained at d/B ratio of 1.36B and 1.02B for geogrid and geotextile respectively.
2. The soil reinforced with geotextile behave differently from the geogrid. The improvement in bearing capacity increases with increasing in reinforcement layers. Optimum number of reinforcement layers was obtained at N=4 for geogrid reinforced soil

and N=3 for geotextile reinforced soil.

3. The width of reinforcement also played a crucial role in amassing maximum benefit from reinforcements. A substantial improvement in performance of reinforcement was found when the width of reinforcements was equal to 5 times the width of footing for geogrid and geotextile.
4. For the foundation construction point of view, Geogrid was the best performing material. Although, geotextile performed better at higher settlement ratios, geogrid provided better reinforcement for lower settlement ratios for which the structures are usually designed for.
5. Regression analysis showed that the most significant parameter for computation of the ultimate bearing capacity of the soil foundation is number of layers of reinforcements.
6. The vertical spacing between reinforcements is a more significant parameter than the Normalized Stiffness and the Normalized Tensile Strength of the geosynthetic.
7. The numerical model for computation of the ultimate bearing capacity includes number of layers of geosynthetics, Initial layer spacing and the normalized Tensile Strength.

References:

1. Binquet, J., and Lee, K. L.: Bearing capacity tests on reinforced earth slabs. Journal of Ge-otechnical and Geoenvironmental Engineering (101), (1975(b)).
2. El Sawwaf, M. A.: Behavior of strip footing on geogrid-reinforced sand over a soft clay slope. Geotextiles and Geomembranes 25(1), 50-60 (2007).
3. Rowe, R. K., and Soderman, K. L.: Stabilization of very soft soils using high strength geosynthetics: the role of finite element analyses. Geotextiles and Geomembranes 6(1-3), 53-80 (1987).
4. Mandal and Sah: Beating Capacity Tests on Geogrid-Reinforced Clay, Geotextiles and Geomembranes 327-333 (1992).
5. Latha G. M., and Somwanshi A.: Bearing capacity of square footings on geosynthetic reinforced sand (27), 281–294 (2009).
6. Thallak, S., G., Saride, S., and Dash, S., K.: Performance of surface footing on geocell-reinforced soft clay beds, GeotechGeolEng (25), 509–524 (2007).
7. Tafreshi, M. S. N, and Dawson, A.: Comparison of bearing capacity of a strip footing on sand with geocell and with planar forms of geotextile reinforcement (28), 72–84 (2010).

8. Chen, Q., Abu-Farsakh, M., Sharma, R., and Zhang, X.: Laboratory investigation of behavior of foundations on geosynthetic-reinforced clayey soil, Transportation Research Record: Journal of the Transportation Research Board 28-38. (2007).
9. Biswas, A., Krishna, A.M., and Das, S.K.: Behavior of Geosynthetic Reinforced Soil Foundation Systems Supported on Stiff Clay Subgrade, Int. J. Geomech 1532-3641 (2016).
10. Guido, V. A., Chang, D. K., & Sweeney, M. A.: Comparison of geogrid and geotextile reinforced earth slabs. Canadian Geotechnical Journal, 23(4), 435-440 (1986).

Kochi Chapter

Indian Geotechnical Conference
IGC 2022
15th – 17th December, 2022, Kochi

Stability Analysis of Rainfall-InducedLandslide Using Numerical Modelling

Akash Bhardwaj[1] and Amit Kumar Shrivastava[2]

[1] PG Scholar, Delhi Technological University, New Delhi, India
[2] Professor, Delhi Technological University, New Delhi, India
akash1491998@gmail.com

Abstract. Landslides triggered by rainfall are very frequent in India, especially in the Himalayanregion. despite various attempts, they are still occurring and causing heavy loss of life and civilization. As it is known that natural calamities are inexorable but the reduction in damage caused by them is possible., through preventive measures. For the prevention of landslides, we need to adopt various mitigation techniques but before that, an analysis of a slope’s stability is required to find a critical surface. In this study, a failed slope is considered in the Shimla districtof Himachal Pradesh, and its stability before and after rainfalls of different intensities (throughoutthe year) was studied using numerical modelling. The repeated slope failure necessitates a numerical approach to comprehend the instability components because there have been no previous stability investigations of this slope failure. According to the results, the failure of this slope was primarily caused by rain during the monsoon. Before rainfall, the slope's F.O.S. was more than 1, indicating that it was stable; but, following rainfall, it drops to 0.801,0.578, and 0.576, respectively. This study further in the future can be used in designing a landslide early warning system.

Keywords: Slope stability, rainfall-induced landslides, GeoStudio, Slope/W, Seep/W.

1 Introduction

The Himalayas is one of India's most prone areas to landslides, making it the ideal location to research all types of mass movements and slope failures that occur in nature. The Himalayas Mountain belt is geologically younger and comprises tectonically unstable geological formations. Himalayan province alone contributes to nearly 30% of the world’s total loss due to landslides (Li, 1990; Dahal et al, 2009). Rainfall-induced landslides are most common in the Himalayan region, especially in the monsoon season which results in heavy loss of life and property. The principle behind rainfall catalysing landslide is that when rainfall water infiltrates via pores present in the soil, it leads to the generation of positive pore water pressure which leads to decrement in effective stress thus resulting in a reduction of soil’s strength and ultimately leading to slope failure or landslide. Himachal Pradesh is one of the most landslide-prone states in India, mostly because of the deadly combination of (Unstable Himalayan formation and heavy rainfall). To prevent the loss due to rainfall-induced landslides, we have to study the stability criterion for which numerical techniques have been proven a reliable tool to study the stability of slopes and the effect of rainfall on them.

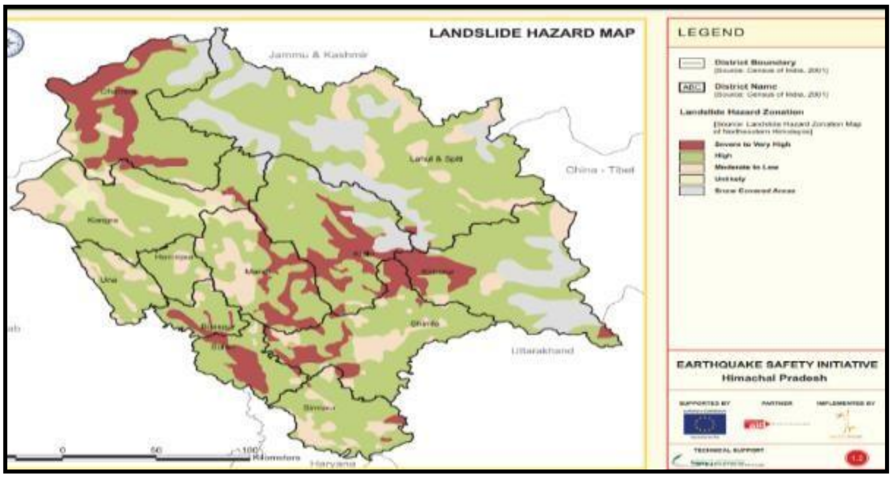


Fig. 1. Landslide Hazard Map of Himachal Pradesh (Source: HPSDMA)

[1] Kotropi soil is tested chemically and geotechnically as part of the preliminary investigation. Helical soil nails with a diameter of 20 mm and a length of 6 m are used to stabilise the failed slope in the presence of favourable prevailing soil conditions. Determining the factor of safety using the limit equilibrium approach, which is additionally checked by numerical modelling using a finite element subroutine PLAXIS 2D. [2] The inherent qualities of soil materials that affect the stability of the current slope have been identified through the geotechnical study. To measure the connection between precipitation and slope collapse, an event-specific antecedent rainfall threshold has been proposed. To show the situation of pre- and post-failure stability of the slope,a two-dimensional limit equilibrium method has also been used. [3]A tiny catchment in Niihama city on Shikoku Island in western Japan was chosen because it had a historyof seven slope failures brought on by severe rainfall brought on by a storm in October 2004. Following extensive fieldwork and a series of laboratory experiments to calculate hydro-mechanical parameters in saturated and unsaturated conditions, seepage and slope stability modelling of these slope failures were carried out in the GeoStudio environment using the precipitation data of 1920 October 2004. In silty sand, the pore water pressure was quickly changing, according to the seepage modelling results, and larger topographic hollows were shown to have greater maximum pore water pressures.[4] has carried out an examination of slope stability using the Mohr-Coulomb and Hoek-Brown failure criteria. The comparable traits for the slope stability analysis are identified in this work. It is established that employing an inadequate approximation of the confining stress causes major mistakes in the present conversion relationships.[5] The coupled model is established between internal erosion and unsaturated flow. It investigates how internal erosion affects slope stability and pore water pressure profiles. There is parametric research on hydraulic and erosion parameters. The findings of the numerical example demonstrate that internal erosion occurs mostly in the area inside the wetting front, which speeds up the wetting front's advance and reduces slope stability.[6] investigates rain-soaked soil-related landslides that occurred in Seoul, Korea. Used laboratory, field, and numerical methods to study landslides brought on by rainfall. The utilised approach is suited for simulating landslides, according to a

significant correlation between the numerical results and the analysed data.[7] Even in coarse-grained soils, pore pressure builds up because the movement does not allow for volume change, which results in the liquefaction of the soil mass. This results in a reduction in soil shear strength, making the slope unstable.[8] The main flaw of limit equilibrium approaches, which only satisfy statics equations, is that they fail to take strain and displacement compatibility into account. This has two detrimental effects. One is that it is unable to account for local differences in safety factors, and the second is that computed stress distributions are frequently erroneous.

As there have been very few studies conducted in this area, more research is required to demonstrate the viability and applicability of numerical modelling in this Himalayan Mountain belt region, which is particularly vulnerable to rainfall-induced landslides. To minimise the limitations of limit equilibrium analyses, the Morgenstern–Price method is used in this study which satisfies both force and moment equilibrium. The goal of this study is to determine whether numerical modelling can accurately simulate landslides in the Himalayan region under consideration, where rainfall is the primary cause of landslides. This study explains the role of rainfall in slope instability, which can be used to develop a rainfall-intensity duration model.

2 Study Location

The research area is close to Mishnu road near Bhajawa village of Shimla district of Himachal Pradesh, India (Fig.2). The place is shown in the satellite image retrieved below by Google Earth. The area is vulnerable to landslides brought on by rains.

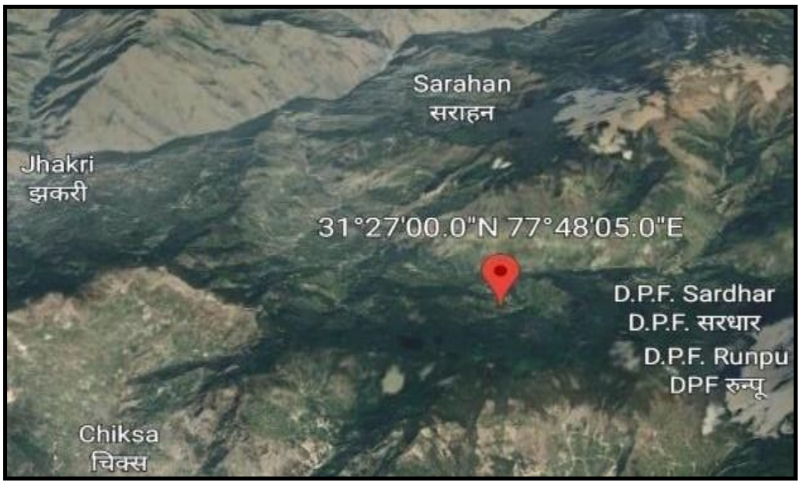


Fig. 2. Showing the location of the research area

2.1 Description of Study Area

To conduct the study of stability analysis a failed slope's data is taken from the Bhukosh Portal of the Geological Survey of India which is 30-33 m (on a 1/3 scale) in height with a slope angle of 43° (Fig.3). This slope has already failed during the monsoon season of the year 2016. Loose and heavily worn quartz mica schist and gneiss were found on the slope. The grain size distribution of slope matter is nonuniform and contains rock pieces embedded in the soil.

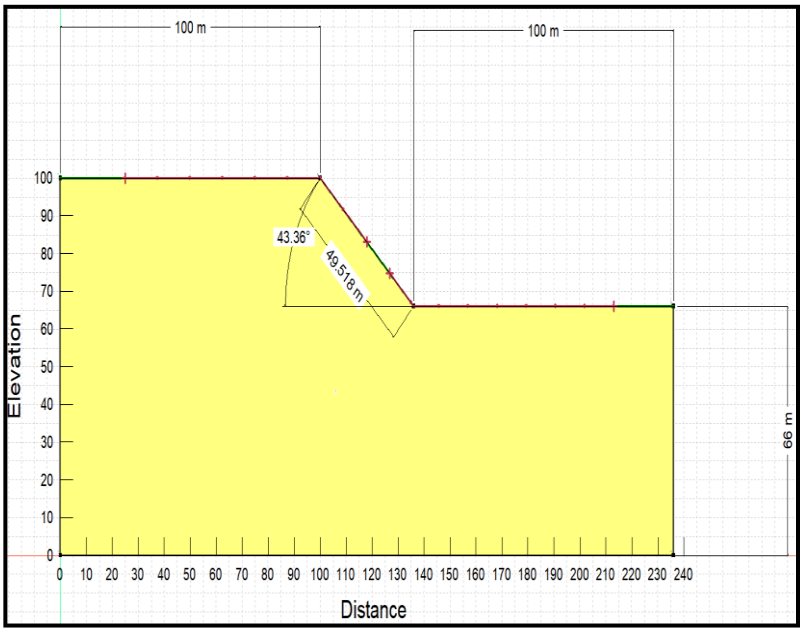


Fig. 3. Geometry of Study Slope

2.2 Material Properties and Tests

The landslide was separated into three equal portions along the landslide slope to collect samples from the site: the higher, middle, and lower sections. At various depths of 0.5 m, 1 m, and 1.5 m, soil samples are taken from each area using the core cutter method in open pits. Compaction, direct shear test as per IS 2720(Part 13):1986, and consolidometer tests are only a few examples of experiments that are carried out. Modelling in Finite Element analysis is based on the outcomes of these parameters. Results of these tests give the value of the Natural moisture content as 12%, Saturated unit weight of soil as 16 kn/m^3, Cohesion as 12 kn/m^2, and Angle of internal friction 32° under UU Condition of the triaxial test as per IS 2720(Part 12):1981, and Coefficient of permeability as 0.0026 m/hr.

2.3 Rainfall Characteristics

The southwest monsoon, which is a result of the orographic precipitation conditions, is primarily responsible for the rainfall in this research area. The southwest monsoon occurs from June through September, with the heaviest rainfall occurring in July. The Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) provided data on the study's monthly precipitation variation (Fig.4). For the simulation of slope conditions during the monsoon period, rainfalls of three different intensities are considered. The reason for considering 6.6 mm/day,33 mm/day, and 100 mm/day rainfall intensities is that 6.6 mm/day is the highest monthly rainfall intensity value for the monsoon period, and 100 mm/day is the maximum daily rainfall in July and 33mm/day is rainfall intensity over which there is no significant change in F.O.S of slope indicating complete failure.

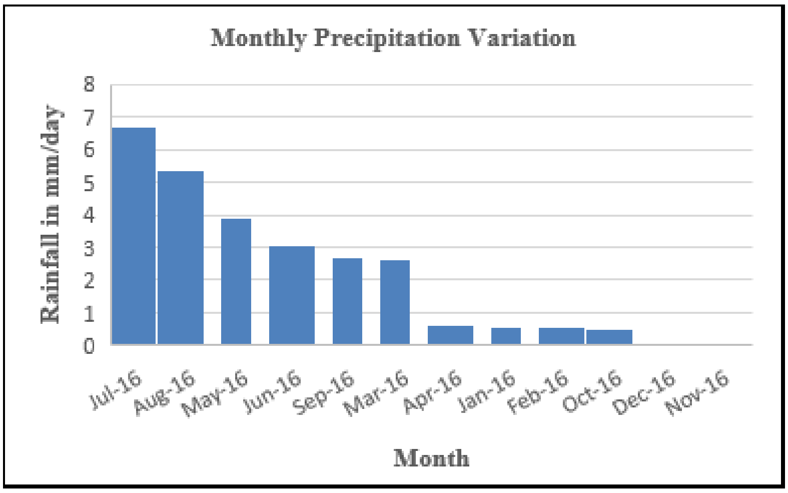


Fig. 4. Showing monthly precipitation variation of the year 2016 (Source: CHIRPS)

3 Methodology

3.1 Numerical Modelling

In this study to examine the soil slope, GeoStudio 2020 Software is used. Complete numerical modelling is performed in three phases. In the first step Slope/W tool which is a method based on the limit equilibrium approach was used to examine slope stability before the rainfall then using the finite element approach, Seep/W was used to model the rainfall, and the results were obtained from Seep/W again used in Slope/W to examine the stability of saturated slope following heavy rainfall.

3.2 Seepage Analysis During Rainfall

Based on the 2D finite element approach using SEEP/W, we can obtain Pore water pressure generated by rainfall of appropriate intensity concerning stated material property, slope geometry, and corresponding starting and boundary conditions. The mechanism behind its work is that it solves Darcy’s equation for a given slope condition by using a numerical discretization technique and executes water flow governing equations for the calculation of 2D seepage (Paswan & Shrivastava,2022).

$$\frac{\partial}{\partial x}\left(k_x \frac{\partial H}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial H}{\partial y}\right) + q = m_w^2 Y_w \frac{\partial H}{\partial t} \qquad (1)$$

3.3 Stability Analysis

The slope’s stability is examined using the GeoStudio software's Slope/W tool and is based on the limit equilibrium approach. Although there are other ways to calculate a slope's factor of safety, the Morgenstern-Price approach is what we'll be using in this research. This approach is used because it has the benefit of taking into account both force and moment equilibrium.

3.4 Geometry Modelling

The study used numerical analysis to create four geometry models. The first model is for slope stability study before rainfall, while the other three models are for investigation of slope stability following rainfalls with intensities of 6.6mm/day, 33mm/day, and 100mm/day, respectively. The reason behind choosing these intensities of rainfall is to consider all three possibilities of minimum, moderate and maximum rainfall. Two boundary conditional models are presented for demonstration.

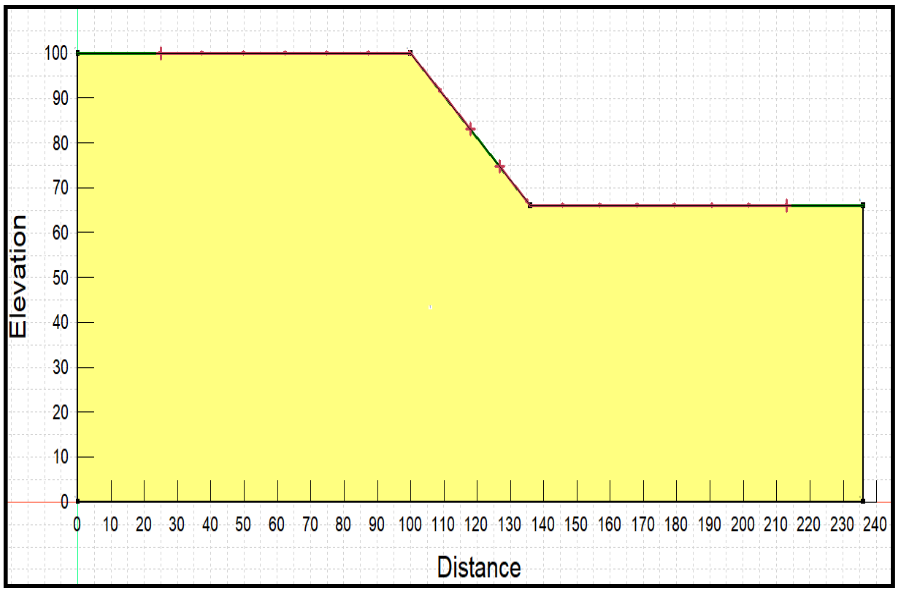


Fig. 5. Showing model of the unsaturated slope before rainfall

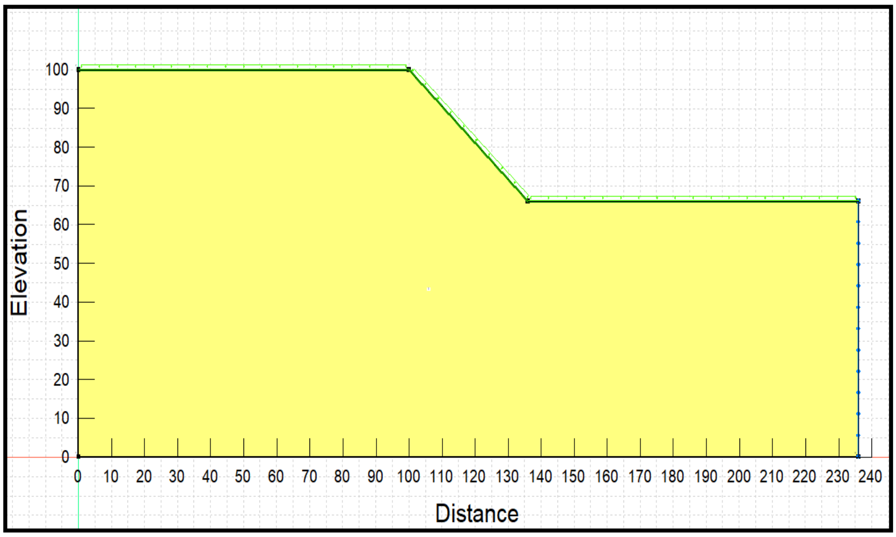


Fig.6. Showing model of the saturated slope after rainfall

4 Results and Discussion

4.1 Before Rainfall

This section deals with the outcomes of numerical modelling. To study the effect of rainfall on the stability of the slope, a factor of safety for the slope before rainfall is examined and it is found to be more than 1 (Fig.7). which justifies that the slope is stable when there is no rainfall. This analysis is also giving a critical slip surface on which factor of safety is minimum or we can say critical. Thus, information can be utilised for future mitigation purposes.

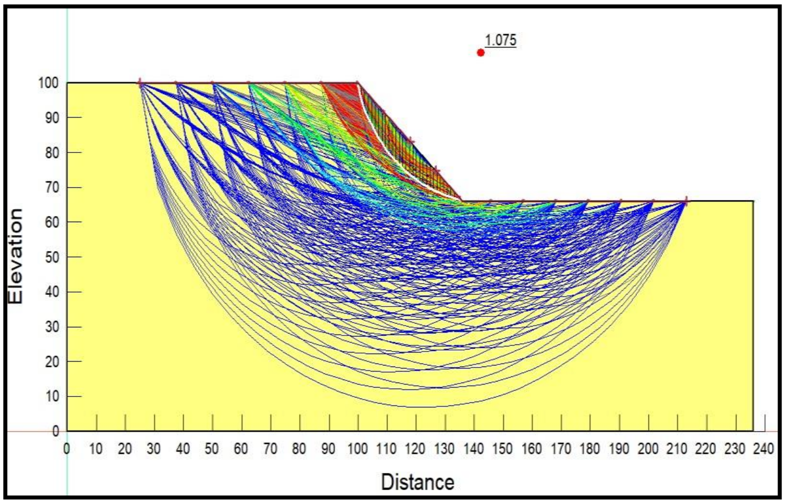


Fig. 7. Shows that the F.O.S of the slope is more than 1 before rainfall

4.2 After Rainfall

Now the analysis is carried out using different rainfall intensities and boundary conditions to observe the variation in factor of safety due to infiltration of the rainwater. is done in 2 parts, firstly seepage analysis is performed in the Seep/W tool, and then its results are used in Slope/W to find the stability of slopes. (Fig.8,9,10) Showing factor of safety for slopes with different rainfall intensities and it is very conclusive from them that F.O.S is decreasing with an increase in rainfall intensity. The factor of safety is 0.801 for rainfall intensity of 6.6mm/day and further decreasing to 0.578 for rainfall intensity of 33mm/day which is understandable as more infiltration of rainfall water causes the development of more pore water pressure resulting in a decrease in value of effective stress and ultimately leads to decrease in value of F.O.S but after rainfall intensity of 33mm/day, it is found that there is no significant change in values of F.O.S and possible reason of this is that slope is completely failed at rainfall intensity of 33mm/day and any further increase in rainfall intensity is not causing any difference in the value of F.O.S that's why F.O.S safety at 100mm/day rainfall intensity is also coming out to be 0.576.

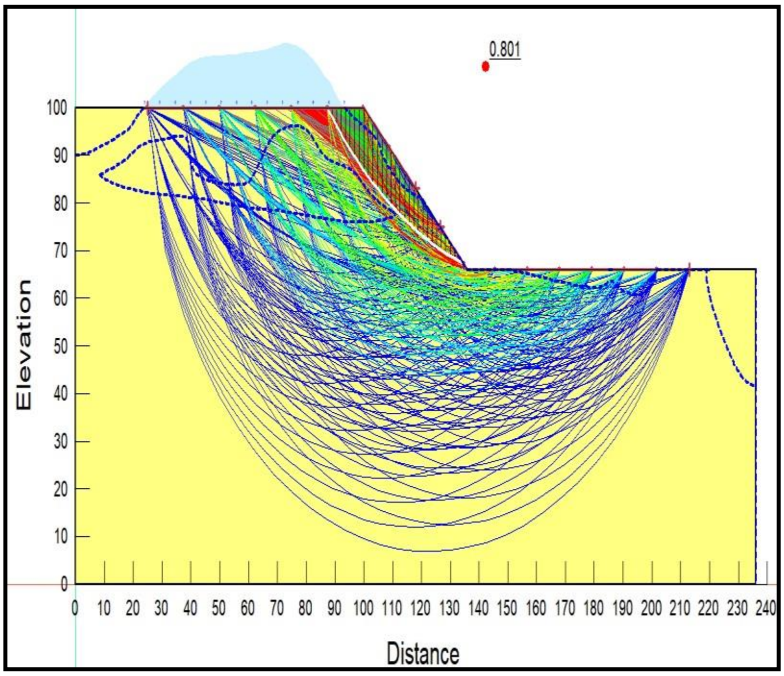


Fig. 8. Shows that the F.O.S of the slope is 0.801 for rainfall intensity of 6.6mm/day

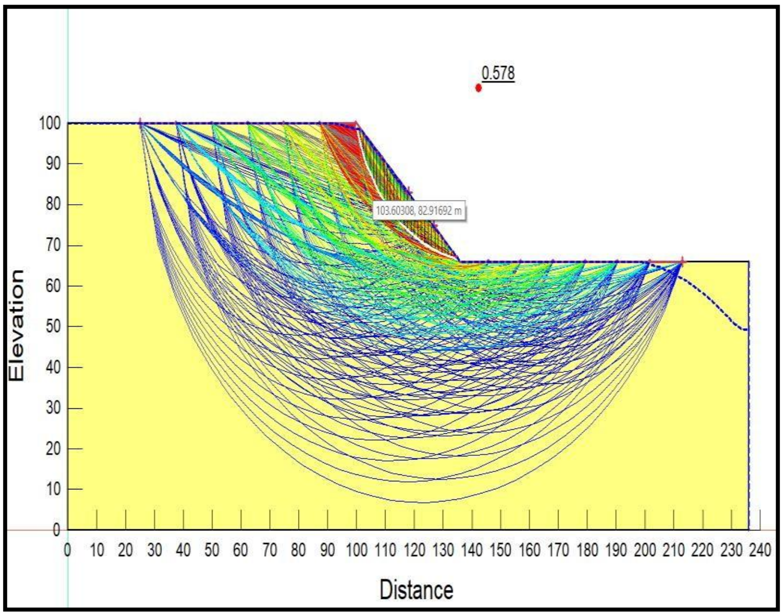


Fig. 9. Shows that the F.O.S of the slope is 0.578 for rainfall intensity of 33mm/day

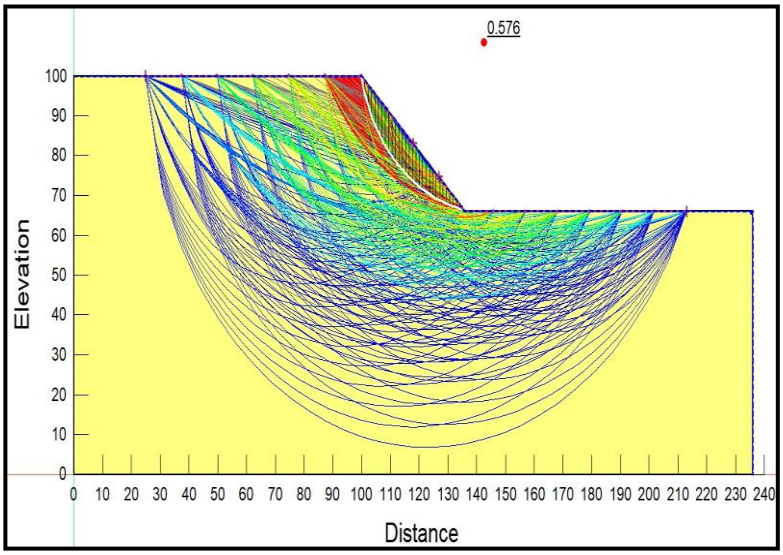


Fig. 10. Shows that the F.O.S of the slope is 0.576 for rainfall intensity of 100mm/day

Figure 11 is showing a plot of shear mobilisation and shear resistance before implementing rainfall intensity on the slope model. As we can see the value of shear resistance is more than shear mobilisation, which indicates that the value of forces that participate in slope failure is less than forces that are contributing to the stability of the slope, hence our slope is stable before rainfall.

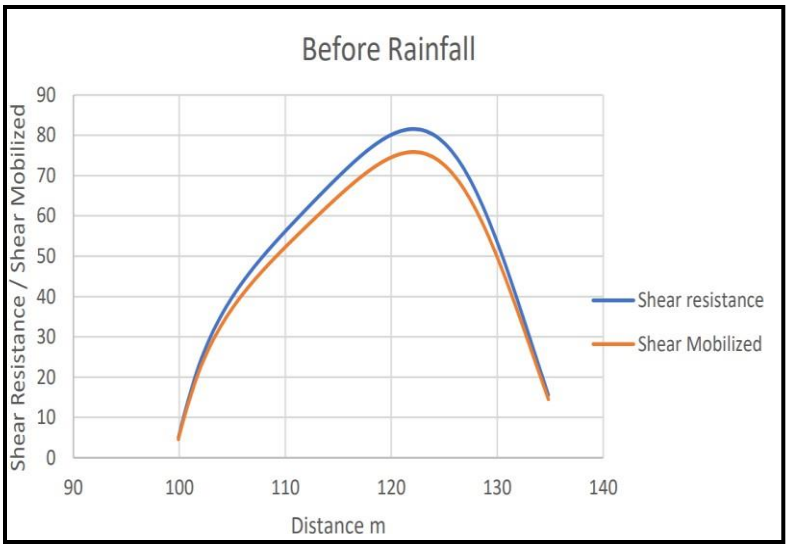


Fig. 11. Showing Shear Mobilised vs Shear Resistance comparison before rainfall

Figure 12 is showing a plot of shear mobilised and shear resistance after implementing rainfall intensity of 6.6mm/day on the slope model. As we can see, the value of shear mobilisation is more than shear resistance now, which indicates that the value of forces that participate in slope failure is more than forces that are contributing to the stability of the slope, ultimately failing the slope.

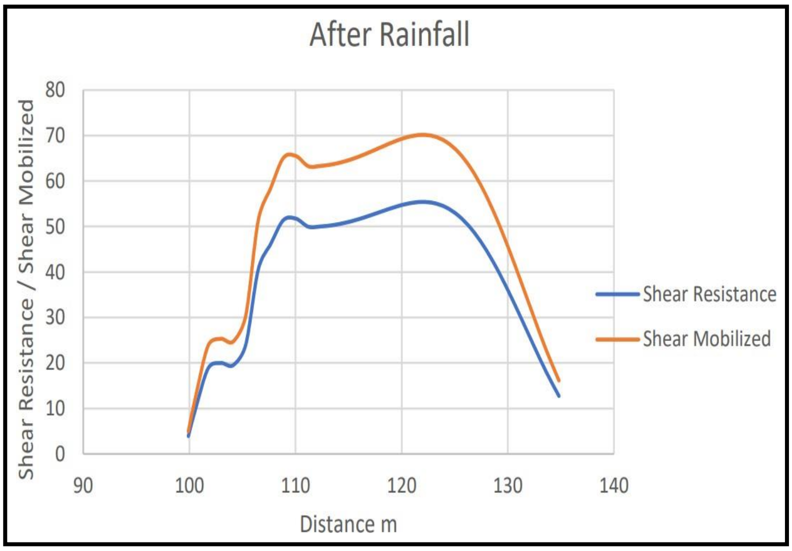


Fig. 12. Showing Shear Mobilised vs Shear Resistance comparison after rainfall

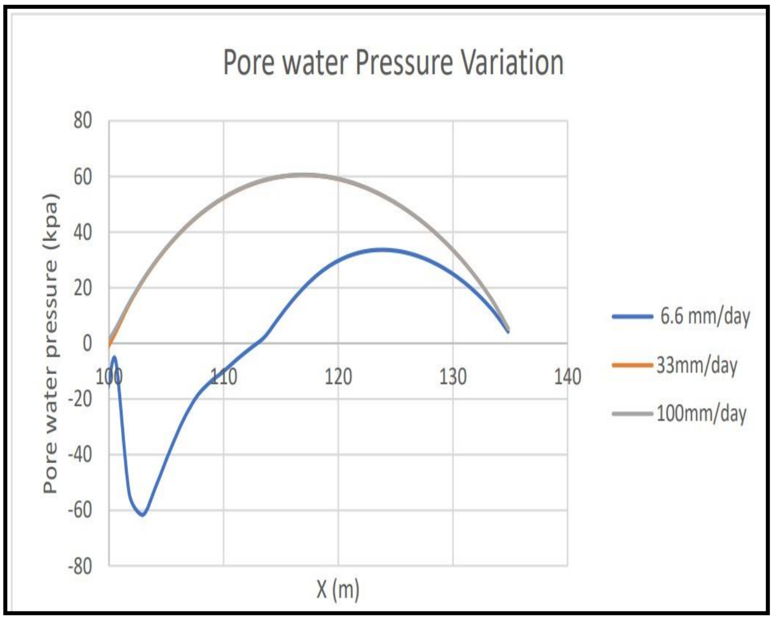


Fig. 13. Shows an increase in pore water pressure with an increase in rainfall intensity and hence causing slope instability

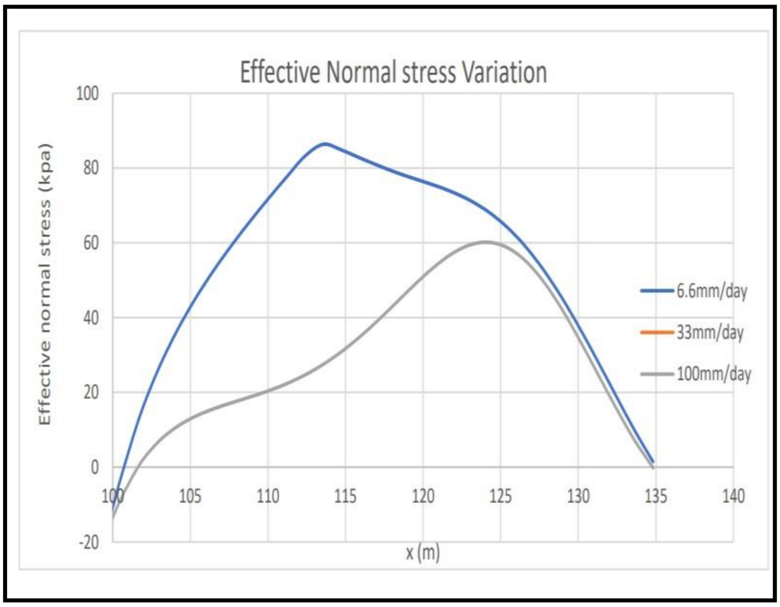


Fig. 14. Shows a decrease in effective normal stress with an increase in rainfall intensity and hence causing slope instability

5 Conclusions

In this study, an investigation is carried out to study the effect of rainfall intensity on a soil slope. The major findings of this study are:

(1) Factor of safety before rainfall or pre-monsoon was greater than 1, indicating a stable slope.
(2) Factor of safety for the slope after rainfall of intensity of 6.6mm/day was 0.801 which is less than 1 and hence indicates an unstable slope.
(3) Factor of safety for the slope after rainfall of intensity of 33mm/day was 0.578 which is very less than 1 and hence shows the critical failure condition.
(4) Factor of safety for the rainfall intensity of 100mm/day is 0.576, indicating that the slope has already completely failed before the such high intensity of rainfall, hence showing no significant reduction in the safety factor value.
(5) As this study confirms that rainfall is the main culprit behind landslides and slope instability in this area, this study can be used as a base to determine a rainfall intensity–duration threshold model which can be further used as a landslide early warning system for this location.

Future Scope and Limitation

There is a need for more studies for the validation and comparison of results obtained from numerical simulation with actual physical modelling results. This study only talks about the effect of rainfall intensities on slope stability but not about any remedial technique.

Acknowledgment

This study is funded by Delhi Technological University, Delhi for the MTech course. The authors are grateful to Research Scholar Mr. Abhishek Prakash Paswan who supported the author through this study.

References

1. Sharma, P., Rawat, S., and Gupta, A. K. (2018). "Study and remedy of Kotropi landslide in Himachal Pradesh, India." Indian Geotechnical Journal, 49(6), 603–619.
2. Singh, A. K., Kundu, J., and Sarkar, K. (2017). "Stability Analysis of a recurring soil slope failure along NH-5, Himachal Himalaya, India." Natural Hazards, 90(2), 863–885, doi:10.1007/s11069-017-3076-z.
3. Acharya, K. P., Bhandary, N. P., Dahal, R. K., and Yatabe, R. (2014). "Seepage and slope stability modelling of rainfall-induced slope failures in topographic hollows." Geomatics, Natural Hazards, and Risk, 7(2), 721–746.
4. Rafiei Renani, H., and Martin, C. D. (2019). "Slope stability analysis using equivalent Mohr-Coulomb and Hoek–Brown criteria." Rock Mechanics and Rock Engineering, 53(1), 13–21.
5. Zhang, L., Wu, F., Zhang, H., Zhang, L., and Zhang, J. (2017). "Influences of internal erosion on infiltration and slope stability." Bulletin of Engineering Geology and the Environment, 78(3), 1815–1827.
6. Jeong, S., Lee, K., Kim, J., and Kim, Y. (2017). "Analysis of rainfall-induced landslides on unsaturated soil slopes." Sustainability, 9(7), 1280.
7. Sassa K (1985) The mechanism of debris flows. In: Proceedings, 11th international conference on soil mechanics and foundation engineering, San Francisco, vol 1, pp 1173–1176
8. Krahn, J. (2003). "The 2001 R.M. Hardy Lecture: The Limits of Limit Equilibrium Analysis." *Canadian Geotechnical Journal*, 40(3), 643–660.
9. Singh, K., and Sharma, A. (2022). "Road Cut Slope Stability Analysis at KOTROPI landslide zone along NH-154 in Himachal Pradesh, India." Journal of the Geological Society of India, 98(3), 379–386.
10. Rahardjo, H., Ong, T. H., Rezaur, R. B., and Leong, E. C. (2007). "Factors controlling instability of homogeneous soil slopes under rainfall." Journal of Geotechnical and Geoenvironmental Engineering, 133(12), 1532–1543.
11. Sarkar, K.; Singh, A.K.; Niyogi, A.; Behera, P.K.; Verma, A. K.; Singh, T. N. (2016). The assessment of slope stability along NH-22 in Rampur-Jhakri Area, Himachal Pradesh. Journal of the Geological Society of India 88(3):387-393. (IF: 1.459/2020)
12. Panchal, S., and Shrivastava, A. K. (2022). "Landslide hazard assessment using analytic hierarchy process (AHP): A case study of national highway 5 in India." *Ain Shams Engineering Journal*, 13(3), 101626.
13. Panchal, S., and Shrivastava, A. K. (2020). "Application of analytic hierarchy process in landslide susceptibility mapping at regional scale in GIS Environment." *Journal of Statistics and Management Systems,* 23(2), 199–206.
14. Behera, P.K.; Sarkar K.; Singh, A.K.; Verma, A. K.; Singh, T. N. (2016). Dump slope stability analysis – A case study. Journal of the Geological Society of India 88(6):725- 735. (IF: IF: 1.459/2020)

15. Dahal RK, Hasegawa S, Nonomura A, Yamanaka M, Masuda T, Nishino K (2008a) Failure characteristics of rainfall-induced shallow landslides in granitic terrains of Shikoku Island of Japan, Env Geol, online first. doi:10.1007/s00254-008-1228-x
16. Dahal RK, Kafle KR (2003) Landslide triggering by torrential rainfall, understanding from the Matatirtha landslide, south-western outskirts of the Kathmandu valley. In: Proceedings of one day international seminar on Disaster mitigation in Nepal, Nepal Engineering College and Ehime University, pp 44–56.
17. Collins, Brian & Znidarcic, Dobroslav. (2004). Stability Analyses of Rainfall Induced Landslides. Journal of Geotechnical and Geoenvironmental Engineering - J GEOTECH GEOENVIRON ENG. 130. 10.1061/(ASCE)1090-0241(2004)130:4(362).
18. Paswan, Abhishek & Shrivastava, Amit. (2022). "Stability Analysis of Rainfall induced landslides".
19. Paswan, A. P., and Shrivastava, A. (2022). "Modelling of rainfall-induced landslide: A threshold-based approach." Arabian Journal of Geosciences, 15(8).
20. Giannecchini R (2006) Relationship between rainfall and shallow landslides in the southern Apuan Alps (Italy). Nat Hazards Earth Sys Sci 6:357–364
21. GeoStudio (2005) GeoStudio Tutorials include student edition lessons, 1st edn. Geo-Slope International Ltd., Calgary.
22. Gupta, R. P., and Joshi, B. C. (1990). "Landslide hazard zoning using the GIS approach—a case study from the Ramganga catchment, Himalayas." *Engineering Geology*, 28(1-2), 119–131.
23. Talukdar P, Bora R, Dey A (2018) Numerical investigation of hill slope instability due to seepage and anthropogenic activities. Indian Geotech J 48(3):585–594. https://doi.org/10.1007/s40098- 017-0272-4

J Mater Sci: Mater Electron (2023)34:694

Structural, thermal, and luminescence kinetics of $Sr_4Nb_2O_9$ phosphor doped with Dy^{3+} ions for cool w-LED applications

Ravina Lohan[1], A. Kumar[1], Mukesh K. Sahu[3], Anu Mor[2], V. Kumar[4], Nisha Deopa[1,*], and A. S. Rao[2]

[1] *Department of Physics, Chaudhary Ranbir Singh University, Rohtak Bypass Road, Jind, Haryana 126102, India*
[2] *Department of Applied Physics, Delhi Technological University, Bawana Road, New Delhi 110 042, India*
[3] *Department of Computer Science Engineering, Graphic Era (Deemed to be University), Haldwani Campus, Haldwani, Uttarakhand 263139, India*
[4] *Department of Electronics and Communication Engineering, Graphic Era (Deemed to be University), Dehradun, Uttarakhand 248002, India*

Received: 19 October 2022
Accepted: 3 February 2023

ABSTRACT

In this study, a series of white-light-emitting strontium niobium oxide {$Sr_4Nb_2O_9$: $x$$Dy^{3+}$ (x = 0.01, 0.03, 0.05, 0.07, and 0.10 mol)} phosphors were synthesized via solid-state reaction approach and analyzed by using XRD, SEM, diffuse reflectance, photoluminescence (PL), and temperature-dependent PL (TDPL) spectroscopy. The cubic structure of $Sr_4Nb_2O_9$ microparticles was identified by inspecting the diffraction pattern of the freshly generated phosphor. The formation of heterogeneous microstructures, including some aggregation, was seen in the SEM image of $Sr_4Nb_2O_9$ phosphor. Diffuse reflectance and PL were examined for varying dopant ions concentration to explore the optical luminescence characteristics of $Sr_4Nb_2O_9$ phosphor materials. Further, absorption spectra were obtained by using the diffuse reflectance spectra. The PL spectra of as-prepared phosphor exhibit two peaks at 483 nm and 580 nm, corresponding to the $^4F_{9/2} \rightarrow {}^6H_{15/2}$ and $^4F_{9/2} \rightarrow {}^6H_{13/2}$ transitions, respectively. By correlating absorption and PL spectra, Judd oflet parameters were evaluated. The intensity of PL spectra of as-prepared phosphors increases up to 7 mol% and beyond that, it decreases. It is also affirmed that the PL intensity is maximum for the 7 mol% Dy^{3+} ions-doped $Sr_4Nb_2O_9$ sample. The PL lifetime of the level $^4F_{9/2}$ was evaluated by exciting the Dy^{3+} ions at 350 nm. By applying the Inokuti-Hirayama model to decay curves, the energy transfer process was explored. The CIE chromaticity coordinates of all the as-prepared phosphors lie in the white zone of the chromaticity diagram. The PL intensity remains 71% at 200 °C that of at room temperature, indicating the phosphor's exceptional

Address correspondence to E-mail: nispectro999@gmail.com

https://doi.org/10.1007/s10854-023-10055-z

Published online: 07 March 2023

thermal stability. From the aforesaid findings, it is observed that $Sr_4Nb_2O_9$: Dy^{3+} phosphor is a promising choice for lighting and luminescent devices.

1 Introduction

Over several years, the crystalline luminescent materials incorporated with lanthanide or transition metal ions have been gained extensive consideration owing to their potential utility in diversified fields of optoelectronics applications. The crystalline luminescent material such as phosphor is one of the special classes of luminescent materials, consisting of an inert host lattice doped with activator ions in small concentrations and the host provides a suitable site for the accommodation of dopant [1]. Typically, some materials like oxides, oxynitrides, nitrides, and sulfides doped with a small amount of (RE) rare-earth (4f) or transition (3d) metal ions can be taken as host materials.

Researchers intensively investigated the luminescent materials doped with RE ions due to the vast area of applications like in optical temperature sensors, lighting, drug delivery, and solar cells [2, 3]. At the current time, lighting equipment is based on white-light-emitting diodes (w-LEDs) which have numerous benefits like reduced energy consumption, extended lifespan, and environmentally friendly nature over incandescent lamps, tungsten bulbs, and fluorescence. Replacing these traditional lamps/bulbs, w-LEDs developed a new generation of light sources that attract researchers to this field [4]. The commonly produced w-LED is fabricated using blue-light-emitting InGaN LED chip layered with YAG: Ce^{3+}, which transforms part of blue light into yellow light and results in a cool white light generation. Due to the lack of red components, problems such as poor color rendering index, halo effect, and thermal quenching occur [5]. Numerous research works have been investigated on UV/n-UV LED coated with multicolor emitting phosphors as these have high color rendering index (CRI) (> 90). However, these UV/n-UV LED excitable phosphors have low luminous efficiency because of different degradation schedules of red, green, and blue-emission. So, a prerequisite candidate for the manufacturing of w-LEDs with UV/n-UV chips is single-phase phosphor [6, 7]. The currently active research field is based on luminescent phosphor, to increase the efficiency and variety of uses of currently available SSL devices.

The commercially available phosphor (pumped with UV/n-UV light) based on sulfide and nitride has some drawbacks like low efficiency and less chemical stability, whereas the oxide-based phosphors are an excellent choice for numerous applications like SSL and display devices because of their extraordinary moisture resistance, chemical inertness, chemical stability, wide band gap, cost effectiveness, and eco-friendliness [8–11]. Currently, niobates are considered rising luminescent hosts because of exceptional properties, including diverse crystal structure, wide transparency range, non-linear property, good chemical stability, and high electro-optical and high mechanical performance [12], which results in extensive applications in microwave resonators, pyroelectric field, ferroelectric, photocatalytic, and photorefractive device field [13]. Alkaline earth niobates are categorized as an excellent material having high scientific and technological importance because of the exceptional photocatalytic, ionic conductive, non-linear optical, photorefractive, and piezoelectric properties used in the applications like delay lines in filters, acoustic transducers, beam deflector, and optical modulator [14, 15]. Owing to the above-mentioned scientific properties offered by alkaline earth niobates, it could be an excellent host for SSL technology.

The RE ions-induced phosphors have been prepared and utilized for many applications. In RE ions, the trivalent dysprosium (Dy^{3+}) ion has the capability of producing white light due to the presence of red, blue, and yellow emission bands owing to 4f–4f transitions. Due to the Stark splitting effect of the peaks, these Dy^{3+} transitions might be broad or divided into multiple peaks [16, 17]. The yellow band's strength is dependent on the host's crystal field environment, which helps generate white light by varying the Y/B ratio in Dy^{3+} ions-doped host [16]. Recently, work on Dy^{3+} ions-doped oxide-based phosphors has been reported by various groups. Recently, J.Y Si and his group investigated the luminescence properties of host-sensitized $MgNb_2O_6$-based phosphors. They observed three emission bands at 488, 583, and 668 nm for the transitions $^4F_{9/2} \rightarrow {}^6H_{15/2}$, $^4F_{9/2} \rightarrow {}^6H_{13/2}$, and $^4F_{9/2} \rightarrow {}^6H_{11/2}$, respectively, which lie in the blue, yellow, and red regions [18]. Ge Zhu et.al developed the novel niobate phosphors $SrBaNb_4O_{12}$:Re^{3+} (Re = Eu, Dy, Sm, and Pr). The electronic and luminescence properties show a strong connection with rare-earth ions concentration. The observed CIE coordinate for the $SrBaNb_4O_{12}$:Dy^{3+} is (0.418, 0.433) [19]. Mete Kaan Ekmekci and his group studied the Dy^{3+}-doped $CoNb_2O_6$ phosphor. The observed CIE coordinates and CCT values confirmed that the as-prepared phosphors are a potential candidate for the W-LEDs applications [20]. Bin Deng et.al studied the luminescent stability of Dy^{3+} ions-activated $KCa_2Nb_3O_{10}$ and the observation affirmed the potential of the phosphor for the W-LEDs utilization [21]. Moreover, all the recent trends in the field of white phosphor are summarized in Table 1. However, as far as we know, work based on PL characteristics of Dy^{3+} ions-doped $Sr_4Nb_2O_9$ phosphor has not been done. Therefore, this motivated us to prepare Dy^{3+} ions-doped $Sr_4Nb_2O_9$ phosphor to understand its photonic properties.

In the present work, strontium niobium oxide ($Sr_4Nb_2O_9$) phosphor doped with the various Dy^{3+} ions concentrations has been synthesized via solid-state reaction (SSR) and then characterized through different techniques such as X-ray diffraction (XRD), scanning electron microscopy (SEM), energy-dispersive spectroscopy (EDS), Fourier transform infrared spectroscopy (FT-IR), PL, time-resolved photoluminescence (TRPL), and temperature-dependent photoluminescence (TDPL). The optical band gap was evaluated using the diffuse reflectance (DRS) UV–vis spectrum and the Kubelka–Munk function. Further, J–O parameters have been evaluated by transforming diffuse reflectance spectra into absorption spectra. The efficient excitation and emission bands are obtained. The TRPL, CIE chromaticity coordinates, and thermal stability of Dy^{3+} ions-doped $Sr_4Nb_2O_9$ phosphor have been studied for the direct utility in n-UV/blue-pumped display and lighting devices.

2 Experimental

2.1 Materials and synthesis procedure

The samples $Sr_4Nb_2O_9$ (SNB) phosphors doped with Dy^{3+} ions have been manufactured through a solid-state reaction at a high temperature. The precursor materials $SrCO_3$ (98%), Nb_2O_5 (99.9%), and Dy_2O_3 (99.9%) were used, having high purity for the phosphor synthesis. The stoichiometric ratio according to $Sr_4Nb_2O_9$: xDy^{3+}(x = 0.0, 0.01, 0.03, 0.05, 0.07, and 0.10 mol) was taken by weighing the chemicals on an electric balance and mixed in an agate mortar for 2 h until the homogeneous mixture was obtained, and then transferred to an alumina crucible. The crucible was then placed in an electric furnace for sintering (at 1350 °C) for 7 h. This step was followed by a period during which the samples were allowed to cool to get an ambient temperature. Ultimately, the processed samples were crushed one more time before being characterized.

2.2 Characterization techniques

Thermogravimetric analysis (TGA) and differential thermal analysis (DTA) curves for mixture powder were recorded using instruments (Setaram, LABSYS evo). The structural analysis has been examined by XRD method through (Bruker, model-D8 Advance) diffractometer accoutered with filter (nickel) and radiation source (Cu K_α) having wavelength $\lambda = 1.5406$ Å within the range $20° \leq 2\theta \leq 80°$. The

Table 1 Summary of some current white phosphors

Sample	Key parameters	Value	References
$InNbTiO_6$:$0.06Dy^{3+}$	Thermal stability	50%	[17]
$MgNb_2O_6$:$0.015Dy^{3+}$	Lifetime	0.77 ms	[18]
$SrBaNb_4O_{12}$:$0.01Dy^{3+}$	CIE color coordinates	(0.418, 0.433)	[19]
$SrNb_2O_6$:$3Dy^{3+}$	Quantum efficiency	32.51%	[22]
Ca_3WO_6:$0.04Dy^{3+}$	Activation energy	0.105 eV	[23]

These current white phosphors encouraged us to synthesize the phosphor which should have high thermal stability(> 50%) [17], less lifetime (< 0.77 ms) [18], CIE close to the (0.33, 0.33), high quantum efficiency(> 32.51%) [22], and higher activation energy(> 0.105 eV) [23].

SEM and the EDS elemental analysis were examined with the help of EVO MA 10 VPSEM. FT-IR spectra of the as-prepared phosphors were recorded with the help of Nicolet IS50 FT-IR apparatus. The DRS has been carried out with a spectrophotometer (Jasco, V-770) having deuterium and halogen light sources. The PL excitation and PL emission spectra were recorded using a spectrofluorophotometer accoutered with a Xenon lamp (300 W) as an excitation source (Jasco, 8300FP). The Hitachi F 7000 spectrofluorometer along with a microsecond Xenon flash lamp (450 W) was used for the measurement of PL decay. The TDPL was recorded using the ocean optics spectrometer (FLAME-S-XR1-ES) and sample holder with heating assembly.

3 Results and discussion

3.1 Thermal properties

TGA and DTA curves of un-doped SNB phosphor, as displayed in Fig. 1, were utilized to investigate the thermal behavior of the prepared materials. There are three stages of weight loss in the TGA curve. The initial weight loss is obtained between 450 and 650 °C, and the corresponding weight loss is nearly 0.504%. Further, the peak in DTA supports the weight loss of the first stage, which causes evaporation and dehydration of surface water. In the second stage, weight loss of nearly 1.552% between 650 and 800 °C is ascribed because of the release of CO_2 gas. The third weight loss (∼ 9.028%) is between 800 and 1400 °C. The endothermic peak observed at 1084 °C could be because of the formation of $Sr_4Nb_2O_9$ phase. Beyond 1084 °C, no weight loss has been observed. The total weight loss is 11.084% in $Sr_4Nb_2O_9$ sample up to the calcination temperature 1400 °C from the TGA profile. Hence, the $Sr_4Nb_2O_9$ sample was sintered at different temperatures (> 1084 °C) to obtain the crystalline phase.

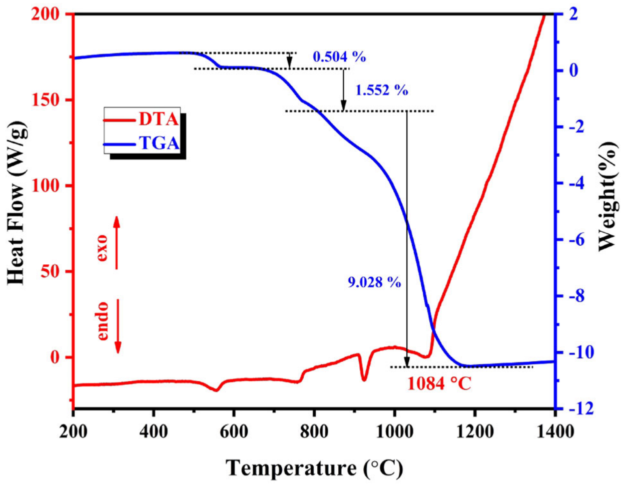


Fig. 1 TGA-DTA curve for un-doped SNB sample

3.2 Crystal structure and crystallite size analysis

The XRD patterns of un-doped and doped (7 mol%) SNB samples follow the standard data (Card No. 048-0558), as shown in Fig. 2. The XRD patterns affirm that the as-prepared sample is crystallized in cubic structure with Fm3m space group and the lattice parameters of $a = b = c = 8.207$ Å, $\alpha = \beta = \gamma =$ 90.00°, and V = 553.52 Å^3. The diffraction patterns for Dy^{3+} ions-doped SNB samples are also analogous to the standard data, which signifies that the Dy^{3+} ions are effectively absorbed in the SNB host without any structural changes. The ionic radius of Dy^{3+} ion is smaller than the ionic radius of Sr^{2+} ion. So, the Dy^{3+} can effectively replace the Sr^{2+} ion and occupy 4c/4d site. For this reason, the Dy^{3+} ions are doped in

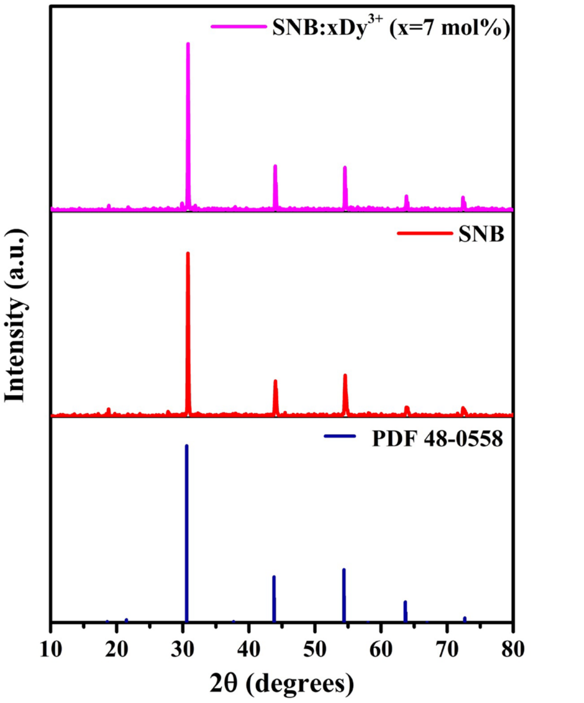


Fig. 2 XRD patterns of un-doped SNB and SNB: xDy^{3+} (x = 7 mol%) samples and compared with standard data

$Sr_4Nb_2O_9$ phosphor, and accordingly, the amount of Sr^{2+} precursor is reduced in the raw materials. The proximity in the ionic radii of Dy^{3+} and Sr^{2+} ions allows us to contemplate the substitution of Dy^{3+} ions at the sites of Sr^{2+} ions in $Sr_4Nb_2O_9$, host lattice. The XRD pattern recorded confirms the effective doping of the Dy^{3+} ions in the host lattice without any structural changes. Using the diffraction data, the average crystallite size (*D*) is evaluated using the Debye-Scherer equation available in the literature [24]. The values of full width at half maximum (FWHM) and Bragg's diffraction angle, derived from the XRD patterns to estimate the crystallite size, are represented by β and θ, respectively. The average crystallite size for Dy^{3+} (7 mol%) ions-doped SNB sample is 43.65 nm.

3.3 Morphological study

The SEM image of the as-prepared SNB sample is taken at different resolutions to explore the morphological characteristics and presented in Fig. 3a, b.

The SEM image displays the spherical shape and smooth surface of SNB particles that are appropriate for phosphor-converted w-LED. The particles show agglomeration due to the high-temperature sintering of phosphor with a size in micron (1–1.5 μm). It is reported that agglomeration is caused by smaller particles with a larger ratio of surface to volume than bulk particles [25]. As a result, smaller particles can have a lot of surface energy; thus, to reduce it, smaller particles clump together to form larger particles. The structure of the crystal, size, chemical composition, and the morphology of the luminescent materials are strong factors in concluding their properties and applications. The morphology affects the PL properties through geometric effects. The peculiar shapes of particles and the roughness of their surface bring about scattering which leads to less absorption and decreased the luminous efficiency.

3.4 Energy dispersive spectroscopy (EDS)

The elemental composition of un-doped SNB and SNB:7 mol% Dy^{3+} samples are analyzed using the EDS spectrum and elemental mapping. The EDS spectrum of un-doped SNB sample is exhibited in Fig. 4a. The EDS spectrum affirms the peaks of Sr, Nb, and O elements in the as-prepared SNB sample.

Also, the inset data of Fig. 4a show the element's weight% and atomic% in the SNB sample, which confirms the preparation of the SNB host. The elemental mappings are displayed in Fig. 4b, and the findings unveiled that the constituents Sr, Nb, and O are evenly distributed in the as-prepared SNB sample. Figure 5 demonstrates the SEM image, SEM–EDS mapping and EDS spectrum of the SNB:7 mol% Dy^{3+} sample to study its morphology and elemental composition. As depicted in Fig. 5a, the SNB:7 mol% Dy^{3+} sample exhibit spherical and smooth micron particles. Figure 5b–e shows the elemental mapping and results revealed that the elements Sr, Nb, O, and Dy are evenly distributed over the grains. The EDS spectrum is shown in Fig. 5f which revealed that the spectrum shows strong peaks of Sr, Nb, and O elements along with the Dy tiny peaks and without any impurity elemental peaks, in SNB: 7 mol% Dy^{3+} phosphor demonstrating that the Dy^{3+} ions are successfully incorporated in the host matrix. The compositional elements are displayed in the inset of Fig. 5f.

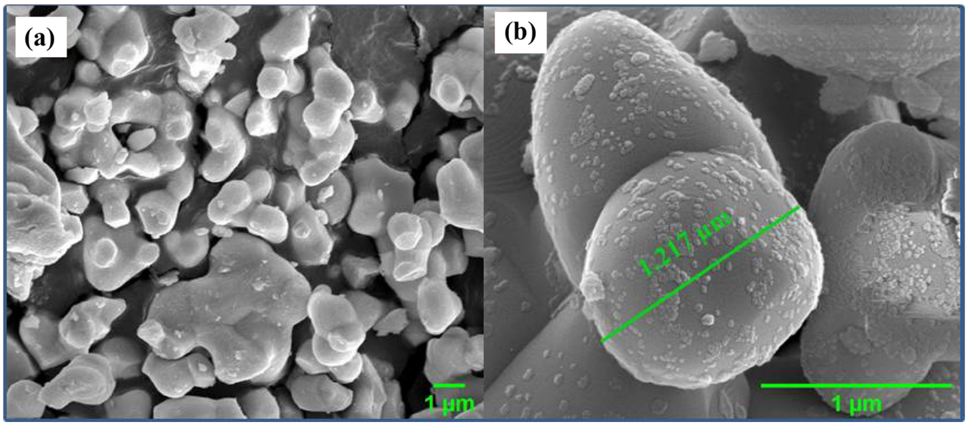


Fig. 3 **a**, **b** SEM images of un-doped SNB sample

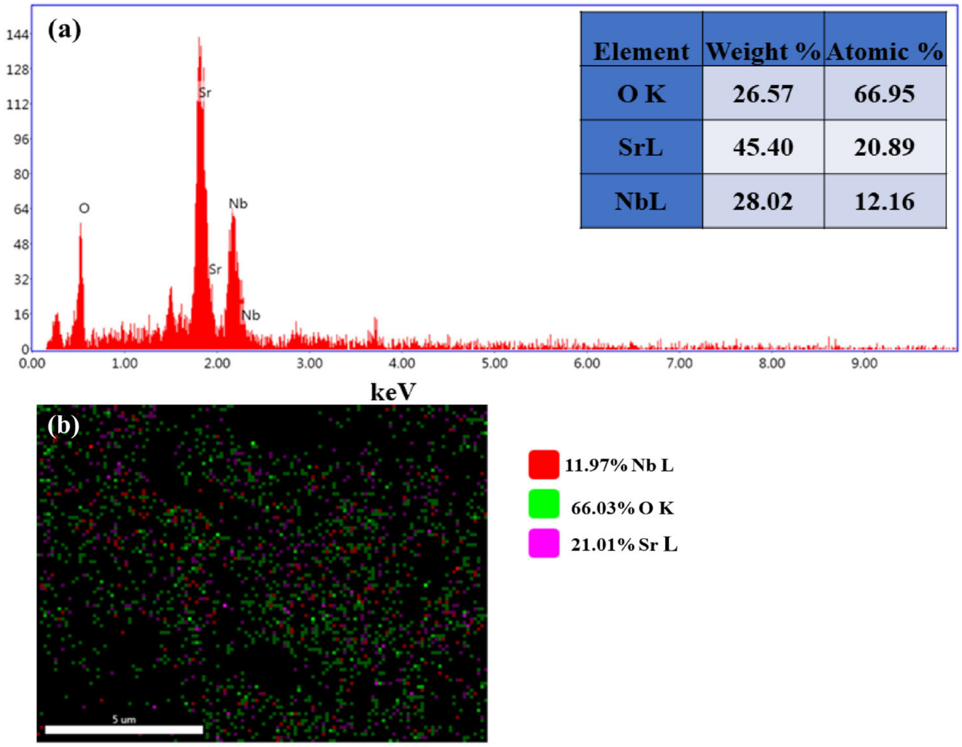


Fig. 4 **a** EDS spectrum of un-doped SNB sample [Inset shows the elements weight and atomic percent]. **b** SEM–EDS mapping of un-doped SNB sample

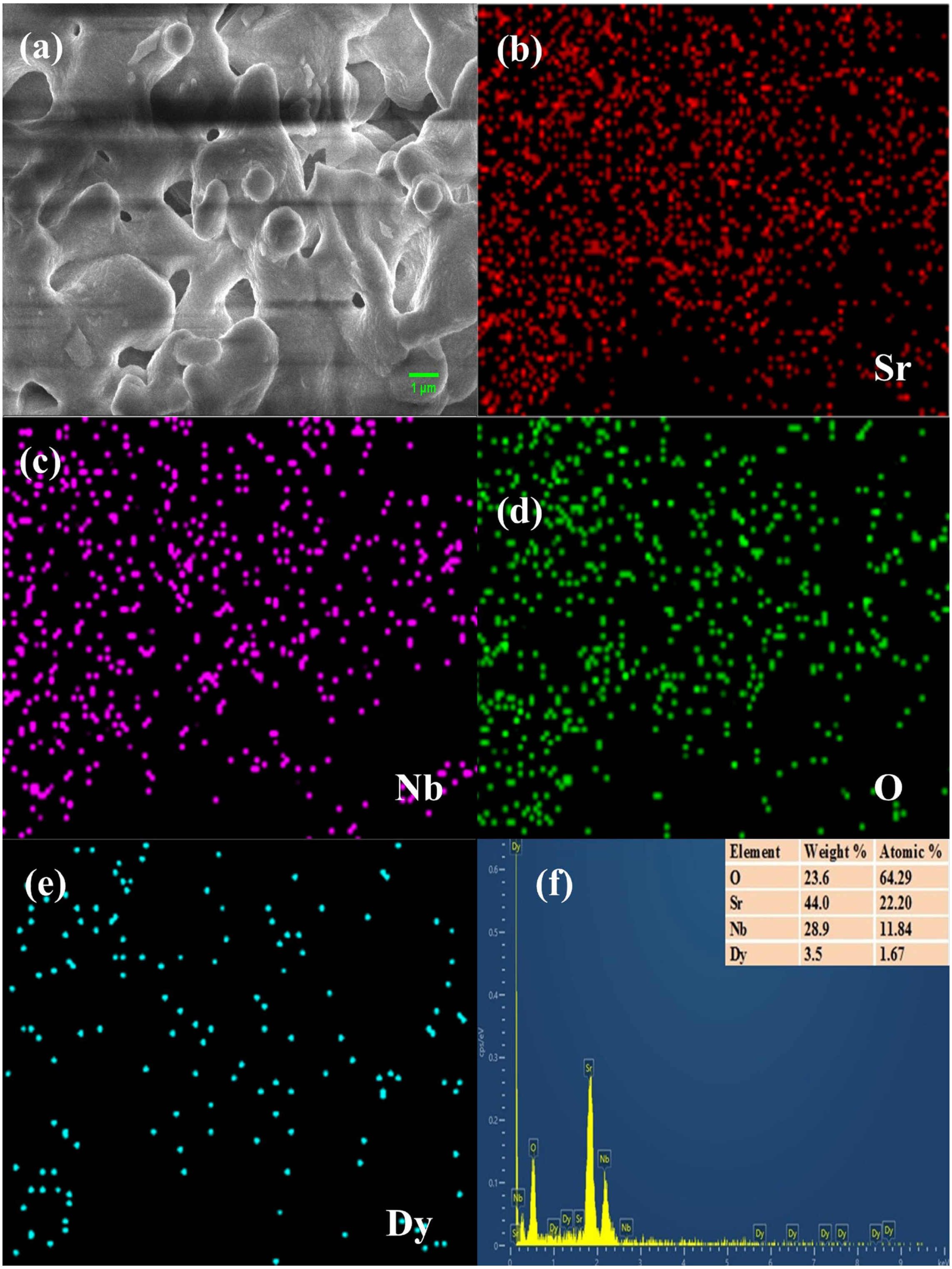
3.5 FT-IR spectral analysis

The FT-IR spectra are recorded from 500 to 2500 cm^{-1} to investigate the as-prepared sample's chemical bonding and molecular structure. Figure 6 demonstrates the FT-IR spectra of all the as-prepared Dy^{3+} ions-doped SNB samples that contain six shoulders at 599.7, 663.4, 863.9, 1447.8, 1980.5, and 2322.8 cm^{-1}, are tabulated in Table 2. The anti-symmetric stretching band at 599.7 cm^{-1} is eventually related to Sr–O vibrations [26]. The vibrations from 663.4 cm^{-1} strongly indicate the existence of NbO_6 octahedra in this work [27]. The energy band at 863.9 cm^{-1} is ascribed to the Nb–O stretching in the NbO_6 octahedron [28]. The intense band at 1447.8 cm^{-1} is ascribed to the absorption of water molecules because of the contact of the sample with the environment [29]. The asymmetric stretching of carbonates (CO_3^{2-}) is observed at 1980.5 cm^{-1} [25]. The band at 2322.8 cm^{-1} is assigned to the gaseous CO_2 [30]. The band position moves to a lower wavenumber with the increase in the Dy^{3+} ions concentration. This might be attributed to the variation of the environment around the NbO_6 octahedra. Therefore, FT-IR scattering is sensitive to local structural changes such as the coordination state and defects, and when foreign ions are incorporated into the crystal structure, the position of the bands is shifted.

3.6 Diffuse reflectance study and optical band-gap measurement

The DRS is measured in between 250 and 2500 nm for the SNB:$x$$Dy^{3+}$ (x = 1, 3, 5, 7, and 10 mol%) samples, presented in Fig. 7. The spectra consist of five inhomogeneous band edge peaks and contain the absorption edge at 283 nm, which can be set as the absorption wavelength for the SNB host. The peaks centered at 737, 793, 891, 1073, and 1266 nm corresponding to the absorption transitions of Dy^{3+} ions from the $^6H_{15/2}$ ground state to $^6F_{3/2}$, $^6F_{5/2}$, $^6F_{7/2}$, $^6F_{9/2}$ + $^6H_{7/2}$, $^6F_{11/2}$ + $^6H_{9/2}$ states, respectively [31]. The fewer absorption peaks at lower wavelengths result from the 4f–4f transition of Dy^{3+} ions [32]. The spectra indicate that the Dy^{3+} ions-doped SNB phosphor effectively absorbs in near-UV and NIR regions.

The absorption spectra are transformed from the DRS using the Kubelka–Munk function mentioned in the literature [33, 34]. The optical band gap (E_g) of SNB: $x$$Dy^{3+}$ (x = 1, 3, 5, 7, and 10 mol%) is determined with the help of the equation given in the reported literature [35, 36]. A graph between $h\nu$



Element	Weight %	Atomic %
O	23.6	64.29
Sr	44.0	22.20
Nb	28.9	11.84
Dy	3.5	1.67

Fig. 5 **a** SEM images. **b**–**e** SEM–EDS mapping. **f** EDS spectrum of SNB: 7 mol% Dy^{3+} phosphor [Inset shows the elements weight and atomic percent]

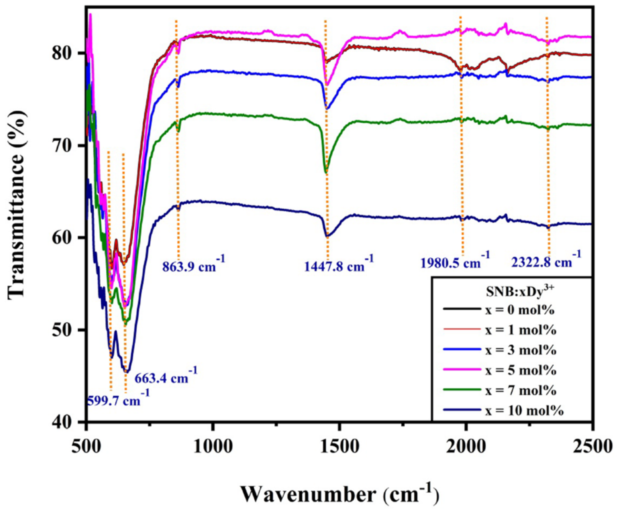


Fig. 6 FT-IR spectra of un-doped SNB and SNB: $x\mathrm{Dy}^{3+}$ samples

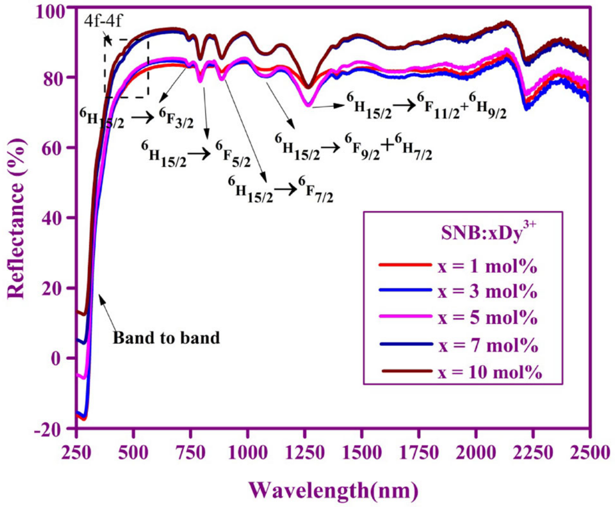


Fig. 7 Diffuse reflectance spectra of SNB:$x\mathrm{Dy}^{3+}$ phosphors

Table 2 Assignments of FT-IR peak positions of Dy^{3+} ions-doped SNB phosphors

Band position (cm^{-1})	Band Assignment	References
599.7	Sr–O vibration	[25]
663.4	Existence of NbO_6 octahedra	[26]
863.9	Nb–O stretching in NbO_6 octahedron	[27]
1447.8	Absorption of water molecules	[28]
1980.5	Asymmetric stretching of carbonate ($\mathrm{CO}_3{}^{2-}$)	[25]
2322.8	Gaseous CO_2	[29]

(incident energy) and $[F(R)\,h\nu]^2$ (Tauc Plot) is plotted for SNB:$x\mathrm{Dy}^{3+}$ samples, which are shown in Fig. 8. The direct band gap of SNB is calculated by extrapolating the slope equal to zero, and results reveal that with increasing Dy^{3+} ions concentration, optical band gap decreases, which comes out to be in the range 4.005–4.038 eV and mentioned in Table 3.

3.7 Bonding parameters

The DRS is taken and transformed to absorption spectra by using the Kubelka–Munk function and are used to obtain the nephelauxetic ratios (β) and bonding parameters (δ). The β and δ give information about the bonding nature of Dy^{3+} ions with the surroundings using the absorption spectra. The β values and bonding parameters are determined using the equation mentioned in the reported work [37, 38] and mentioned in Table 3.

The sign of the bonding parameter exhibits the nature of Dy^{3+} ions with the surroundings. The positive sign is assigned to covalent nature, while the negative sign is assigned to an ionic nature [39]. The positive values of δ for SNB: $x\mathrm{Dy}^{3+}$ (x = 1, 3, 5, 7 and 10 mol%) phosphors show the covalent nature of Dy^{3+} ions with the SNB matrix.

3.8 Judd-Ofelt analysis

Judd-Ofelt (J-O) theory is a basic technique for the evaluation of different radiative parameters of luminescent materials. The J-O theory in correlation with the absorption spectra is used to evaluate experimental (f_{exp}) as well as calculated (f_{cal}) oscillator strength values of SNB: $x\mathrm{Dy}^{3+}$ (x = 1, 3, 5, 7 and 10 mol%) phosphors [40]. The f_{exp} is calculated by the following equation [41]:

$$f_{exp} = \frac{2.303mc^2}{N_A \pi e^2}\int \varepsilon(v)dv = 4.32\times10^{-9}\int \varepsilon(v)dv \qquad (1)$$

where m is single electronic mass, c represents the speed of light, N_A represents Avogadro Number, $\varepsilon(v)$ represents the transition's molar absorbance at v (cm^{-1}) wavenumber and dv is the half of the band width of the absorption peak at v wavenumber.

The J-O theory describes the f_{cal} in transition from the ground state to the excited state (ψ to ψ′) as [42]:

$$f_{cal} = \frac{8\pi^2 mcv}{3h(2J+1)} \frac{(n^2+2)^2}{9n} \sum_{\lambda=2,4,6} \Omega\lambda \left| \left\langle \psi J \| U^{(\lambda)} \| \right\rangle \psi' J' \right|^2 \tag{2}$$

here h, m, J, n and v is the Planck's constant, the mass of an electron, the total angular momentum of the ground state, the index of refraction and the wave number of the transition respectively. Ω_λ ($\lambda = 2, 4, 6$) represents the intensity parameters of J-O, $\|U^{(\lambda)}\|$ is the doubly reduced matrix elements of the unit tensor operator with rank λ and are independent of the matrix (host) which are evaluated from absorption transition. The least square fit method is applied to find r.m.s deviation by using the equation given in the literature [43]. The smaller values of r.m.s. deviation between f_{cal} and f_{exp} validate the J-O theory. The oscillator strength is in direct relationship with the probability of absorption between the energy levels of the ground state and excited state. The smaller values of r.m.s deviation show the best fit between calculated (fcal) and experimental (fexp) oscillator strength. The oscillator strengths with r.m.s. deviation and intensity parameters of SNB: xDy^{3+} (x = 1, 3, 5, 7 and 10 mol%) phosphors are listed in Tables 4 and 5, respectively.

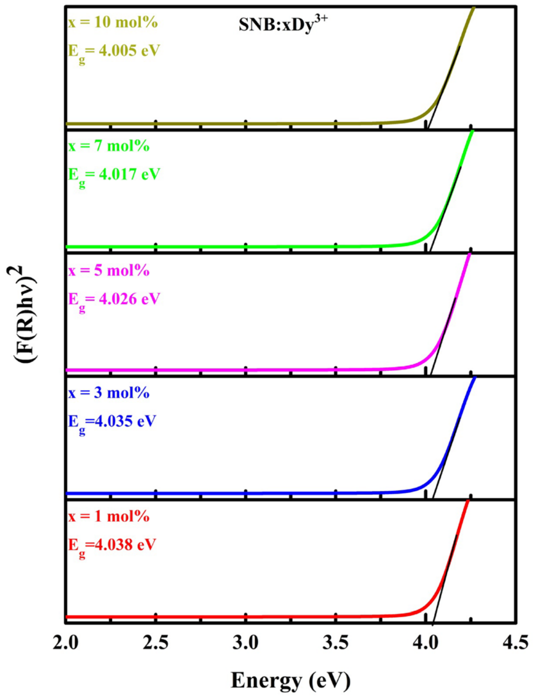


Fig. 8 Tauc plots of SNB:xDy^{3+} phosphors

$\Omega_2 > \Omega_6 > \Omega_4$ is the trend followed by the Ω_λ values of Dy^{3+}-doped SNB phosphor. These parameter values are of great importance because of their use to calculate the asymmetry in the environment of RE ions, the nature of bonding, and the bulk properties. The Ω_2 parameter exhibits covalent character dominance while structural dependency can be explained by the Ω_4 and Ω_6 parameters [44]. The highest value of Ω_2 among all the Ω_λ parameters reveals the bond between RE and ligand ions to be extremely covalent with asymmetric RE site [45]. This favors the outcomes drawn from the analysis of bonding parameters. The σ (root-mean-square value of oscillator strengths) can be depicted as the best fit between f_{exp} and f_{cal} and is calculated using the formula from the literature [46, 47].The values of f_{cal} and f_{exp} are in good agreement and confirmed by the small value of σ.

3.9 Photoluminescence properties

It is essential to identify the excitation wavelength for a better understanding of the emission spectra of the as-prepared SNB phosphors. The excitation spectrum of SNB:xDy^{3+} phosphors are recorded at the emission wavelength of 580 nm. The photoluminescence excitation (PLE) spectra are presented in Fig. 9, containing the eight peaks located at 324, 336, 350, 365, 389, 423, 454, and 472 nm corresponding to the transition of Dy^{3+} ions from the ground state ($^6H_{15/2}$) to $^6P_{3/2}$,

Table 3 Direct band gap (E_{gd}), Nephelauxetic (β) ratio, and Bonding (δ) parameter of Dy^{3+} ions-doped SNB phosphors

Parameters	SNB: 1 mol% Dy^{3+}	SNB: 3 mol% Dy^{3+}	SNB: 5 mol% Dy^{3+}	SNB: 7 mol% Dy^{3+}	SNB: 10 mol% Dy^{3+}
Direct band gap (E_{gd})	4.038	4.035	4.026	4.017	4.005
$\overline{\beta}$	0.9785	0.9812	0.9795	0.9796	0.9799
δ	0.0218	0.0191	0.0209	0.0208	0.0204

Table 4 Experimental (f_{exp} × 10^{-6}), calculated (f_{cal} × 10^{-6}) oscillator strengths and r.m.s deviation (δ_{rms} × 10^{-6}) of Dy^{3+} ions-doped SNB phosphors

Transitions from $^6H_{15/2}$ →	SNB:1 mol% Dy^{3+}		SNB:3 mol% Dy^{3+}		SNB:5 mol% Dy^{3+}		SNB:7 mol% Dy^{3+}		SNB:10 mol% Dy^{3+}	
	f_{exp}	f_{cal}	f_{exp}	f_{cal}	f_{exp}	f_{cal}	f_{exp}	f_{cal}	f_{exp}	f_{cal}
$^6F_{11/2}$ + $^6H_{9/2}$	0.018	0.018	0.037	0.037	0.545	0.545	0.492	0.492	0.061	0.061
$^6F_{9/2}$ + $^6H_{7/2}$	0.005	0.005	0.011	0.011	0.018	0.017	0.013	0.013	0.016	0.016
$^6H_{7/2}$	0.006	0.006	0.015	0.015	0.022	0.024	0.017	0.018	0.022	0.023
$^6H_{5/2}$	0.005	0.004	0.011	0.009	0.016	0.014	0.012	0.011	0.015	0.013
$^6H_{3/2}$	0.001	0.001	0.002	0.002	0.003	0.003	0.001	0.002	0.002	0.002
δ_{rms}(× 10^{-6})	± 0.001		± 0.001		± 0.001		± 0.001		± 0.001	

Table 5 Comparison of Judd–Ofelt intensity parameters Ω_2, Ω_4 and Ω_6 (10^{-22} m^2) for Dy^{3+} ions-doped SNB phosphors

Glass name	Ω_2	Ω_4	Ω_6	Order
SNB: 1 mol% Dy^{3+}	1.494	0.0452	0.545	$\Omega_2 > \Omega_6 > \Omega_4$
SNB: 3 mol% Dy^{3+}	3.268	0.2987	1.333	$\Omega_2 > \Omega_6 > \Omega_4$
SNB: 5 mol% Dy^{3+}	4.789	0.4347	2.044	$\Omega_2 > \Omega_6 > \Omega_4$
SNB: 7 mol% Dy^{3+}	4.380	0.4061	1.602	$\Omega_2 > \Omega_6 > \Omega_4$
SNB: 10 mol% Dy^{3+}	4.880	0.4754	1.802	$\Omega_2 > \Omega_6 > \Omega_4$

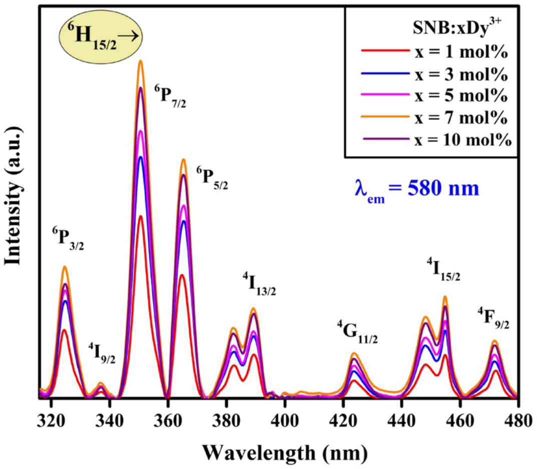


Fig. 9 PLE spectra of SNB:$x$$Dy^{3+}$ phosphors under 580 nm emission

$^4I_{9/2}$, $^6P_{7/2}$, $^6P_{5/2}$, $^4I_{13/2}$, $^4G_{11/2}$, $^4I_{15/2}$, and $^4F_{9/2}$, states, respectively. The highly intense peak at 350 nm that corresponds to the $^6H_{15/2}$ → $^6P_{7/2}$ transition is considered the excitation wavelength for recording the emission spectra and the peaks position is in good agreement with the published literature [31, 48, 49]. The excitation spectral intensity varies with the concentration of Dy^{3+} ions in SNB phosphor. The optimum excitation intensity is obtained for 7 mol% Dy^{3+} ions-doped SNB phosphor.

The emission spectrum of SNB:7 mol% Dy^{3+} phosphor under 324, 350, and 365 nm excitation wavelengths is depicted in Fig. 10. Each spectrum contains three bands at 483, 580, and 672 nm corresponding to the electronic transitions $^4F_{9/2}$ → $^6H_{15/2}$, $^4F_{9/2}$ → $^6H_{13/2}$, and $^4F_{9/2}$ → $^6H_{11/2}$, respectively, and among these, the band at 580 nm is highly intense. The excited wavelength 350 nm is considered as the optimized wavelength for the as-prepared phosphors as the spectrum at 350 nm has higher intensities for all the peaks in comparison to other excited wavelengths.

To determine Dy^{3+} ions concentration effect on the PL intensity, emission spectra are recorded for

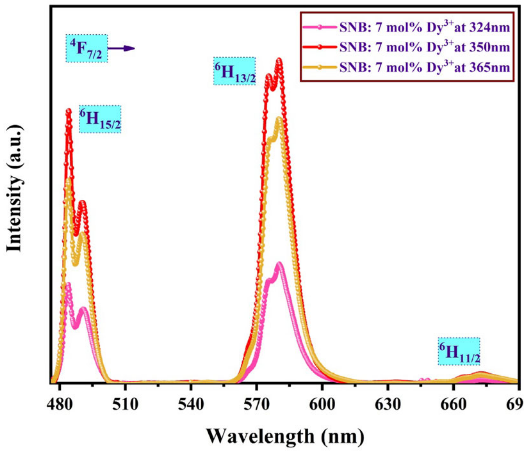


Fig. 10 Emission spectra of SNB:7 mol% Dy^{3+} phosphor under various pumping wavelengths (324, 350, 365 nm)

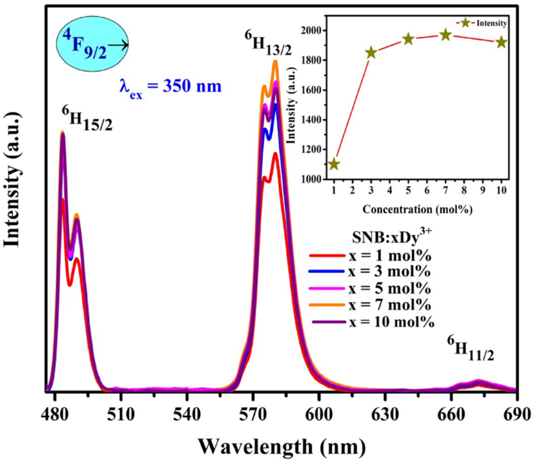


Fig. 11 Emission spectra monitored at 350 nm for SNB:xDy^{3+} phosphors [Inset shows the variation of intensity with the Dy^{3+} ions concentrations in SNB phosphors]

variable Dy^{3+} ions-doped SNB phosphors. Figure 11 illustrates the PL spectra of SNB:xDy^{3+} (x = 1, 3, 5, 7, and 10 mol%) phosphors at the excitation wavelength of 350 nm. The spectra contain two intense bands at 483 nm, 580 nm, and one weak band at 672 nm corresponding to the electronic transitions $^4F_{9/2} \rightarrow {}^6H_{15/2}$, $^4F_{9/2} \rightarrow {}^6H_{13/2}$, and $^4F_{9/2} \rightarrow {}^6H_{11/2}$, respectively. The intense yellow band at 580 nm is related to the low symmetry and hypersensitive transition of Dy^{3+} ions. On the other hand, the blue band at 483 nm is insensitive to the environment with the high symmetry of Dy^{3+} ions [50, 51]. Ionic radii of Dy^{3+} ions (0.912 Å) are lower than the Sr^{2+} ions (1.18 Å) surrounded with six atoms, allowing Dy^{3+} ions can efficiently replace Sr^{2+} sites with low symmetry.

Furthermore, it can be readily observable that $^4F_{9/2} \rightarrow {}^6H_{15/2}$ transition splits into a maximum of J + 1/2 (J is total angular momentum) stark manifolds in the blue-emission region due to the crystal field effect caused by surrounding ions [52]. Furthermore, the Y/B ratio, also known as the asymmetry ratio, is the relative measure of band intensities at 580 nm to 483 nm that is used to assess the effectiveness of Dy^{3+} ions-doped phosphors. The asymmetry ratio is found close to 1, indicating the appropriate white emission for as-prepared Dy^{3+} ions-doped phosphor [53]. The Y/B ratio for this investigation is in the range of 1.23–1.54, indicating that Dy^{3+} ions are placed at non-inversion symmetry positions in the host matrix, as shown in Table 6. With the increment of Dy^{3+} ions' concentration, the positions of the peak remain unaffected with an enhancement of emission peak intensity, and the maximum intensity is for Dy^{3+} ions (7 mol%)-doped SNB phosphor which is regarded as optimum concentration. The luminescence intensity diminishes as the concentration of Dy^{3+} ions rises further due to the phenomena of concentration quenching [54, 55]. The intensity fluctuations with doping concentration are depicted in the inset of Fig. 11 and confirm the concentration quenching for the transition ($^4F_{9/2} \rightarrow {}^6H_{13/2}$) at 7 mol% for the as-prepared Dy^{3+} ions-doped SNB phosphor.

Table 6 Colorimetric parameters along with Y/B ratio for SNB:xDy^{3+} phosphor with excitation of 350 nm

SNB: xDy^{3+}	Y/B ratio	Color coordinates		CCT(K)
		X	Y	
x = 1 mol%	1.35	0.358	0.380	4646
x = 3 mol%	1.39	0.368	0.387	4396
x = 5 mol%	1.25	0.374	0.392	4254
x = 7 mol%	1.54	0.375	0.393	4242
x = 10 mol%	1.23	0.365	0.383	4467

The energy level diagram, represented in Fig. 12, shows the emission, excitation, and cross-relaxation process of Dy^{3+} ions in the SNB host matrix.

The difference between the energy levels above $^4F_{9/2}$ is very trivial and results in non-radiative de-excitation; hence, $^4F_{9/2}$ level is populated. Owing to the large energy gap between $^4F_{9/2}$ and $^4F_{3/2}$, multi-phonon relaxation can be neutralized and $^4F_{9/2}$ de-excitation implies the radiative energy transfer via transitions $^4F_{9/2} \rightarrow {}^6H_{15/2}$, $^4F_{9/2} \rightarrow {}^6H_{13/2}$, and $^4F_{9/2} \rightarrow {}^6H_{11/2}$ [56]. The cross-relaxation mechanism is possible through the non-radiative decay mentioned below:

$$\mathrm{CR1}: {}^4F_{9/2} + {}^6H_{15/2} \rightarrow {}^6F_{3/2} + \left({}^6F_{9/2} + {}^6H_{7/2}\right),$$

$$\mathrm{CR2}: {}^4F_{9/2} + {}^6H_{15/2} \rightarrow {}^6F_{3/2} + \left({}^6F_{11/2} + {}^6H_{9/2}\right).$$

The distance between the nearest Dy^{3+} ions is reduced with the increment in dopant concentration, leading to a non-radiative transfer of energy between the dopant ions. Therefore, as the concentration of Dy^{3+} ions increases after an optimum concentration (7 mol%), the probability of non-radiative transfer increases and finally results in a decrement in luminescence efficiency. Hence, there is necessary to calculate the mean distance (R_c) between the

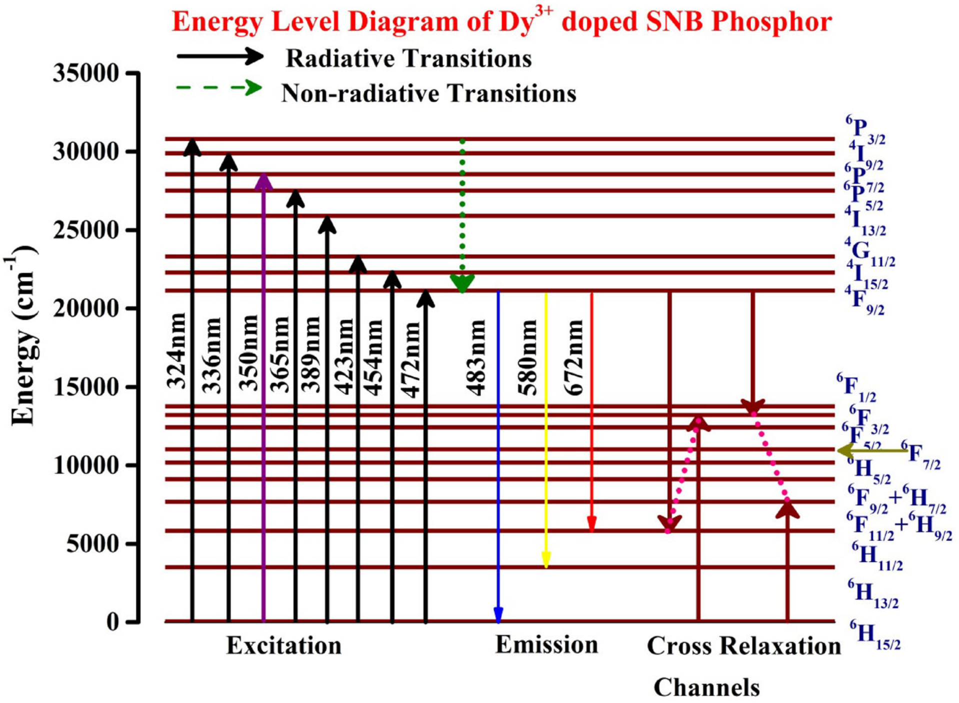


Fig. 12 Energy level diagram of Dy^{3+} ions in the SNB:xDy^{3+} phosphors

neighboring dopant ions at which non-radiative energy transfer takes place in SNB:xDy^{3+} (x = 1, 3, 5, 7, and 10 mol%) phosphor. Blasse proposed the relation for the mean distance calculation, which is given as follows [57]:

$$R_c = 2\left(\frac{3V}{4\pi N X_c}\right)^{\frac{1}{3}}. \tag{3}$$

The unit cell volume is V, the dopant ions optimum concentration is X_c, and the number of cationic sites in the host lattice is N. The critical distance of energy transfer in SNB:xDy^{3+} between the nearest Dy^{3+} ions according to the values V = 553.52 Å^3, N = 2, and X_c = 0.07 is 19.62 Å. The de-excitation of Dy^{3+} ions is broadly divided into radiative and non-radiative processes. Radiative transfer is not an effective quenching process for the majority of rare-earth ions while the non-radiative process includes internal relaxation and multipolar interactions between the ions. Exchange-type interaction takes place when the critical distance is less than 5 Å and multipolar interaction when it exceeds than 5 Å. Consequently, it proceeds through multipolar interaction [58]. Electric multipolar transitions are responsible for the concentration quenching mechanism as the critical distance is substantially bigger than the exchange interaction distance. Therefore, the mechanism for the concentration quenching may be due to electric multipolar interactions among the closest Dy^{3+} ions in SNB phosphor.

3.10 Radiative parameters

The radiative parameters of SNB:xDy^{3+} (x = 1, 3, 5, 7, and 10 mol%) phosphors can be evolved from the integration of calculated J–O parameters and the experimentally obtained results. J–O theory states that the transition probability (radiative) from the stable state (ground) to the excited state is the sum of the dipole transition (electric) probabilities and the dipole transition (magnetic) probabilities.

The A_R (radiative transition probability), τ_R, A_T (total radiative transition probability), and β_R for the SNB:xDy^{3+} (x = 1, 3, 5, 7, and 10 mol%) phosphors are evaluated using the literature [59, 60] and tabulated in Table 7. The A$_R$ evaluated with the help of J–O theory describe that the radiative transition from state $^4F_{9/2}$ to lower levels affects the luminescence performance of the as-prepared SNB phosphor. The value of the branching ratio and radiative transition probability is higher for the $^4F_{9/2}$ → $^6H_{11/2}$ transition. The radiative transition probability (A$_R$) will give the probability of a certain transition from the excited state to a lower state by radiative decay. Radiative probability is the inverse of a radiative lifetime. The luminescence efficiency is the ratio of the observed

Table 7 Transition probability (A_R) (s^{-1}), luminescence branching ratio (β_R), and total transition probability (A_T) (s^{-1}) and radiative lifetime (τ_R) (ms) for the observed emission transitions of Dy^{3+} ions-doped SNB phosphors

Transition	A_R	β_R	A_T	τ_R
SNB: 1 mol% Dy^{3+}				
$^4F_{9/2} \rightarrow {}^6H_{15/2}$	1.23	0.0057	214.62	4659
$^4F_{9/2} \rightarrow {}^6H_{13/2}$	1.23	0.0215		
$^4F_{9/2} \rightarrow {}^6H_{11/2}$	4.60	0.1405		
SNB: 3 mol% Dy^{3+}				
$^4F_{9/2} \rightarrow {}^6H_{15/2}$	3.09	0.0138	223.88	4466
$^4F_{9/2} \rightarrow {}^6H_{13/2}$	10.58	0.0473		
$^4F_{9/2} \rightarrow {}^6H_{11/2}$	30.75	0.1374		
SNB: 5 mol% Dy^{3+}				
$^4F_{9/2} \rightarrow {}^6H_{15/2}$	4.74	0.0204	231.95	4311
$^4F_{9/2} \rightarrow {}^6H_{13/2}$	15.75	0.0679		
$^4F_{9/2} \rightarrow {}^6H_{11/2}$	31.27	0.1348		
SNB: 7 mol% Dy^{3+}				
$^4F_{9/2} \rightarrow {}^6H_{15/2}$	3.75	0.0164	229.26	4361
$^4F_{9/2} \rightarrow {}^6H_{13/2}$	13.81	0.0602		
$^4F_{9/2} \rightarrow {}^6H_{11/2}$	31.21	0.1361		
SNB: 10 mol% Dy^{3+}				
$^4F_{9/2} \rightarrow {}^6H_{15/2}$	4.77	0.0203	235.18	4251
$^4F_{9/2} \rightarrow {}^6H_{13/2}$	17.44	0.0741		
$^4F_{9/2} \rightarrow {}^6H_{11/2}$	31.63	0.1345		

lifetime to the radiative lifetime, and the higher value of this ratio corresponds to higher efficiency. The branching ratios indicate the occurrence of stimulated emission from the excited states. The larger value of the branching ratio reveals that the as-prepared phosphor is excellent for luminescent applications [61].

3.11 PL decay study

PL Decay curves for SNB:xDy^{3+} phosphor under λ_{em} = 580 nm and λ_{ex} = 350 nm at room temperature are depicted in Fig. 13.

The recorded decay curves are fitted best with the double exponential equation given as follows [62]:

$$I(t) = I_0 + A_1 \exp\left(-\frac{t}{\tau_1}\right) + A_2 \exp\left(-\frac{t}{\tau_2}\right), \tag{4}$$

where $I(t)$ is the PL intensity at time t (in second) and I_0 is the intensity at the initial time. A_1 and A_2 represent the fitting constants. τ_1 and τ_2 denote the 2 decay lifetimes for the double exponential equation. The average decay time (τ_{avg}) is calculated with the formula given in the literature [63]. The lifetime values decrease with an increase in dopant concentration, i.e., Dy^{3+} ions concentration in the range of 0.64–0.55 ms, shown in Fig. 13, due to the non-radiative migration of energy. The quantum efficiency (η_{QE}) is the ratio of the number of the emitted photon to the number of absorbed photons, and in this case, it is defined as the ratio of an observed lifetime (τ) and the τ_r of the $^4F_{9/2}$ energy level. The η_{QE} can be evaluated with the equation mentioned below as follows:

$$\eta_{QE} = \frac{\tau}{\tau_r}. \tag{5}$$

The lifetime values and the quantum efficiencies are presented in Table 8. The quantum efficiencies of the as-prepared phosphors lie in the range 65.21–71.12%. The η_{QE} for optimized Dy^{3+} (7 mol%) ions-doped SNB phosphor is found to be 71.12% and higher in comparison to other reported work [22].

The Inakutti-Hirayama (I-H) model has been applied to the decay curves for optimized Dy^{3+} (7 mol%) ions-doped SNB phosphor to find out the involved energy transfer process. The decay curve of SNB:xDy^{3+} (x = 7 mol%) has been fitted for s = 6, s = 8, and s = 10 using the equation given below [64], and the curve shows the best fit for s = 6 as shown in Fig. 14:

$$I_t = I_0 * \exp\left(-\frac{t}{t_0} - Q * \left(\frac{t}{t_0}\right)^{\frac{3}{s}}\right), \tag{6}$$

where, I_t addresses the luminescent intensity at time t, t_o represent the donor lifetime in the absence of the acceptor, and Q is the parameter of energy transfer. The "s" illustrates the type of interaction between the

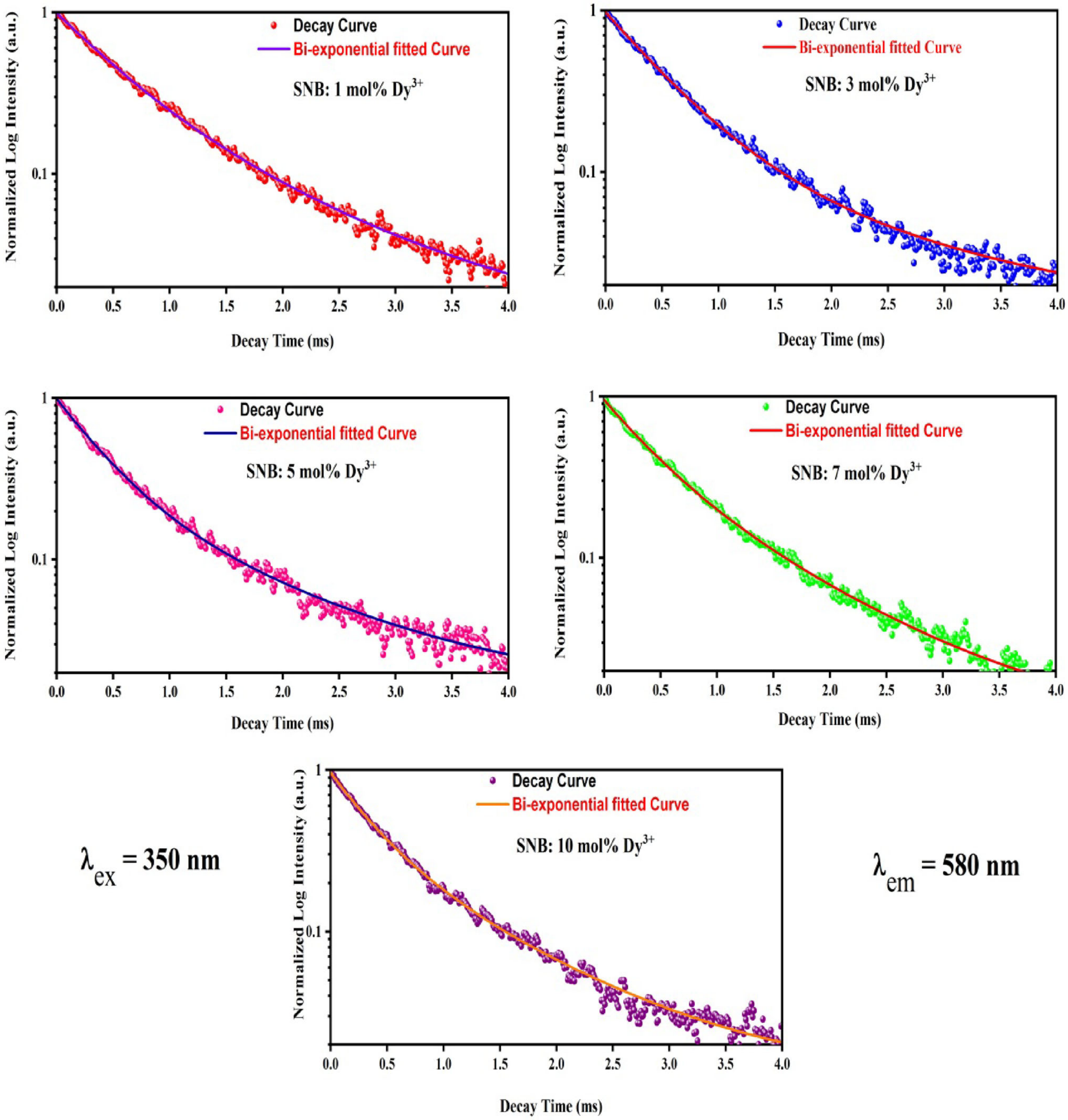


Fig. 13 PL decay profile for SNB: $x\mathrm{Dy}^{3+}$ phosphor at λ_{ex} = 350 nm and λ_{em} = 580 nm

activator ions in a particular host. For the present sample, the best I–H fit has been noticed for $s = 6$ implying that the d–d (dipole–dipole) interaction dominates among Dy^{3+} ions in SNB phosphor. The I–H fit corresponding to $s = 6$ for SNB:$x\mathrm{Dy}^{3+}$ phosphor is shown in Fig. 15. The energy transfer parameter (Q) was calculated using the formula available in the literature [65, 66], and the value of Q comes out to be 0.6309 for Dy^{3+} (7 mol%) ions-doped SNB phosphor.

3.12 CIE coordinates

For the confirmation of the emission color of SNB:$x\mathrm{Dy}^{3+}$ (x = 1, 3, 5, 7, and 10 mol%) phosphors, the chromaticity coordinates have been evaluated

Table 8 Photoluminescence lifetime and quantum efficiency of the $^4F_{9/2}$ states for the SNB phosphors

SNB: xDy^{3+}	Lifetime (ms)	Quantum efficiency
x = 1 mol%	0.6493	69.98%
x = 3 mol%	0.6263	70.21%
x = 5 mol%	0.6131	70.96%
x = 7 mol%	0.6129	71.12%
x = 10 mol%	0.5544	65.21%

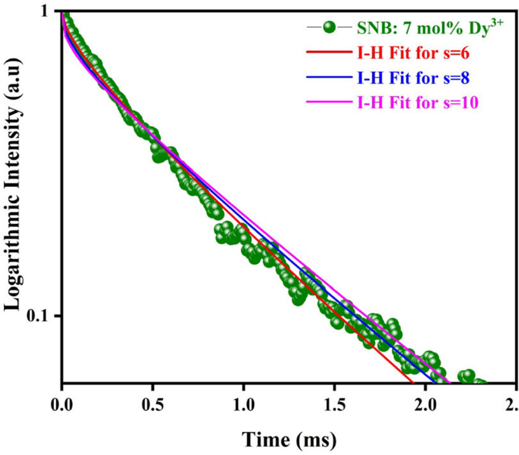


Fig. 14 I–H model fit of SNB:xDy^{3+} (x = 7 mol%) sample for s = 6, 8, 10

from the emission spectra under excitation of 350 nm wavelength. The CIE chromaticity coordinates are calculated to be (0.358, 0.380), (0.368, 0.387), (0.374, 0.392), (0.375, 0.393), and (0.365, 0.383), for SNB:xDy^{3+} (x = 1, 3, 5, 7, and 10 mol%) phosphors, respectively, and these are close to the standard white point (0.333, 0.333) as shown in Fig. 16. The CCT (Correlated color temperature) parameter help in evaluating the quality of light of a source and estimated by applying the McCamy's polynomial formula available in the literature [67].

The CCT values generally estimate the light's "warmness" and "coolness" and for less than 3200 K it is warm and for greater than 4200 K, it is cool. The calculated CCT values for various Dy^{3+} ions-doped SNB phosphors are 4646, 4396, 4254, 4242, and 4467 K and lie in the region of cool white luminescence.

The CIE chromaticity coordinates and the CCT values are listed in Table 6. The calculated CIE coordinates expose the application of Dy^{3+} ions-doped SNB phosphors in the lighting and display devices. The color purity is evaluated by the formula mentioned in the literature [68]. The color purity for Dy^{3+} (7 mol%) ions-doped SNB phosphor is evaluated to be 25.65% under the excitation wavelength of 350 nm, and it has been corroborated with previous reported work CaSrAl$_2$SiO$_7$:Dy^{3+} (37.72%), Sr$_9$Al$_6$O$_{18}$:Dy^{3+} (32.23%), and Sr$_9$Al$_6$O$_{18}$:Dy^{3+} (79.92%) [69–71].

3.13 Temperature-dependent PL properties

In general, UV/Blue LED chips are coated by phosphors, which generate heat around 120–150 °C during the operation, and as a consequence, the efficiency of lighting devices drops. Hence, thermal quenching is an essential characteristic of lighting devices. PL spectra collected at various temperatures are used to investigate the thermal quenching performance of Dy^{3+} (7 mol%) ions-doped SNB phosphor. Figure 17 depicts the TDPL spectra at the excitation wavelength of 350 nm for Dy^{3+} (7 mol%) ions-doped SNB phosphor.

The spectra show that as the temperature rises (30 to 200 °C), the emission strength decreases because of the non-radiative transition increment, and hence, the rate of intensity reduction alters over the emission peaks. With the increment in temperature, the PL intensity decreases because of thermal quenching although PL peaks position remains unaltered, prompting the good color stability of the phosphors [72]. The thermal characteristics of SNB:xDy^{3+} phosphors are concerned with Nb^{5+} energy states. There are two quenching mechanisms for the phosphor, one responsible for the displacement among the ground and excited state, the second one is associated with the thermal stimulated ionization process from the 5d level to the conduction band (CB) [73]. The light of a particular wavelength excites the electron to the CB and thereafter is captured by the trap, subsequently liberating the electrons to CB and Dy^{3+} ions recapturing them. Hence, the thermal stability is related to the ionization process, which can be influenced by band structure. With the increase in temperature, the electrons are easily excited to the CB and resulting in the decrement of the band gap. Hence, more energy from the excited state is lost and weakens the thermal quenching. The phosphors PL intensity decreases with the increase in temperature up to 200 °C known as temperature-dependent luminescence quenching and attributed to enhanced electron–phonon

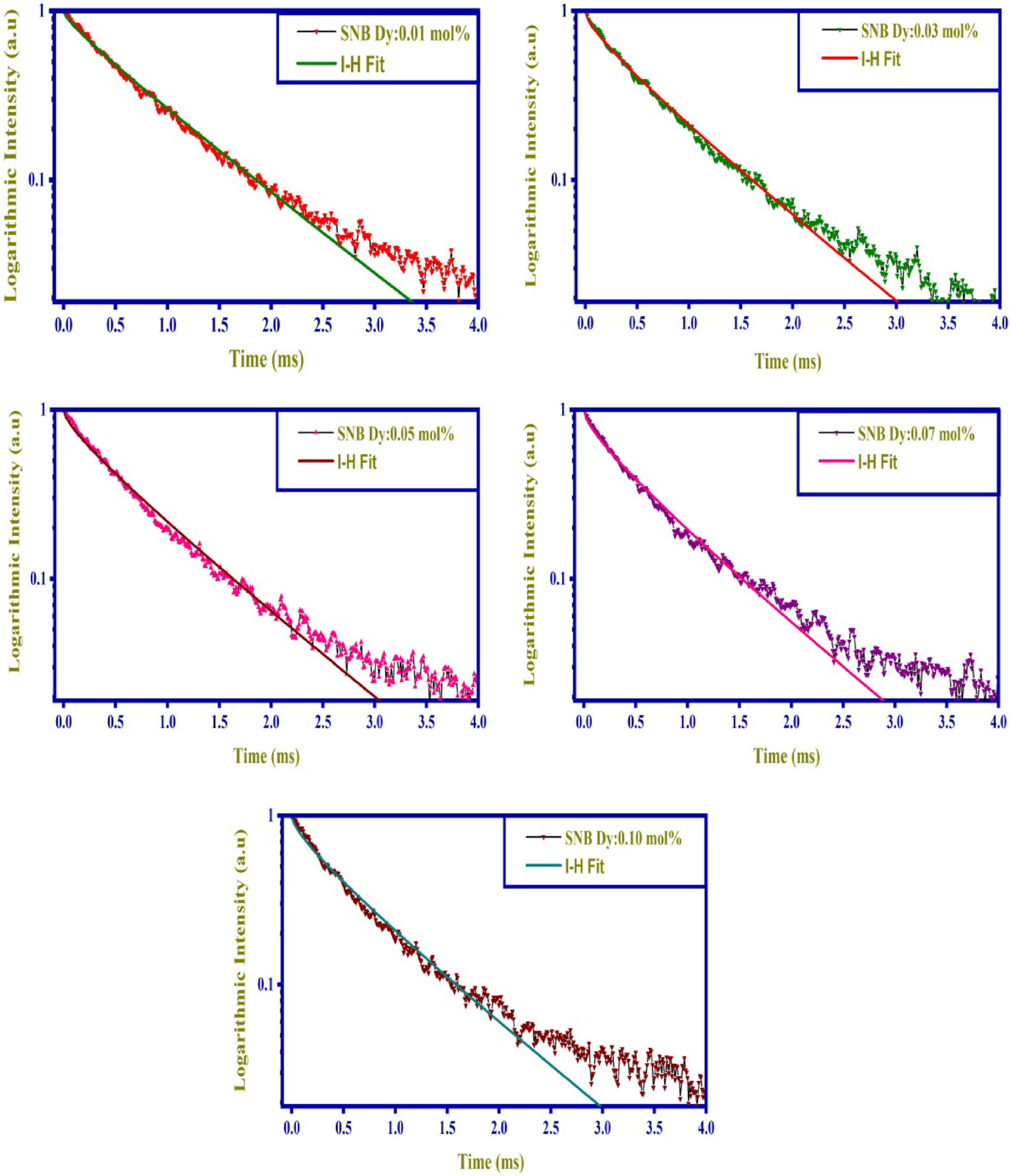


Fig. 15 I–H model fit of SNB:xDy^{3+} samples

interactions which is dominated by non-radiative transitions rather than efficient ET [74]. The inset of Fig. 17 exhibits the plot between relative integrated intensity and temperature for $^4F_{9/2} \rightarrow {}^6H_{13/2}$ transition, taking the intensity of PL spectra at 30 °C as 100% of the optimized sample. The PL intensity of the as-prepared SNB phosphor at 150 and 200 °C comes out to be 83% and 71%, respectively, in

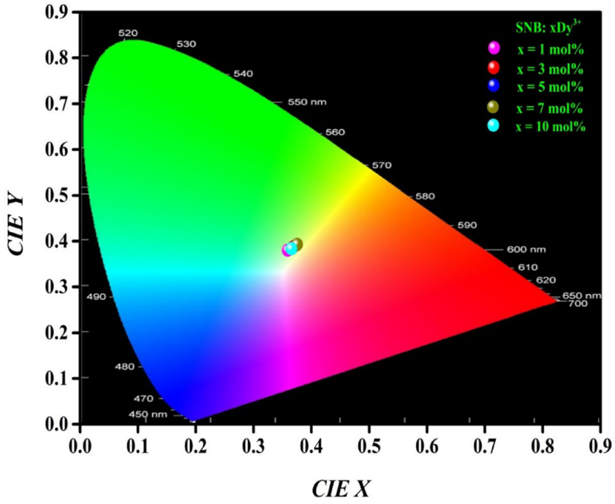


Fig. 16 CIE chromaticity diagram of SNB:xDy^{3+} phosphor

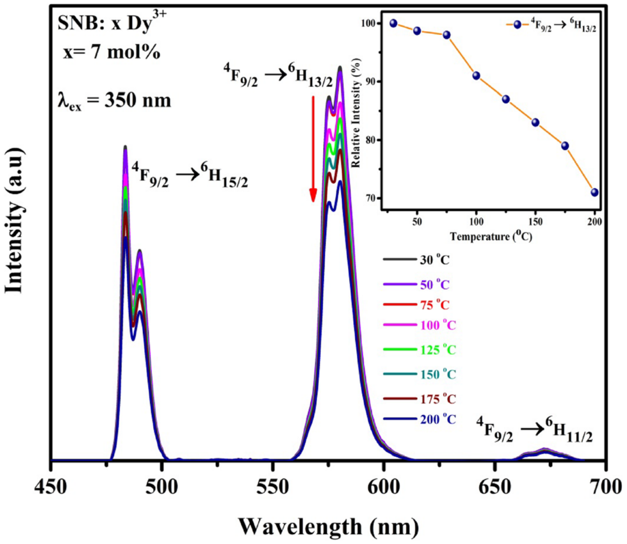


Fig. 17 Temperature-dependent PL properties of 7 mol% Dy^{3+} ions-doped SNB phosphor [Inset displays the relative emission intensity variation for $^4F_{9/2} \rightarrow {}^6H_{13/2}$ transition with temperature in the range from 30 to 200 °C]

contrast to the values at 30 °C, which is higher than the other cited work [75, 76]. The system returns to the ground state by means of a non-radiative multiphonon relaxation which increases with the increment of the temperature and finally results in thermal quenching. The above results conclude that the as-prepared SNB phosphor has exquisite thermal stability. The ΔE (activation energy) has been computed by using the Arrhenius equation mentioned below [77, 78]:

$$I_T = \frac{I_o}{1 + Cexp\left(-\frac{\Delta E}{K_B T}\right)}, \tag{7}$$

where I_o and I_T represent the emission intensity at 30 °C and at a particular temperature T, respectively. K_B and C denote the Boltzmann constant and arbitrary constant, respectively. For determining ΔE, the graph between the ln[(I_o/ I_T)-1] and 1/ $K_B T$ is plotted. Figure 18 represents the plot with linear fitting and the slope comes out to be -0.135. Hence, the ΔE can be 0.135 eV approximately for Dy^{3+} (7 mol%) ions-doped SNB phosphor, which is higher in comparison to the previously reported work [23, 79]. Aforementioned TDPL results indicate that as-prepared phosphor is thermally stable.

4 Conclusion

In conclusion, Dy^{3+} ions-doped SNB phosphors were effectively produced using a solid-state reaction technique. The peaks in the XRD pattern can be identified to be a cubic $Sr_4Nb_2O_9$ crystal structure, well matched with the JCPDS card no. 048-0558 of SNB phosphor. The average crystallite size of Dy^{3+} (7 mol%) ions-doped SNB phosphor was 43.65 nm. The SEM image manifests agglomerated particles with smooth surfaces having spherical shapes and sizes in the range 1–1.5 µm. Using the EDS spectrum, all of the elements in the desired samples were revealed. The sample's optical band gaps (E_g) lie in the range 4.005–4.038 eV, which were determined using DRS. J–O intensity parameters were evaluated

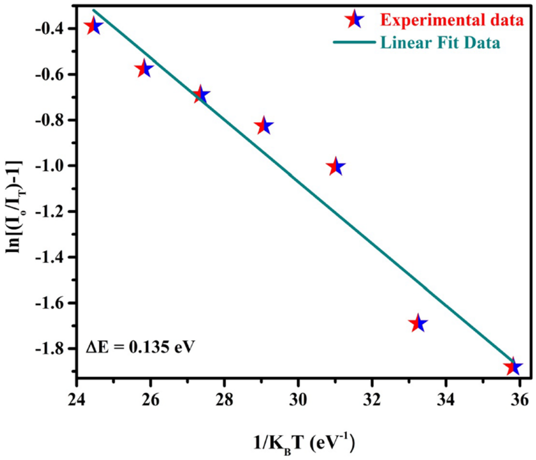


Fig. 18 Plot between ln[(I_0/I_T)-1] versus 1/K_BT

by transforming DRS spectra into absorption spectra. The excitation spectra affirm the several excitation bands in UV/n-UV and blue regions. At the most intense excitation peak (350 nm), recorded emission spectra contain two significant peaks in yellow and blue regions arising out of the transitions $^4F_{9/2} \rightarrow {}^6H_{13/2}$ and $^4F_{9/2} \rightarrow {}^6H_{15/2}$, respectively. The optimal concentration of Dy^{3+} ions in the Dy^{3+} ions-doped SNB phosphor was found to be 7 mol%. Various radiative parameters have been evaluated by correlating absorption and emission spectra. The d–d interaction was shown to be the reason of the concentration quenching, which was also confirmed by the I–H model. The quantum efficiency of the Dy^{3+} (7 mol%) ions-doped SNB sample comes out to be 71.12%. The samples' CIE coordinates lie in the cool white region. The effect of temperature on the PL characteristics shows that the intensity endures to 71% at 200 °C that of to 30 °C. The above-mentioned results reflect that the Dy^{3+} ions-activated SNB phosphor can serve as a host for white LEDs and other luminous devices.

Acknowledgements

One of the authors, Ms. Anu is very much thankful to CSIR/UGC for the award of a Junior Research Fellowship (JRF) (UGC Ref.1352: CSIR-UGC NET DEC. 2018) to her.

Author contributions

RL (First Author): Conceptualization, Methodology, Writing—original Draft. AK: Data Curation, Writing—review & editing. MKS: Formal Analysis, Data Curation. A: Formal Analysis. VK: Resources. ND (Corresponding Author): Supervision, Methodology, Software, Validation, Writing—review & editing. ASR: Writing—review & editing.

Funding

No funding was received to assist with preparation of this manuscript. The authors declare they have no financial interest.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. The datasets are presented in the main manuscript.

Declarations

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.

Research involving in human and animal participants No human participants and animals are involved in this research.

References

1. G.B. Nair, S. Tamboli, S.J. Dhoble, H.C. Swart, Structural and luminescence properties of thermally stable cool-white light emitting $NaCaPO_4$: Dy^{3+} phosphor. Optik **219**, 165026 (2020)
2. L. Han, H. Yao, S. Miao, S. Wang, J. Zhao, T. Sun, C. Guo, Z. Ci, C. Wang, Morphology controllable synthesis and upconversion luminescence of $Gd_4O_3F_6$: Ho^{3+}, Yb^{3+} for temperature sensing. J. Lumin. **197**, 360–369 (2018)
3. L. Zhao, P. Xu, F. Fan, J. Yu, Y. Shang, Y. Li, L. Huang, R. Yu, Synthesis and photoluminescence properties of Sm^{3+} and Dy^{3+} ions activated double perovskite Sr_2MgTeO_6 phosphors. J. Lumin. **207**, 520–525 (2019)
4. G. Lu, B. Deng, Y. Zhang, Y. Wang, Y. Lin, K. Jiang, Y. Shao, D. Zhang, R. Yu, Eu^{3+}-activated $Sr_3LaNb_3O_{12}$ red-emitting phosphors with excellent color stability for high color rendering w-LEDs. J Mater Sci: Mater Electron **33**, 17855–17867 (2022)
5. N. Deopa, A.S. Rao, Photoluminescence and energy transfer studies of Dy^{3+} ions doped lithium lead alumino borate glasses for w-LED and laser applications. J. Lumin. **192**, 832–841 (2017)
6. C. Wei, D. Xu, Z. Yang, J. Li, X. Chen, X. Li, J. Sun, A novel orange–red emitting phosphor Sr_2LuTaO_6:Sm^{3+} for WLEDs. J Mater Sci: Mater Electron **30**, 9303–9310 (2019)
7. F.B. Xiong, H.F. Lin, Z. Ma, Y.P. Wang, H.Y. Lin, X.G. Meng, H.X. Shen, W.Z. Zhu, Luminescence properties of a novel red-emitting phosphor $LaBMO_6$: Pr^{3+} (M =W, Mo). Opt. Mater. **66**, 474–479 (2017)
8. M.K. Sahu, M. Jayasimhadri, Conversion of blue emitting thermally stable $Ca_3Bi(PO_4)_3$ host as a color tunable phosphor

via energy transfer for luminescent devices. J. Lumin. **227**, 117570 (2020)

9. M.K. Sahu, H. Kaur, B.V. Ratnam, J. Suresh Kumar, M. Jayasimhadri, Structural and spectroscopic characteristics of thermally stable Eu^{3+} activated barium zinc orthophosphate phosphor for white LEDs. Ceram. Int. **46**, 26410–26415 (2020)
10. J. Chen, S. Zhao, Z. Zhao, M. Liao, S. Pan, J. Feng, D. Zhu, W. Pang, J. Lin, Z. Mu, The structure and luminescence properties of blue–green-emitting Sr_2YNbO_6: Bi^{3+} phosphors. J. Lumin. **239**, 118336 (2021)
11. R.N. Perumal, G. Subalakshmi, E. Varadarajan, S. Sadhasivam, Optical properties of Eu^{3+} activated $SrLa_2O_4$ red-emitting phosphors for WLED applications. J Mater Sci: Mater Electron. **29**, 2638–2644 (2018)
12. J. Hou, P. Chen, G. Zhang, Y.Z. Fang, W. Jiang, F. Huang, M. Zifeng, Synthesis, structure and photoluminescence properties of tetragonal tungsten bronze-type Eu^{3+}-doped $K_2LaNb_5O_{15}$ niobate phosphor. J. Lumin. **146**, 97–101 (2014)
13. Y. Han, S. Wang, H. Liu, L. Shi, J. Zhang, Z. Zhang, Z. Mao, D.J. Wang, Z.F. Mu, Z.W. Zhang, Y. Zhao, Synthesis and luminescent properties of a novel deep-red phosphor Sr_{2-}GdNbO$_6$: Mn^{4+} for indoor plant growth lighting. J. Lumin. **220**, 116968 (2020)
14. T. Jia, Z. Ci, Q. Wu, G. Zhu, C. Wang, Y. Wang, ECS. J. Solid State Sci. Technol. **4**(5), 78–82 (2015)
15. B. Liu, C. Shi, Z. Qi, Potential white-light long phosphor Dy^{3+} doped aluminates. Appl. Phys. Lett. **86**, 191111-1–191113 (2005)
16. G. Liu, B. Jacquier, *Spectroscopic properties of rare earths in optical materials* (Tsinghua University Press, China, 2005)
17. L. Su, X. Fan, Y. Liu, Z. Wu, G. Cai, X. Wang, Structure and luminescent properties of new $Dy^{3+}/Eu^{3+}/Sm^{3+}$-activated $InNbTiO_6$ phosphors for white UV-LEDs. Opt. Mater. **98**, 10940 (2019)
18. J.Y. Si, S.Y. Song, N. Liu, G.M. Cai, L.M. Su, Synthesis and photoluminescence of host-sensitized $MgNb_2O_6$ based phosphors. J. Lumin. **198**, 10–18 (2018)
19. G. Zhu, Z. Li, F. Zhou, M. Gao, C. Wang, S. Xin, Novel layered niobate phosphors $SrBaNb_4O_{12}$: Re^{3+} (Re= Eu, Dy, Sm and Pr): crystal structure, electronic structure and luminescence property Investigation. J. Lumin. **211**, 76–81 (2019)
20. M.K. Ekmekci, M. Ilhan, L.F. Güleryüz, A. Mergen, Study on molten salt synthesis, microstructural determination and white light emitting properties of $CoNb_2O_6$:Dy^{3+} phosphor. Optik **128**, 26–33 (2017)
21. B. Deng, J. Jiang, W. Chen, A. Zhang, Z. Liang, F. Li, F. Zeng, G. Zhang, New Dy^{3+} -activated $KCa_2Nb_3O_{10}$ yellow-emitting phosphors for w-LEDs application: preparation and optical properties. Inorg. Chem. Commun. **145**, 110051 (2022)
22. M. İlhan, İÇ. Keskin, Analyzing of Judd-Ofelt parameters and radioluminescence results of $SrNb_2O_6$: Dy^{3+} phosphor synthesized via molten salt method. Phys. Chem. Chem. Phys. **22**, 19769–19778 (2020)
23. D. Xu, Z. Yang, J. Sun, X. Gao, J. Du, Synthesis and luminescence properties of double-perovskite white emitting phosphor Ca_3WO_6: Dy^{3+}. J. Mater. Sci. Mater. Electron. **8**, 8370–8377 (2016)
24. L. Alexander, H.P. Klug, Determination of crystallite size with the x-ray spectrometer. J. Appl. Phys. **21**, 137–142 (1950)
25. A.K. Bedyal, A.K. Kunti, V. Kumar, H.C. Swart, Effects of cationic substitution on the luminescence behavior of Dy^{3+} doped orthophosphate phosphor. J. Alloys compd. **806**, 1127–1137 (2019)
26. I.P. Sahu, D.P. Bisen, N. Brahme, R.K. Tamrakar, Generation of white light from dysprosium-doped strontium aluminate phosphor by a solid-state reaction method. J. Electron. Mater. **45**, 2222–2232 (2016)
27. M. Afqir, A. Tachafine, D. Fasquelle, M. Elaatmani, J.C. Carru, A. Zegzouti, M. Daoud, S. Sayouri, T.D. Lamcharfi, M. Zouhairi, Structural, electric and dielectric properties of Eu-doped $SrBi_2Nb_2O_9$ ceramics obtained by co-precipitation route. Process. Appl. Ceram. **12**, 72–77 (2018)
28. S. Lanfredi, D.H.M. Genova, I.A.O. Brito, A.R.F. Lima, M.A.L. Nobre, Structural characterization and Curie temperature determination of a sodium strontium niobate ferroelectric nanostructured powder. J. Solid State Chem. **184**, 990–1000 (2011)
29. M. Afqir, A. TachaÞne, D. Fasquelle, M. Elaatmani, J.C. Carru, A. Zegzouti, M. Daoud, Synthesis and characterizations of Ho_2O_3 modified $SrBi_2Nb_2O_9$ ceramics. Chin. J. Phys. **56**, 1158–1165 (2018)
30. D. Nelis, D. Mondelaers, G. Vanhoyland, A. Hardy, K. Van Werde, H. Van den Rul, M.K. Van Bael, J. Mullens, L.C. Van Poucke, J. D'Haen, Synthesis of strontium bismuth niobate ($SrBi_2Nb_2O_9$) using an aqueous acetate–citrate precursor gel: thermal decomposition and phase formation. Thermochim. Acta **426**, 39–48 (2005)
31. W.T. Carnall, P.R. Fields, K. Rajnak, Electronic energy levels in the trivalent lanthanide aquo ions. I. Pr3+, Nd3+, Pm3+, Sm3+, Dy3+, Ho3+, Er3+, and Tm3+. J. Chem. Phys. **49**, 4424 (1968)
32. A.K. Bedyal, V. Kumar, R. Prakash, O.M. Ntwaeaborwa, H.C. Swart, A near-UV-converted $LiMgBO_3$: Dy^{3+} nanophosphor: surface and spectral investigations. Appl. Surf. Sci. **329**, 40–46 (2015)

33. M.K. Sahu, M. Jayasimhadri, K. Jha, B. Sivaiah, A.S. Rao, D. Haranath, Synthesis and enhancement of photoluminescent properties in spherical shaped Sm^{3+}/Eu^{3+} co-doped $NaCaPO_4$ phosphor particles for w-LEDs. J. Lumin. **202**, 475–483 (2018)
34. Z. Xia, Y. Zhang, M.S. Molokeev, V.V. Atuchin, Structural and luminescence properties of yellow-emitting $NaScSi_2O_6$: Eu^{2+} phosphors: Eu^{2+} site preference analysis and generation of red emission by codoping Mn^{2+} for white-light-emitting diode applications. J. Phys. Chem. C. **117**, 20847–20854 (2013)
35. R. Lopez, R. Gomez, Band-gap energy estimation from diffuse reflectance measurements on sol-gel and commercial TiO_2: a comparative study. J. Sol. Gel Sci. Technol. **61**, 1–7 (2012)
36. R. Köferstein, L. Jäger, S.G. Ebbinghaus, Magnetic and optical investigations on $LaFeO_3$ powders with different particle sizes and corresponding ceramics. Solid State Ion. **249–250**, 1–5 (2013)
37. D.E. Henrie, Percent covalency and the nephelauxetic effect in lanthanide complexes. J. Mol. Phys. **28**, 415–421 (2006)
38. I. Bulus, S.A. Dalhatu, R. Hussin, W.N. Wan Shamsuri, Y.A. Yamusa, The role of dysprosium ions on the physical and optical properties of lithium boro sulfo phosphate glasses. Int J. Mod. Phys. B **31**(13), 1750101 (2017)
39. Y.A. Yamusa, R. Hussin, W.N. Wan Shamsuri, Effect of Dy^{3+} on the physical, optical and radiative properties of $CaSO_4$–B_2O_3–P_2O_5 glasses. Indian J. Phys. **93**, 15–26 (2019)
40. P.V. Do, V.P. Tuyen, V.X. Quang, N.T. Thanh, V.T. Thai Ha, H. Van Tuyen, N.M. Khaidukov, J. Marcazzó, Y. Lee, B.T. Huy, Optical properties and Judd-Ofelt parameters of Dy^{3+} doped K_2GdF_5 single crystal. Opt. Mater. **35**, 1636–1641 (2013)
41. Manjeet, A. Kumar, Anu, Ravina, N. Deopa, A. Kumar, R.P. Chahal, S. Dahiya, R. Punia, A.S. Rao, Structural, thermal, optical and luminescence properties of Dy^{3+} ions doped zinc potassium alumino borate glasses for optoelectronics applications. J. Non-Cryst. Solids **588**, 121613 (2022)
42. B.R. Judd, Optical absorption intensities of rare-earth ions. Phys. Rev. **127**, 750–761 (1962)
43. G.S. Ofelt, Intensities of crystal spectra of rare-earth ions. J. Chem. Phys. **37**, 511–520 (1962)
44. A. Ichoja, S. Hashim, S.K. Ghoshal, Judd-Ofelt calculations for spectroscopic characteristics of Dy^{3+}-activated strontium magnesium borate glass. Optik **218**, 165001 (2020)
45. A.K. Kunti, N. Patra, S.K. Sharma, H.C. Swart, Radiative transition probability enhancement of white light emitting Dy^{3+} doped and K^+ co-doped $BaWO_4$ phosphors via charge compensation. J. Alloys. Compd. **735**, 2410–2422 (2018)
46. S. Selvi, G. Venkataiah, A. Selvaraj, K. Marimuthu, Structural and luminescence studies on Dy^{3+} doped lead boro–telluro-phosphate glasses. Phys. B Condens. Matter **454**, 72–81 (2014)
47. E. Cavalli, Optical spectroscopy of Dy^{3+} in crystalline hosts: general aspects, personal considerations and some news. Opt. Mater. X **1**, 100014 (2019)
48. P. Dewangan, D.P. Bisen, N. Brahme, S. Sharma, Structural characterization and luminescence properties of Dy^{3+} doped $Ca_3MgSi_2O_8$ phosphors. J. Alloys compd. **10**, 390 (2018)
49. R.N. Perumala, A.X. Lopeza, G. Subalakshmi, Synthesis and multi-colour luminescence spectra of RE^{3+} (RE^{3+}=Eu^{3+}, Sm^{3+}, Dy^{3+}, $Eu^{3+}/Sm^{3+}/Dy^{3+}$) doped $BiLa_2O_4$ phosphors. Optik (Stuttg.) **170**, 125–131 (2018)
50. Hu. Xiaoxue, S. Yi, B. Liang, Hu. Gengqiao, Y. Wang, Synthesis and photoluminescence properties of Er^{3+} and Dy^{3+} doped Na_2NbAlO_5 phosphors. J. Rare Earth **36**, 789–794 (2018)
51. N.P. Rajesh, G. Subalakshmi, C.K. Jayasankar, Investigations on energy transfer and tunable luminescence spectra for single, co-doped and tri-doped RE^{3+}(RE^{3+}= Dy^{3+}, Sm^{3+} and Eu^{3+}) activated $Sr_{1.99}Bi_{0.01}CeO_4$ phosphors. Opt. Mater. **85**, 464–473 (2018)
52. S. Kaur, A.S. Rao, M. Jayasimhadri, Color tunability and energy transfer studies of Dy^{3+}/Eu^{3+} co-doped calcium aluminozincate phosphor for lighting applications. Mater. Res. Bull. **116**, 79–88 (2019)
53. N. Deopa, S. Saini, S. Kaur, A. Prasad, A.S. Rao, Spectroscopic investigations on Dy^{3+} ions doped zinc lead alumino borate glasses for photonic device application. J. Rare Earths **37**, 52–59 (2019)
54. L.W. Xu, H.M. Crosswhite, J.P. Hessler, Fluorescent and dynamic properties of optically excited dysprosium trifluoride. J. Chem. Phys. **81**, 698–703 (1984)
55. D.L. Dexter, J.H. Schulman, Theory of concentration quenching in inorganic phosphors. J. Chem. Phys. **22**, 1063–1070 (1954)
56. T. Krishnapriya, A. Jose, T.A. Jose, C. Joseph, N.V. Unnikrishnan, P.R. Biju, Luminescent kinetics of Dy^{3+} doped $CaZn_2(PO_4)_2$ phosphors for white light emitting applications. Adv Powder Technol. **32**, 1023–1032 (2021)
57. G. Blasse, Energy transfer in oxidic phosphors. Philips Res. Rep. **24**, 131–144 (1969)
58. G. Blasse, B.C. Grabmaier, *Energy transfer in luminescent materials* (Springer, Berlin, Heidelberg, 1994), pp.91–107
59. G. ChinnaRam, T. Narendrudu, S. Suresh, A. SuneelKumar, M.V. Sambasiva Rao, V. RaviKumar, D. KrishnaRao, Investigation of luminescence and laser transition of Dy^{3+} ion in P_2O_5 PbO Bi_2O_3 R_2O_3 (R = Al, Ga, In) glasses. Opt. Mater. **66**, 189–196 (2017)

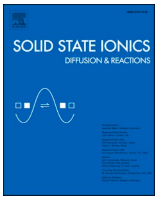
60. H. George, N. Deopa, S. Kaur, A. Prasad, M. Sreenivasulu, M. Jayasimhadri, A.S. Rao, Judd-Ofelt parametrization and radiative analysis of Dy^{3+} ions doped sodium bismuth strontium phosphate glasses. J. Lumin. **215**, 116693 (2019)
61. L. Marek, M. Sobczyk, V.A. Trush, K. Korzeniowski, J. Legendziewicz, Synthesis, structure and radiative and non-radiative properties of a new Dy^{3+} complex with sulfonylamidophosphate ligand. J. Rare Earths **37**, 1255–1260 (2019)
62. K. Jha, M. Jayasimhadri, Spectroscopic investigation on thermally stable Dy^{3+} doped zinc phosphate glass for white light emitting diodes. J. Alloys Compd. **688**, 833–840 (2016)
63. Y. Liu, G. Liu, J. Wang, X. Dong, W. Yu, Single-component and warm-white-emitting phosphor $NaGd(WO_4)_2$: Tm^{3+}, Dy^{3+}, Eu^{3+}: synthesis, luminescence, energy transfer, and tunable color. Inorg. Chem. **53**, 11457–11466 (2014)
64. S. Kaur, A.S. Rao, M. Jayasimhadri, V.V. Jaiswal, D. Haranath, Tb^{3+} ion induced colour tunability in calcium aluminozincate phosphor for lighting and display devices. J. Alloys Compd. **826**, 154212 (2020)
65. Ravina, Naveen, Sheetal, V. Kumar, S. Dahiya, N. Deopa, R. Punia, A.S. Rao, Judd- Ofelt itemization and influence of energy transfer on Sm^{3+} ions activated B_2O_3–ZnF_2–SrO–SiO_2 glasses for orange-red emitting devices. J. Lumin. **229**, 117651 (2021)
66. R.A. Talewara, S.K. Mahamudaa, K. Swapnaa, M. Venkateswarlua, A.S. Rao, Spectroscopic studies of Sm^{3+} ions doped alkaline-earth chloro borate glasses for visible photonic applications. Mater. Res. Bull. **105**, 45–54 (2018)
67. C.S. McCamy, Correlated color temperature as an explicit function of chromaticity coordinates. Color Res. Appl. **17**, 142–144 (1992)
68. G. Zhu, Z. Li, C. Wang, X. Wang, F. Zhou, M. Gao, S. Xin, Y. Wang, Highly Eu^{3+} ions doped novel red emission solid solution phosphors $Ca_{18}Li_3(Bi, Eu)(P_O4_)_14$: structure design, characteristic luminescence and abnormal thermal quenching behavior investigation. Dalton Trans. **48**, 1624–1632 (2019)
69. S. Sharma, N. Brahme, D.P. Bisen, P. Dewangan, Cool white light emission from Dy^{3+} activated alkaline alumino silicate phosphors. Opt. Exp. **26**, 29495–29508 (2018)
70. P. Sehrawat, R.K. Malik, N. Kumari, M. Punia, S.P. Khatkar, V.B. Taxak, Cool-white illumination characteristics of combustion-derived novel single-phase $Sr_9Al_6O_{18}$: Dy^{3+} nanomaterials for NUV induced WLEDs and solar cells. Chem. Phys. Lett. **770**, 138438 (2021)
71. C.M. Mehare, N.S. Dhoble, C. Ghanty, S.J. Dhoble, Photoluminescence and thermoluminescence characteristics of $CaAl_2Si_4O_{12}$: Dy^{3+}new phosphor prepared by combustion synthesis. J. Mol. Struct. **1227**, 129417 (2021)
72. G. Yuan, R. Cui, J. Zhang, X. Zhang, X. Qi, C. Deng, Photoluminescence evolution and high thermal stability of orange red-emitting $Ba_{3-x}Sr_xZnNb_2O_9$: Eu^{3+} phosphors. J. Solid State Chem. **303**, 122447 (2021)
73. D. Cui, Z. Song, Z. Xia, Q. Liu, Luminescence tuning, thermal quenching, and electronic structure of narrow-band red-emitting nitride phosphors. Inorg. Chem. **56**, 11837–11844 (2017)
74. V. Naresh, N. Lee, Energy transfer dynamics in thermally stable single-phase $LiMgBO_3$:Tm^{3+}/ Dy^{3+} phosphor for UV triggered white light-emitting devices. MSEB **271**, 115306 (2021)
75. S. Liumei, X. Fan, Y. Liu, W. Zhanchao, G. Cai, X. Wang, Structure and luminescent properties of new Dy^{3+}/Eu^{3+}/Sm^{3+}-activated $InNbTiO_6$ phosphors for white UV-LEDs. Opt. Mater. **98**, 109403 (2019)
76. T. Wang, H. Yihua, L. Chen, X. Wang, M. He, A white-light emitting phosphor $LuNbO_4$: Dy^{3+} with tunable emission color manipulated by energy transfer from NbO_{43}- groups to Dy^{3+}. J. Lumin. **181**, 189–195 (2017)
77. P. Li, M. Peng, X. Yin, Z. Ma, G. Dong, Q. Zhang, J. Qiu, Temperature dependent red luminescence from a distorted Mn^{4+} site in $CaAl_4O_7$: Mn^{4+}. Opt. Express **21**, 18943–18948 (2013)
78. M.K. Sahu, J. Mula, White light emitting thermally stable bismuth phosphate phosphor $Ca_3Bi(PO_4)_3$: Dy^{3+} for solid-state lighting applications. J. Am. Ceram. Soc. **102**, 6087–6099 (2019)
79. W.R. Liu, C.H. Huang, C.W. Yeh, J.C. Tsai, Y.C. Chiu, Y.T. Yeh, R.S. Liu, A Study on the luminescence and energy transfer of single-phase and color-tunable $KCaY(PO_4)_2$: Eu^{2+}, Mn^{2+} phosphor for application in white-light LEDs. Inorg. Chem. **51**, 9636 (2012)

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

Contents lists available at ScienceDirect

Solid State Ionics

journal homepage: www.elsevier.com/locate/ssi



Study of lithium diffusion properties and electrochemical performance of SnSe/C and SnSe/MWCNT composite anode for Li-ion batteries

Shivangi Rajput [a], Amrish K. Panwar [a,*], Amit Gupta [b]

[a] Department of Applied Physics, Delhi Technological University, New Delhi, India
[b] Department of Mechanical Engineering, Indian Institute of Technology, New Delhi, India

ARTICLE INFO

Keywords:
Lithium-ion battery
Tin selenide
Composite
Anode materials
Carbon black
MWCNT

ABSTRACT

Herein, synthesis and electrochemical analysis of SnSe based alternative alloy anode and its composites with Super P (SnSe/C) and MWCNT (SnSe/MWCNT) have been attempted via a facile high-energy ball milling method. The X-ray diffraction (XRD) of the synthesized materials has been observed to confirm the phase formation. The morphological studies are analyzed using FESEM and TEM to see the shape, size and distribution of particles. In contrast, Energy Dispersive Spectroscopy (EDS) and Raman spectroscopy confirm the homogeneous mixing of SnSe with Super P and MWCNT. Electrochemical analysis of all three compositions has been carried to explore the specific capacity, efficiency and cyclic performance. The lithium-ion diffusion coefficient is estimated using CV at different scan rates, EIS, and GITT for the SnSe, SnSe/C, and SnSe/MWCNT samples. For all three prepared anodes, the Li^+ diffusion coefficient is calculated using EIS, CV, and GITT and it was found to be in the range of 10^{-15} cm^2s^{-1}, 10^{-14} to 10^{-15} cm^2s^{-1} and 10^{-12} to 10^{-14} cm^2s^{-1}, respectively. Among all three anode materials, SnSe/MWCNT showed higher values of Li-ion diffusion coefficient, indicating that the SnSe/MWCNT electrode has superior kinetics over SnSe/C and SnSe.

1. Introduction

With increasing development in the field of the global economy and with the growth of population, several problems, such as climate change and the excessive depletion of fossil fuels, have driven significant worldwide attention toward the development of energy storage devices for renewable sources. In the recent past, Lithium-ion Batteries have gained significant popularity in the field of electronic and energy storage devices because of their outstanding features such as high energy density, low maintenance, very low memory effect, and minimal self-discharge [1]. Furthermore, LIBs have been demonstrated to be one of the most efficient strategies for storing energy for various portable to large-scale devices such as mobile phones, laptops, digital electronics, and electric vehicles [2]. The most vital part of a battery is the electrode material that decides the transportation of charges and storage of energy [3]. Hence, the key to achieving better electrochemical performance is the electrode material [4,5]. Since the commercialization of LIBs in the year 1990–91, the commonly used anode in LIBs was graphite. Still, it has certain limitations, such as low gravimetric and volumetric capacity as well as safety concerns which make graphite inappropriate anode material for the next generation LIBs. On the other hand, carbonaceous electrode materials have the low theoretical capacity and cannot alone be sufficient to meet the modern power requirements [6]. Therefore, developing alternative novel anode materials is of utmost need to aid in high storage capability and safer operability of LIBs. The elements belonging to group IV like Si, Ge and Sn have proven themselves among the favourable anode materials owing to their higher specific capacities in comparison with carbon-based anode material. Out of these anode materials, tin has gained immense attention due to its theoretical capacity of 994 $mAhg^{-1}$, high packaging density and safer thermodynamic potential in comparison to carbonaceous anode materials [7–9].

Metallic Tin undergoes an alloying reaction with lithium leading to the formation of $Li_{4.4}Sn$, incorporating 4.4 Li-ions per unit formula and it also has a higher electrochemical potential vs Li/Li^+ as compared to graphite, which in result enhances its stability and safety as anode in Li-ion batteries. As a result, a considerable amount of research has been done on tin-based anodes. However, the commercialisation of tin-based anodes is still not possible due to their large volume expansion during cycling process leading to particle cracking and capacity deterioration [3,10,11]. To remove these obstacles, different strategies have been adapted such as nano-structuring, and the use of inter-metallics or composite material instead of pure metal. Nanosized anode materials

* Corresponding author.
E-mail address: amrish.phy@dtu.ac.in (A.K. Panwar).

https://doi.org/10.1016/j.ssi.2023.116206
Received 10 January 2023; Received in revised form 4 March 2023; Accepted 10 March 2023
Available online 20 March 2023

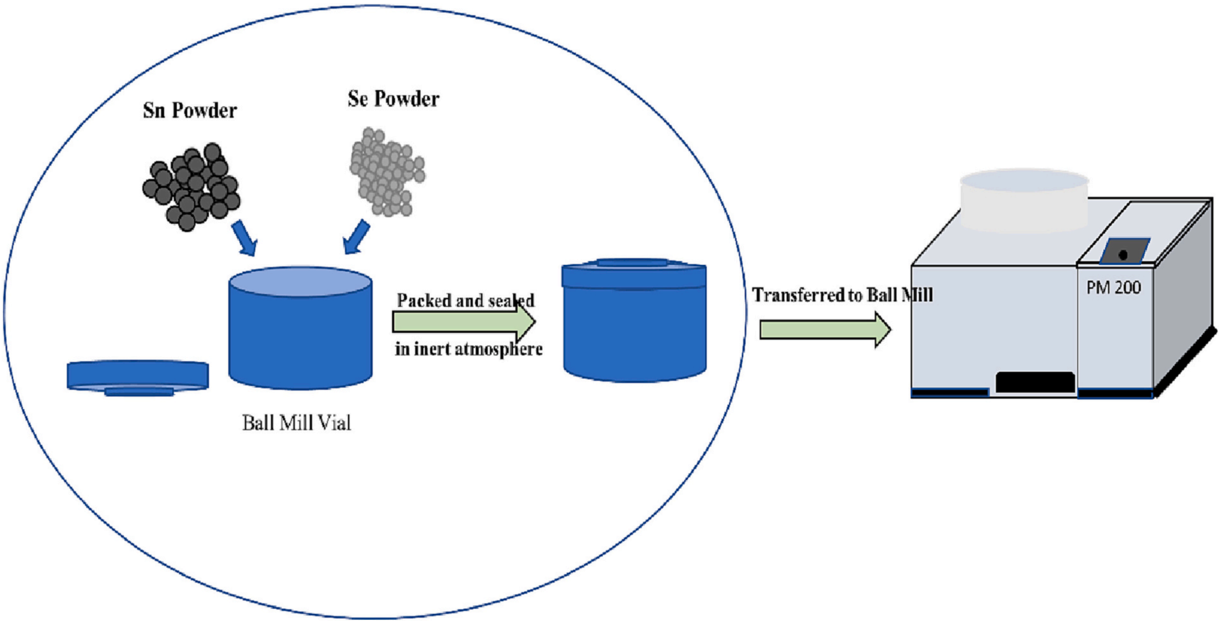


Fig. 1. Schematic diagram for synthesis of SnSe, SnSe/C and SnSe/MWCNT composite.

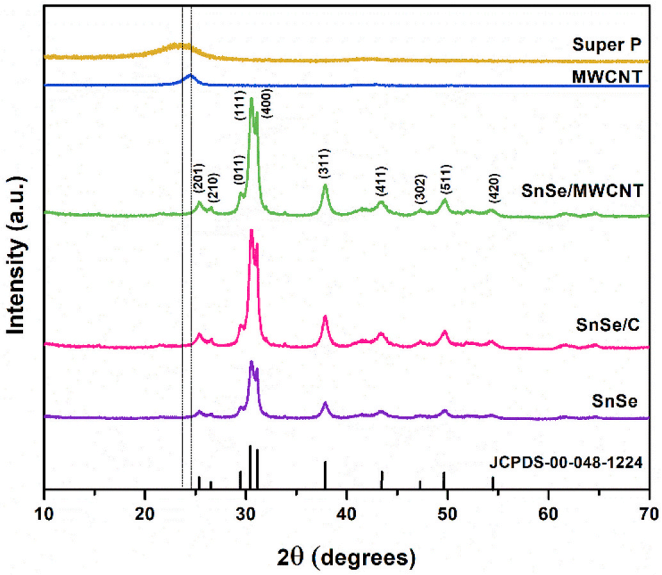


Fig. 2. XRD patterns for SnSe, SnSe/C and SnSe/MWCNT along with super P and MWCNT.

Table 1
Average Crystallite size for SnSe, SnSe/C and SnSe/MWCNT.

Parameter	SnSe	SnSe/C	SnSe/MWCNT
D(nm)	42	23	26
δ (nm^{-2}) × 10^{-3}	0.56	1.89	1.47

could offer a better solution by providing a good strain accommodation, improving the specific capacity and rate capability due to enhanced interfacial area and kinetics of Li-ion diffusion [12]. Sn easily reacts with elements of group VA and group VI A to form binary or ternary compounds (Sn-M, Sn-M-M'), hence, one such approach is to make alloy of Sn with inactive materials that could hinder the expansion/shrinkage of volume during charging or discharging such as Sn—Sb, Sn—Co, Sn—Ni, Sn—Cu etc. Another approach could be composites with conductive materials such as carbon, polypyrrole, polyaniline to enhance electrochemical performance [12,13].

In this study, SnSe alloy and its composites with Super P and Multi Walled Carbon Nanotubes (MWCNTs) have been synthesized via high energy ball mill method using Tin and Selenium metal powder, Super P and MWCNTs as precursors. Super P and MWCNTs are among the most carbonaceous materials with characteristics like high electronic conductivity, great thermal stability and mechanical properties. MWCNTs are widely used to make composites with anode materials owing to its excellent properties which enhances the cyclic performances of LIBs by preventing them from collapsing after long cycles and decreasing the volume expansion during cycling maintaining structural integrity. Super P acts as a buffer when used in anode material to suppress the volume change during lithiation/de-lithiation process.

2. Experimental

2.1. Synthesis

SnSe material was prepared using high energy ball mill (HEBM). For the synthesis, Sn powder (99.9% purity) and Se powder (99.5% purity) were first grounded and then added to a stainless-steel vial containing steel balls. The vial was sealed in glove box in order to create an inert atmosphere. Milling was done effectively for 15 h at a rotary speed of 350 rpm.

After the milling, SnSe particles were obtained. Similarly, SnSe/C and SnSe/MWCNT composite were synthesized using Sn metal powder, Se metal powder, Super P and MWCNTs under the same conditions as mentioned above. The Schematic diagram of synthesis route is shown in Fig. 1.

2.2. Structural and morphological characterization

For the determination of crystal structure, characterization of the synthesized SnSe, SnSe/C, and SnSe/MWCNT samples was facilitated by an X-ray Diffractometer (RIGAKU ULTIMA IV) equipped with CuK$_{\alpha 1}$ radiation having a wavelength of 1.54 Á. The XRD data was observed and recorded between the range of 10° to 70°. The morphologies were investigated using a Hitachi s-3700 scanning electron microscope. For TEM imaging, FEI Tecnai G2 20 S-Twin microscope was used. RAMAN

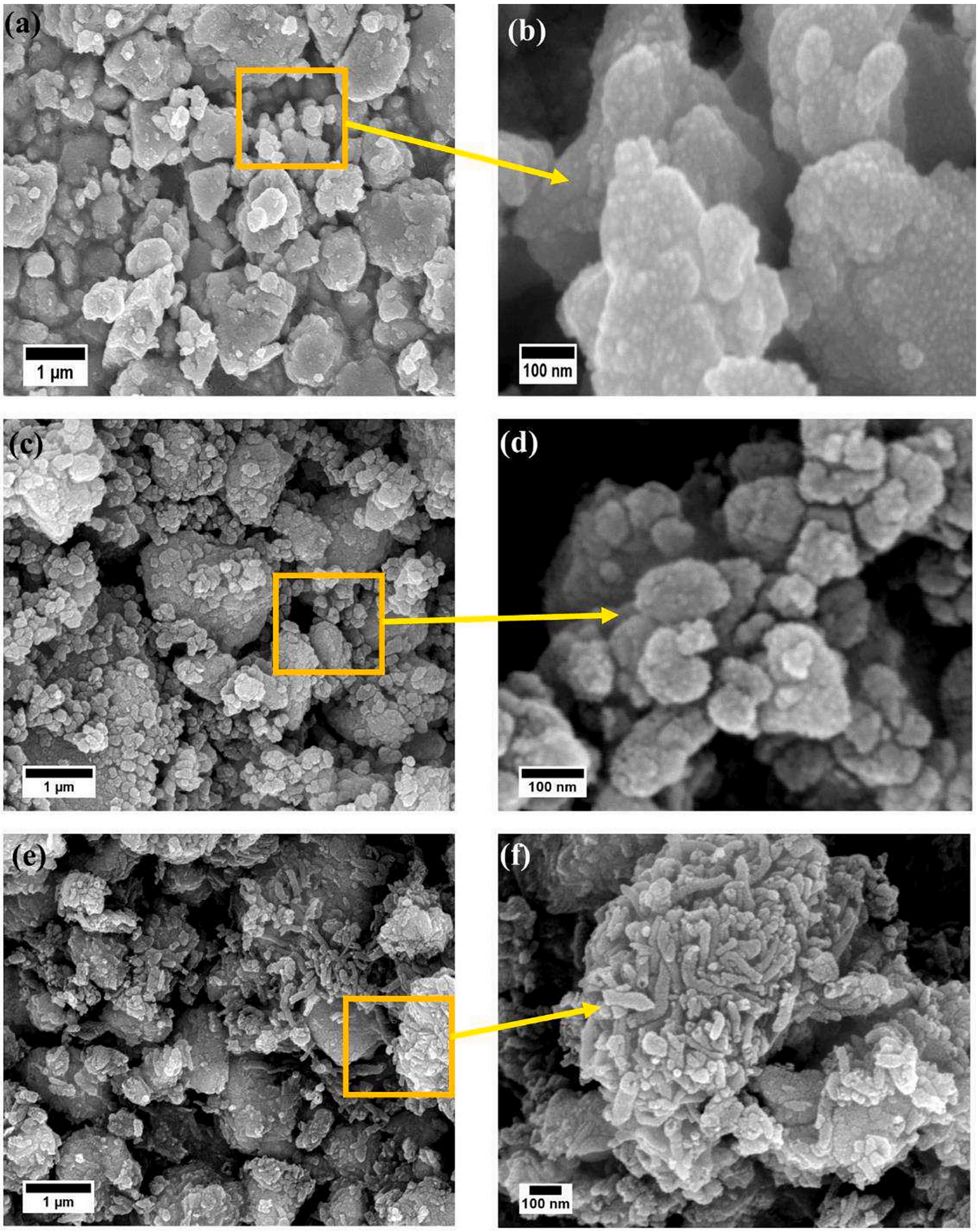


Fig. 3. FESEM micrographs of (a-b) SnSe, (c-d) SnSe/C, (e-f) SnSe/MWCNT.

studies were carried out using Renishaw Laser Spectrometer, Invia II model. For DC electrical measurements, activation energy was measured at a source voltage of 1.0 V, and DC resistance was measured at a varied voltage from −10 V to 10 V using a KEITHLEY 6517A electrometer. For the pellet formation of the prepared samples, they were homogeneously mixed with 2.5 wt% of Polyvinyl Alcohol (PVA), and then circular pellets of a diameter of 10 mm were made using a hydraulic press (pressure ~ 7 tons). The obtained pellets were then annealed at 250 °C for 2 h to evaporate the binder. Electrical contacts on pellets were made by coating both circular sides with silver paste and were then annealed at 150 °C for 1 h to eradicate the moisture from the surface of the pellet.

2.3. Electrochemical characterization

To perform electrochemical measurements, a homogenous slurry comprising active material, Super P and PVDF in the ratio 80:10:10 (by weight) in NMP solvent was prepared. Then, the prepared slurry was evenly spread over a Cu foil using automatic coating unit. Then the coated slurry was kept at 80 °C overnight to dry out the solvent in the slurry. After drying, the disk electrodes of 16 mm diameter were punched and, then the electrodes were pressed between two stainless steel twin rollers to improve adherence between active material and current collector. The half cell (CR2016) was fabricated in Mbraun glove box work station and for the assembly of half-cell lithium chip, celgard 2400 and 1 M solution of LiPF6 dissolved in DMC and EC in the ratio 1:1 (by volume) was used as counter electrode, separator and electrolyte, respectively.

3. Results and discussion

3.1. Structure and morphological analysis

The wide scan XRD patterns of synthesized SnSe, SnSe/C and SnSe/MWCNT samples are shown in Fig. 2 along with the individual super P and MWCNT. All obtained peaks in the XRD pattern are assigned to 25.20^o (201), 26.45^o (210), 29.42^o (011), 30.40^o (111), 31.08^o (400),

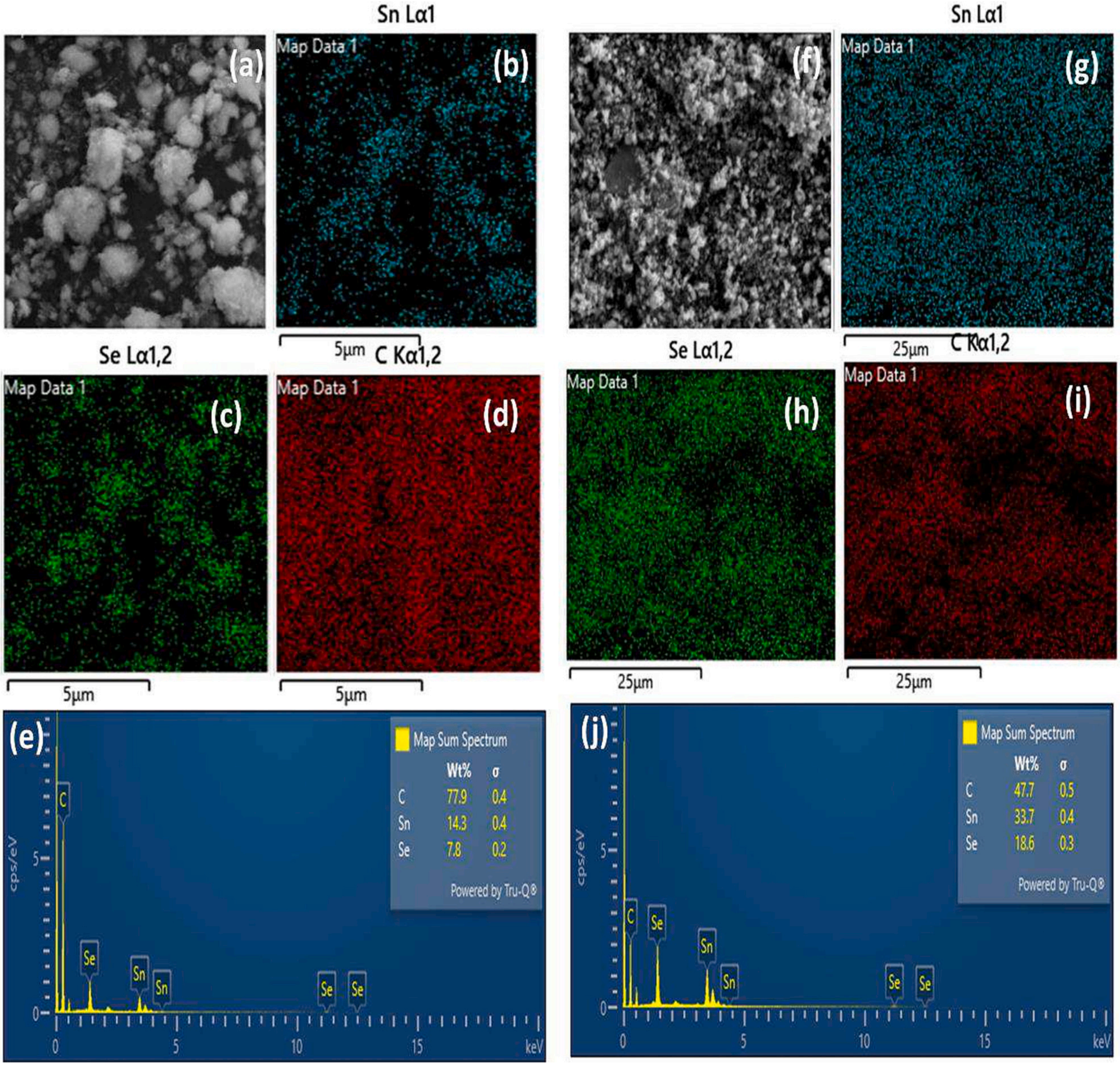


Fig. 4. Elemental mapping images of Sn(blue), Se(green), C (Red) of (a-d) SnSe/C composite, (f-i) SnSe/MWCNT, (e) the result of EDS analysis for SnSe/C, (j) the result of EDS analysis for SnSe/MWCNT. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

37.20^{o} (311), 43.34^{o} (411), 47.30^{o} (302), 49.60^{o} (511) and 54.47^{o} (420) planes of SnSe which matches well with the JCPDS file: 00–048-1224 as an orthorhombic structure with space group, pnma. No extra diffraction peaks were noticed. Hence, the XRD patterns confirm the high purity of the SnSe/C and SnSe/MWCNT composite in proper phase. It also indicates that the addition of Super P or MWCNT has no effect on the phase formation of SnSe and does not affect the crystal structure of SnSe. Also no oxidation reaction took place during the synthesis process [14]. The average crystallite size of the SnSe, SnSe/C and SnSe/MWCNT composites has been calculated using Scherrer's formula given by:

$$D = \frac{k\lambda}{\beta cos\theta} \tag{1}$$

where D, β, λ, k and θ represents crystallite size, FWHM, wavelength of the x-ray radiation, $CuK_{\alpha 1}$ (0.154 nm), shape constant, ~ 0.9 and Bragg's angle, respectively. The dislocation density, δ represents the deficiency in the particles can be calculated using the formula given below [15];

$$\delta = \frac{1}{D^2} \tag{2}$$

Table 1 summarizes the calculated values obtained from the XRD analysis.

Fig. 3 shows the FESEM images of the SnSe, SnSe/C and SnSe/MWCNT. As depicted in Fig. 3(a-b), SnSe consist of irregularly shaped micro-sized particles with the average particle size of 0.65 μm which consist of agglomerated primary nanoparticles. However, the addition of Super P enhances the morphology and regularity of SnSe particles as shown in Fig. 3(c-d). The SnSe/C composite displays the aggregated and nearly spherical shaped particles with average size of 0.1 μm consisting of agglomerated primary nanoparticle. Such morphology is quite good for electrode materials because the agglomeration of primary nanoparticles can increase the adherence between the electrode and the electrolyte, hence enhancing the battery performance. Whereas, SnSe/MWCNT composite shows slightly larger irregular shaped micro-sized particles. The average size of particles was found to be 0.3 μm. MWCNTs have cylindrical thread like structure which gives large

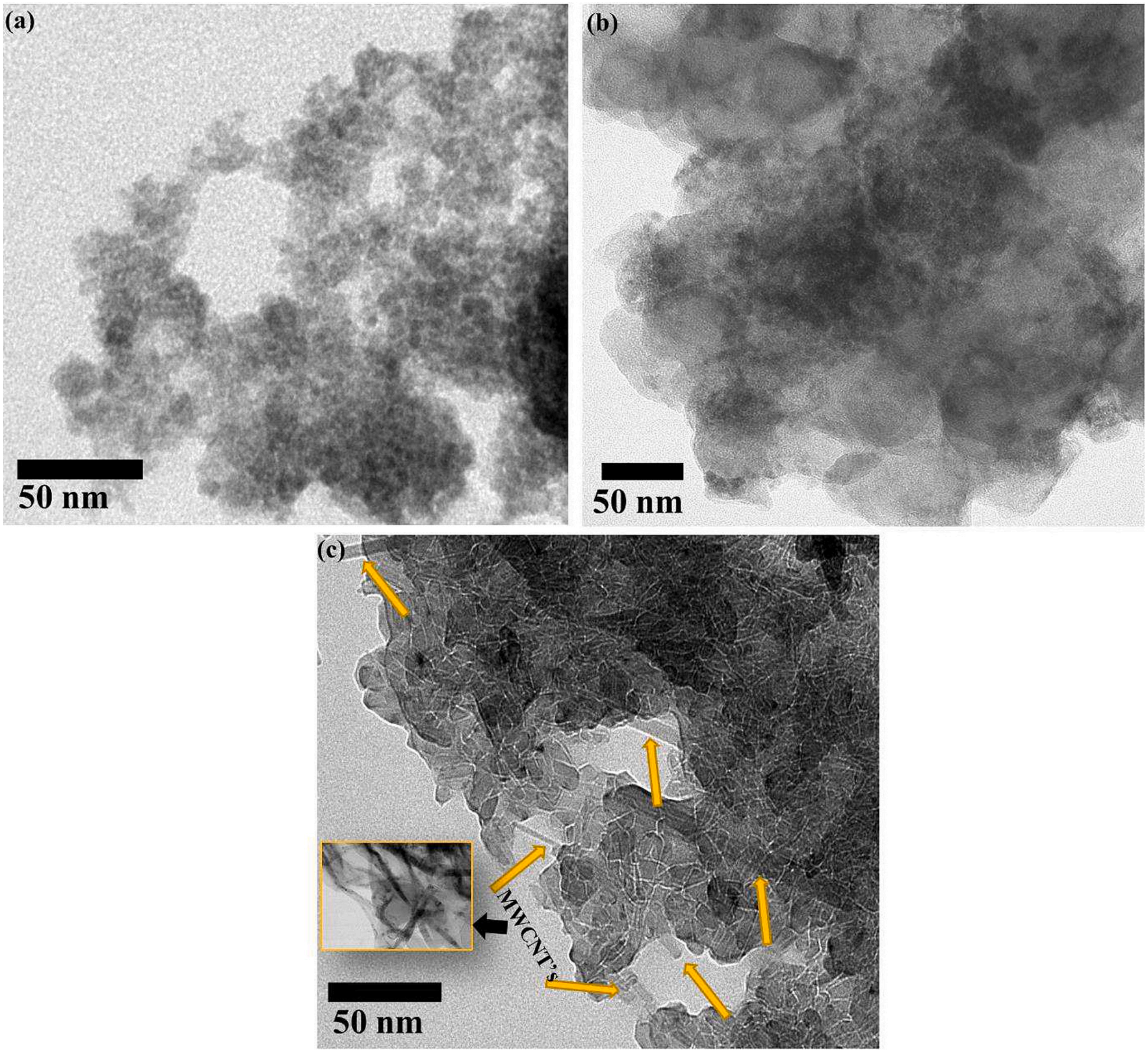


Fig. 5. The TEM images of (a) SnSe (b) SnSe/C composite (c) SnSe/MWCNT.

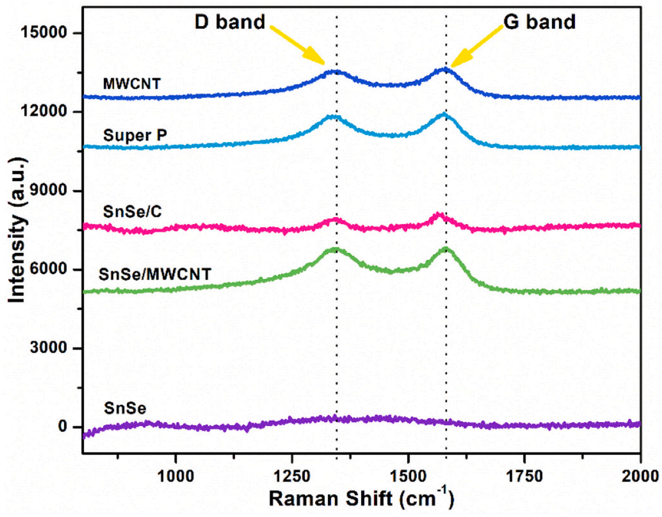


Fig. 6. Raman Spectra of MWCNT, Super P, SnSe, SnSe/C and SnSe/MWCNT.

interlayer spacing hence promoting better exchange of Li-ion in/out of the electrode material. Furthermore, the structure of MWCNTS is such that it does not deform even after multiple charge-discharge, making it as an excellent anode. As can be seen from Fig. 3(d-e), the SnSe/MWCNT composite consist of SnSe particles well surrounded by clusters and bundles of MWCNT and some of the SnSe particles are well covered by MWCNT. This composite structure may help to reduce the damage of SnSe structure during cycling giving excellent electrochemical performance as compared to bare SnSe [15].

Fig. 4 (a-e) depicts the elemental mapping of SnSe/C, images of the EDS mapping proves that SnSe and Super P was homogeneously mixed in the sample. Fig. 4(f-j) shows elemental mapping of SnSe/MWCNT confirming the uniform distribution of Sn, Se and MWCNT throughout the prepared sample. The detailed morphology of SnSe, SnSe/C and SnSe/MWCNT was further investigated using TEM and is shown in Fig. 5. As shown in Fig. 5(a), SnSe composed of primary nanoparticles of size 4–10 nm, whereas the SnSe/C composite consist of agglomerated nanoparticles in the range of 2–6 nm which are well surrounded by clusters of carbon which acts as a buffer and can help in preventing huge volume change during charging and discharging. The TEM images of SnSe/MWCNT as shown in Fig. 5(c), consist of well dispersed MWCNTs in the composite.

To further confirm the presence of carbon and MWCNT in the

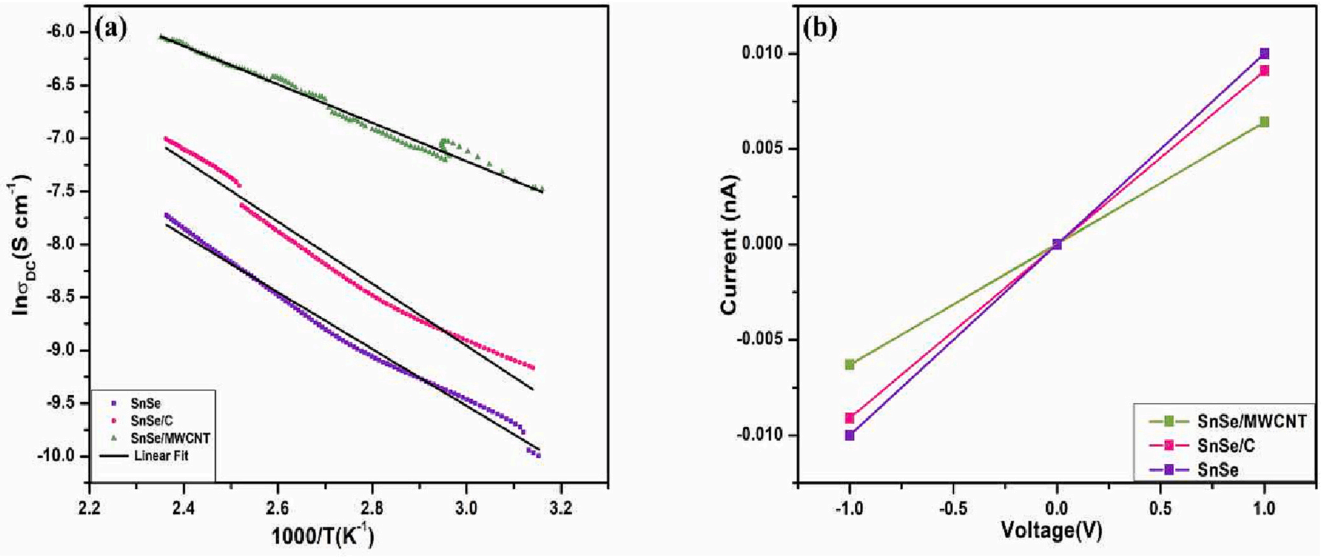


Fig. 7. (a) Variation of conductivity ($\ln\sigma_{DC}$) with temperature (1000/T) for SnSe, SnSe/C and SnSe/MWCNT (b) I-V curve at room temperature for SnSe, SnSe/C and SnSe/MWCNT.

Table 2
Calculated DC conductivity values of (a) SnSe (b) SnSe/C and SnSe/MWCNT.

Sample	Activation Energy (meV)	DC resistance (Ω)	DC Conductivity (S/cm)
SnSe	232	10×10^6	9.76×10^{-9}
SnSe/C	229	9.1×10^6	2.44×10^{-8}
SnSe/MWCNT	157	6.36×10^6	1.86×10^{-7}

composites of SnSe, Raman spectra for the prepared sample were carried out. Raman spectrum of SnSe, SnSe/C, SnSe/MWCNT, Super P and MWCNT are shown in Fig. 6. As can be depicted from the Fig. 6, the Raman spectra of Super P and MWCNT shows two broad peaks close to 1340 cm^{-1} and 1580 cm^{-1}. The carbon D (1340 cm^{-1}) band is due to the disordered structure or defects in carbon, whereas G band (1580 cm^{-1}) is attributed to the stretching of Carbon-Carbon bond and is found in all sp^2 carbon materials. As can be seen from the Raman spectra,. these two D and G bands are present in SnSe/C and SnSe/MWCNT composite, while these bands are absent in bare SnSe confirming the presence of carbon in SnSe/C and SnSe/MWCNT composites [16].

The effect of temperature on electrical conductivity was investigated for the synthesized samples. Fig. 7(a) illustrates an Arrhenius plot, in which ln (σ_{DC}) is plotted against the reciprocal of temperature for SnSe, SnSe/C and SnSe/MWCNT. The slope of curves yields the activation energy for the synthesized samples. The Arrhenius equation in its functional form can be expressed as:

$$\sigma_{DC} = \sigma_o exp\frac{-E_a}{K_B T} \tag{3}$$

Here, K_B, T, and E_a denotes Boltzmann constant, absolute temperature and activation energy, respectively.

The electronic conductivity, σ_{DC} was calculated at room temperature using the following equation:

$$\sigma_{DC} = \frac{1}{R}\left(\frac{L}{A}\right) \tag{4}$$

Where, R, L and A represents resistance of the sample, thickness of the pellet and the area of the pellet, respectively [17]. Fig. 7(b) shows I-V characteristic curve for SnSe, SnSe/C and SnSe/MWCNT at room temperature. The DC resistance value has been calculated using the slope of the I-V Curve. Hence, the DC conductivity has been obtained. It was found that conductivity increases with an increase in temperature, showing a semiconducting behaviour. Among the three tested materials, SnSe/MWCNT have the lowest activation energy and the best conductivity. The calculated values are shown in Table 2.

3.2. Electrochemical analysis

Cyclic Voltammetry study of SnSe, SnSe/C, SnSe/MWCNT anodes were obtained at 0.2 mV/s scan rate from 0.01 to 2.5 V for the first four cycles as depicted in Fig. 8(a-c). During the initial cathodic scan, SnSe, SnSe/C and SnSe/MWCNT show a sharp peak at around 1.02 V, 1.21 V and 1.05 V respectively. This could be due to the conversion reaction of SnSe into Sn and Li_2Se. However, in the proceeding cycles this cathodic peak is shifted to 1.26 V for SnSe/C and SnSe/MWCNT and vanishes in the case of bare SnSe indicating that this conversion reaction is irreversible in case of SnSe anode. On further discharging, another reduction peak was observed at around 0.5 V- 0.6 V. It may occur possibly due to the decomposition of electrolyte and hence forming Solid Electrolyte Interface (SEI) layer on the electrode surface. The peak around 0.06–0.21 V can be ascribed possibly due to the alloying of Sn with Li to form Li_xSn (where $x \geq 4.4$). For initial oxidation scan, the observed peaks between 0.5 V to 0.63 V can be attributed to the de-alloying process of Li_xSn. However, small peaks arising in SnSe/C and SnSe/MWCNT electrodes at 1.8 V and 2.2 V could be due to the re-oxidisation of Sn and Li_2Se to form SnSe and extraction of Li-ion. This is supplemented by the reduction peak at 2.02 V and repetition of these anodic peaks in the subsequent cycles. This implies that in composite anodes, the decomposition of SnSe is a reversible process which is occurring due to the addition of Super P and MWCNT [18]. It is well known that carbon source in the composite can decreases the polarisation promoting the better insertion/de-insertion of Li-ions in the electrode [19]. Moreover, the CV profile of SnSe/C shows suppression of Li_2Se as compared to SnSe and SnSe/MWCNT, this could indicate that at the equilibrium potential range, there is little retardation of the decomposition reaction this could be due to the amorphous carbon (Super P) used in the SnSe/C. However, there are only limited reports related to this mechanism [20]. As can be seen from CV graphs, SnSe/MWCNT shows the best repeatability as compared to SnSe and SnSe/C.

The Lithiation and de-lithiation reaction mechanism of SnSe, SnSe/C and SnSe/MWCNT are as follows:

Lithiation process for SnSe, SnSe/C and SnSe/MWCNT

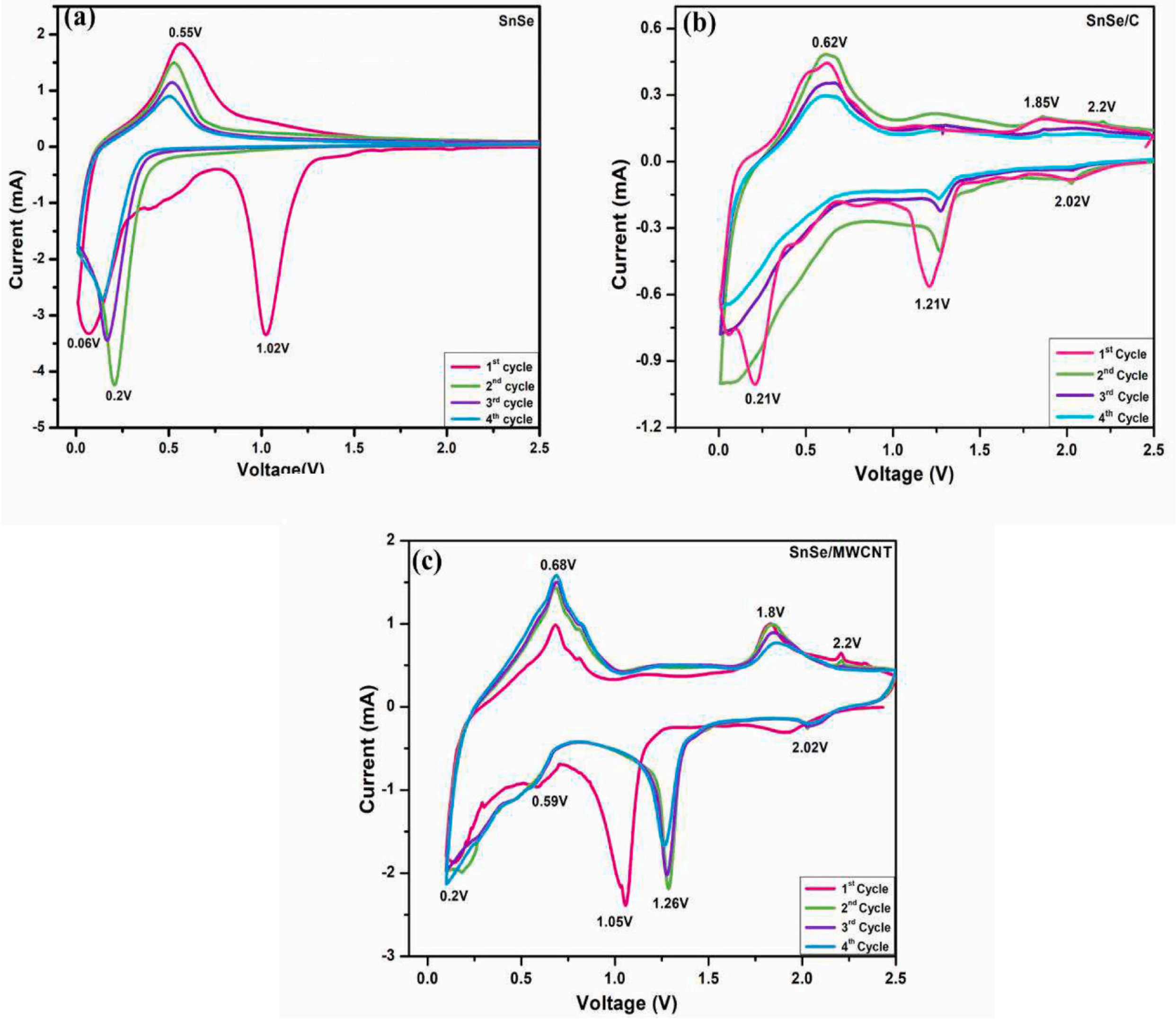


Fig. 8. Cyclic voltammograms (CV) curves of (a) SnSe (b) SnSe/C (c) SnSe/MWCNT at 0.2 mV/s scan rate from 0.01 to 2.5 V.

$$SnSe + 2Li^+ + 2e^- \rightarrow Sn + Li_2Se \tag{5}$$

$$Sn + xLi^+ + xe^- + Li_2Se \rightarrow Li_xSn + Li_2Se \ (0 \le x \le 4.4) \tag{6}$$

De-lithiation for SnSe,

$$Li_xSn + Li_2Se \rightarrow Sn + Li_2Se + xLi^+ + xe^- \tag{7}$$

De-lithiation for SnSe/C and SnSe/MWCNT,

$$Li_xSn + Li_2Se \rightarrow Sn + Li_2Se + xLi^+ + xe^- \tag{8}$$

$$Sn + Li_2Se \rightarrow Li_{2-y}SnSe + yLi^+ + ye^- \tag{9}$$

Fig. 9 (a-c) shows CV data of SnSe, SnSe/C and (c) SnSe/MWCNT, respectively, at various scan rates of 0.05 mV/s, 0.1 mV/s and 0.2 mV/s. This test is used to study the dynamic behaviour, and to calculate the Li-ion diffusion coefficient of SnSe, SnSe/C and SnSe/MWCNT electrodes. With increasing scan rate, the peak intensities of cathodic and anodic reaction also increase [21].

For a diffusion-controlled process, the relation between the square root of scan rate ($\upsilon^{0.5}$) and peak current density (I_P) should be a linear one as shown in Fig. 9(d-f). The Randles-Sevcik eq. (10) is used to determine the relationship and the value of chemical diffusion coefficient [22].

$$I_P = 0.4463n^{3/2}F^{3/2}C_{Li^+} + AR^{-1}T^{-1}D_{(Li^+)^{1/2}}v^{1/2} \tag{10}$$

where I_P, n C_{Li^+}, A, F, R, T, ʋ and $D_{(Li)^+}$ represents peak current density, charge transfer number, Li-ion concentration in electrodes, electrode area, Faraday's Constant, Gas Constant, absolute temperature, scan rate and diffusion coefficient, respectively [21].

Diffusion coefficient was calculated around cathodic and anodic peaks for SnSe, SnSe/C and SnSe/MWCNT. All the calculated values of diffusion coefficient are given in Table 3. From the calculated values, it can be seen SnSe/MWCNT shows the higher diffusion coefficient value as compared to SnSe/C and SnSe implying the best diffusion properties.

In order to investigate kinetic behaviours of bare SnSe, SnSe/C and SnSe/MWCNT, EIS was performed before cycling. EIS measurement is used to examine the prepared electrodes materials since the internal impedance of a cell is a vital characteristic that has a direct effect on its electrochemical performance. Fig. 10(a) shows the Nyquist plots of SnSe, SnSe/C and SnSe/MWCNT at Open Circuit Voltage (OCV) tested from 10 mHz to 100 kHz frequency. Each spectrum at high to medium frequencies consists of a semicircle and a straight inclined line at lower frequencies. EIS analysis has been performed using equivalent circuits with R, C and Q combination to fit the impedance data. The equivalent circuit is shown in Fig. 10(inset), where R_s and R_{ct} denotes the resistance due electrolyte and electrode-electrolyte interface respectively, R_{SEI} and

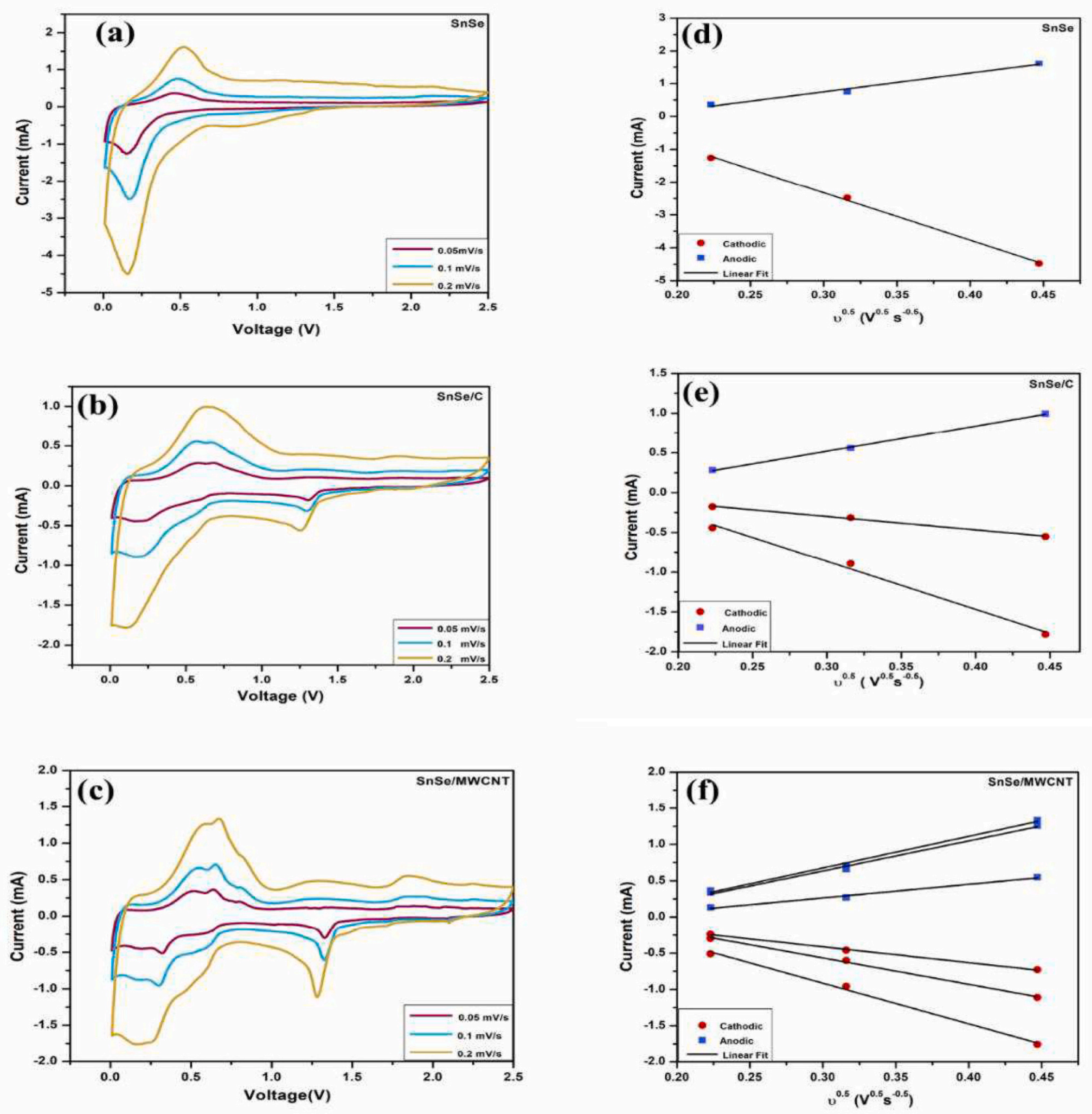


Fig. 9. Cyclic Voltammetry at various scan rate of (a) SnSe (b) SnSe/C and (c) SnSe/MWCNT, cathodic and anodic peak current versus $\upsilon^{0.5}_{,}$ (d) SnSe (e)SnSe/C and (f) SnSe/MWCNT.

Table 3
Calculated $D_{(Li)^+}$ values of (a) SnSe (b)SnSe/C (c) SnSe/MWCNT from cyclic voltammetry.

SnSe	Anodic peak			Cathodic Peak		
$D_{(Li)^+}$ (cm^2s^{-1})	9.67×10^{-15}			2.44×10^{-15}		
SnSe/C	Anodic Peak			Cathodic Peak		
$D_{(Li)^+}$ (cm^2s^{-1})	3×10^{-15}			0.83×10^{-15}	1.08×10^{-14}	
SnSe/ MWCNT	Anodic Peak			Cathodic Peak		
$D_{(Li)^+}$(cm^2s^{-1})	1.05×10^{-15}	5.65×10^{-14}	5.11×10^{-14}	3.9×10^{-15}	1×10^{-14}	1.4×10^{-14}

C_{SEI} may arises due to the transfer of Li-ions through SEI layer, C_{dl} is the double layer capacitance and Z_W corresponds to the Warburg impedance and it corresponds to the tail at low frequency [14,22,23] From EIS graphs, it can be seen that the SnSe/MWCNT has the smallest diameter of the semicircle as compared to SnSe/C and bare SnSe revealing that the charge transfer resistance (R_{ct}) for SnSe/MWCNT composite is much smaller than the SnSe, whereas the R_{ct} value of SnSe/C lies in between SnSe/MWCNT and SnSe. For SnSe/C composite, obtained R_{ct} value could be due to the conducting nature and dispersing effect of carbon, whereas for SnSe/MWCNT there is the easy transfer of Li-ions in the charge transfer process due to the tubular structure of MWCNT which also improved the electrochemical performance of SnSe/MWCNT.

Fig. 10(b) shows the relationship plot between Z' and $\omega^{-0.5}$. The fitting results are shown in Table 4.

The diffusion coefficient in the bulk material can be calculated using

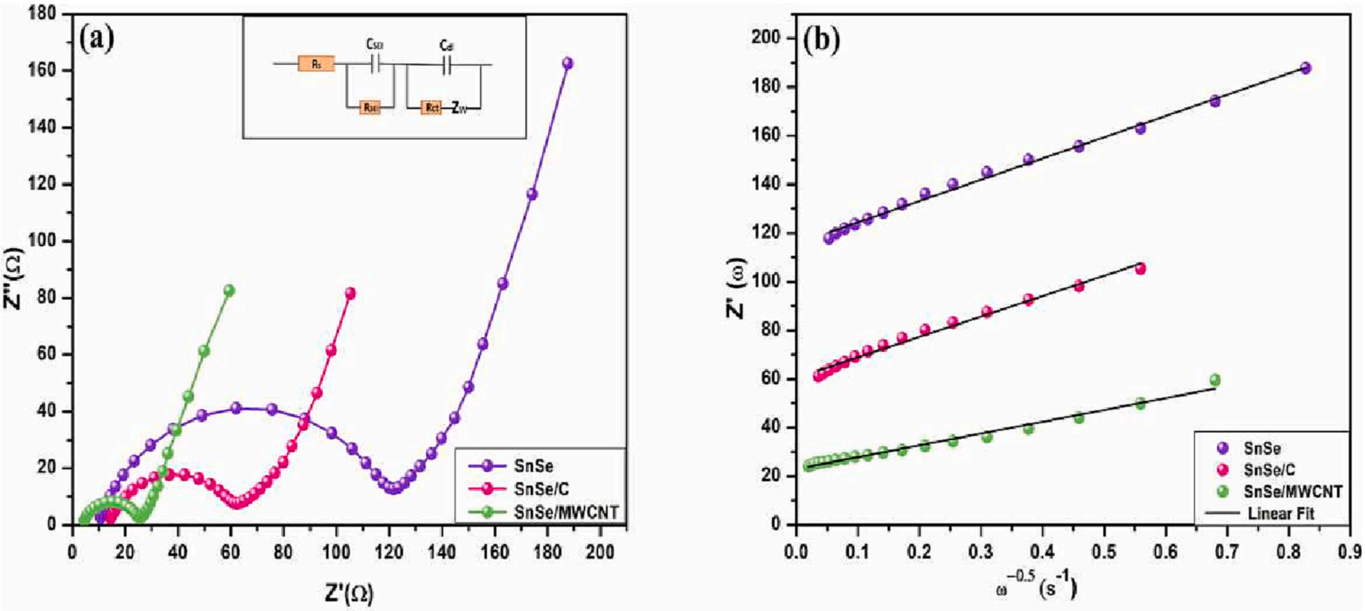


Fig. 10. (a) Nyquist plots of SnSe, SnSe/C and SnSe/MWCNT (b) relationship between real part of resistance (Z') and inverse square root of angular frequency ($\omega^{-0.5}$).

Table 4
Calculated D_{Li^+}values of (a) SnSe (b) SnSe/C (c) SnSe/MWCNT from EIS results.

Parameter	SnSe	SnSe/C	SnSe/MWCNT
R_s (Ω)	12	15	4
R_{ct} (Ω)	110	43	19
$D_{Li^+}(cm^2s^{-1})$	2×10^{-15}	2.19×10^{-15}	6.55×10^{-15}

the given equation:

$$D_{Li^+} = \frac{0.5R^2T^2}{A^2n^2F^2C^2\sigma_\omega{}^2} \quad (11)$$

where D_{Li^+}, R, A, n, T, C, F and σ_ω is the Diffusion Coefficient (cm^2s^{-1}), Boltzmann Constant(8.314 J mol^{-1} K^{-1}), Electrode Area (cm^2), no. of electrons, absolute temperature (K), lithium ion concentration (mol $/cm^3$), Faraday's Constant and Warburg Factor ($\Omega\ s^{-1/2}$), respectively. The relation of Z' and σ_ω is given below:

$$Z' = R_s + R_{ct} + \sigma_\omega \omega^{-1/2} \quad (12)$$

Eq. (12) has been used to calculate the Warburg factor, σ_ω by using the slope of Bode plot. By substituting the values in eq. (11), diffusion coefficient was calculated [24,25]. The calculated values of D_{Li^+} are listed in Table 4. The estimated values of D_{Li^+} from EIS are in accordance with the values calculated from cyclic voltammetry.

Fig. 11(a) displays initial galvanostatic charge-discharge profiles of SnSe, SnSe/C and SnSe/MWCNT in the range 0.01 to 2.5 V (current density = 50 mA/g). During initial discharging, a common plateau was observed around 1.25 V in all three electrodes attributing to the decomposition of SnSe into Li_2Se and Sn followed by a slope at ~0.6 V arising due to the development of the SEI layer and slope around 0.3 V was due to the alloying reaction of metallic Sn with Li. Whereas, during charging a slope at ~0.5 V was noted in all three electrodes attributing to the de-alloying reaction of Li_2Sn. However, another slope is observed at around 1.9 V arising due to the oxidation process of Sn and Li_2Se to

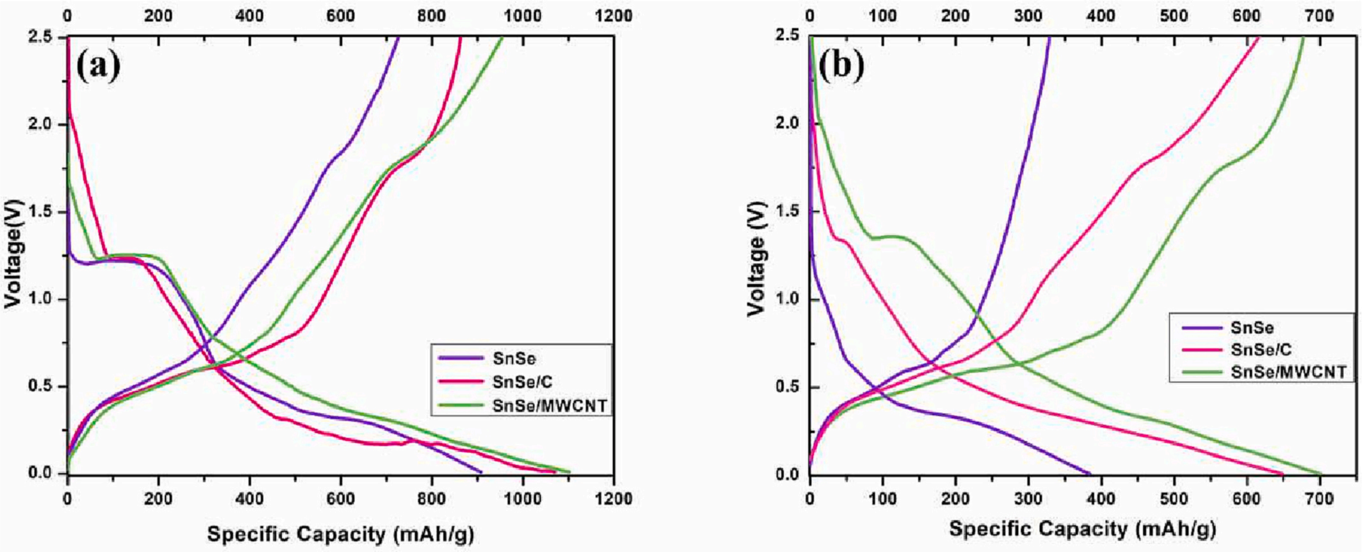


Fig. 11. Galvanostatic charge-discharge profiles of SnSe, SnSe/C and SnSe/MWCNT for (a) 1st cycle (b) 10th cycle.

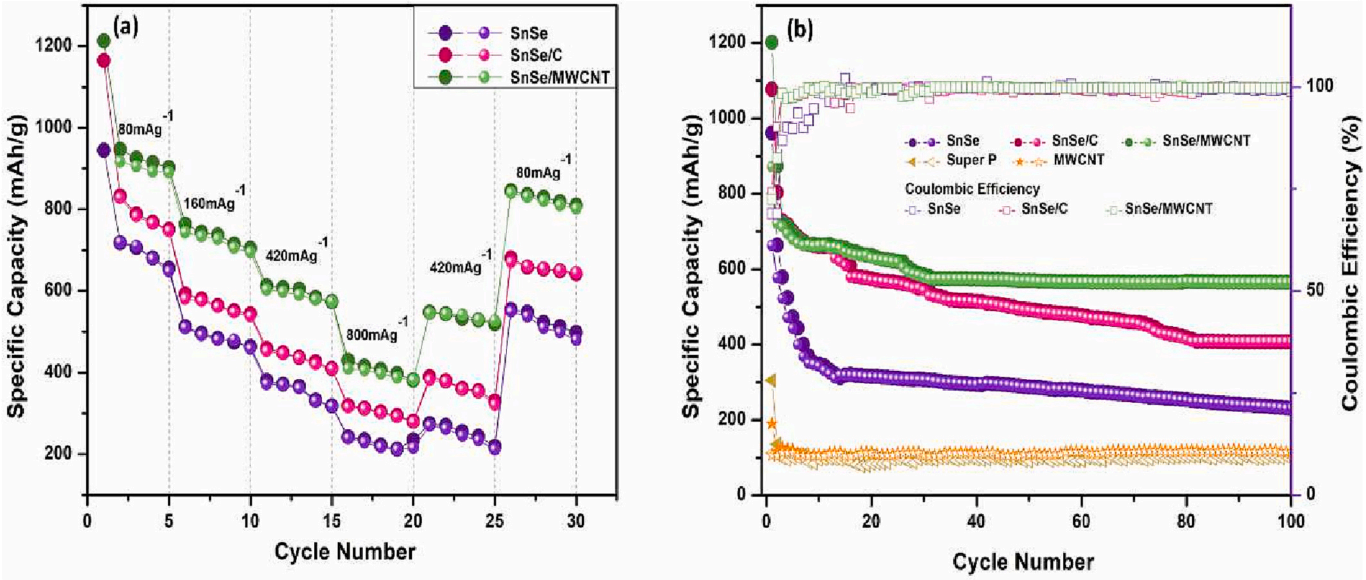


Fig. 12. (a) Rate performance (b) Cycling performance of SnSe, SnSe/C and SnSe/MWCNT (Symbols: ◉ represents charge and ● represents discharge).

Table 5
Specific discharge capacity of SnSe, SnSe/C and SnSe/MWCNT at various current density.

Specific Discharge Capacity (mAh/g)			
Current Density mAg^{-1}	SnSe	SnSe/C	SnSe/MWCNT
80	557	648	799
160	462	537	697
420	315	409	575
800	213	282	381

form SnSe. After initial discharge cycle, the plateau at 1.25 V vanished in SnSe but not in SnSe/C and SnSe/MWCNT indicating that the decomposition of SnSe is reversible in SnSe/C and SnSe/MWCNT. The SnSe/MWCNT electrode showed highest initial discharging-charging capacity of 1107mAh/g-955 mAh/g (coulombic efficiency of 86%), whereas SnSe/C and SnSe electrode shows an initial discharging-charging capacity of 1069 mAh/g – 865 mAh/g (coulombic efficiency of 80%) and 910mAh/g-724 mAh/g (coulombic efficiency of 79%), respectively. Fig. 11(b) shows the charge discharge curve of 10th cycle. SnSe discharge and charge capacity after 10 cycles were recorded to be 384mAh/g and 328mAh/g, whereas SnSe/C and SnSe/MWCNT shows much higher values of charge-discharge of 617mAh/g-647mAh/g and 678mAh/g-700mAh/g respectively even after 10 cycles. The higher capacity of SnSe/C and SnSe/MWCNT electrodes could be due to the lithiation/de-lithiation of Super P and MWCNT and reversible decomposition of SnSe in the composite.

Fig. 12(a) shows the rate performance of SnSe, SnSe/C and SnSe/MWCNT electrodes at different C-rates and the obtained discharge capacities are tabulated in Table 5. It can be observed, SnSe/MWCNT demonstrate best rate capability among all three electrodes at each current density. Fig. 12(b) shows the comparison of cycling behaviour of SnSe, SnSe/C, SnSe/MWCNT, Super P and MWCNT half cells at 500 mA/g. The SnSe electrode shows a rapid decay in capacity maintaining a capacity of 229mAh/g after 100 cycles. The reason of poor performance could be due to the structural and volumetric change during lithiation/de-lithiation of SnSe. Whereas, SnSe/C and SnSe/MWCNT shows better cycling performance maintaining capacity of 405mAh/g and 564mAh/g even after 100 cycles, respectively. The reason of improved cycling behaviour of SnSe/C could be attributed to the good dispersion of SnSe within Super P and the conducting, buffering effect of Super P. Whereas, the reason for better cycling behaviour of SnSe/MWCNT could be attributed to the tubular structure of MWCNTs which can alleviate volume change during cycling and reduce the pulverization of SnSe particles, thus preserving the connectivity of SnSe particles [18].

Moreover, the structure of MWCNTs is such that it decreases the Li-ion diffusion length which leads to enough electrode and electrolyte interphase to absorb Li-ions.

To further investigate the cycling effect on the morphologies of the electrode, FESEM was carried out after cycling. Fig. 13(a-f) shows the FESEM images of SnSe, SnSe/C and SnSe/MWCNT after 100 cycles. As can been seen from the Fig. 13(a-b), there are major cracks visible in the SnSe electrode and SnSe particles have been completely destroyed after 100 cycles. For SnSe/C electrode after 100 cycles, there are few cracks appeared on the SnSe/C electrode and particle morphology have not changed much, whereas SnSe/MWCNT electrode shows no cracking and remains integral though the presence of MWCNT is not visible in the image but there is a presence of a very thin gel like layer on the SnSe/MWCNT electrode after cycling which could be related to the increased capacity [26].

Another method, Galvanostatic Intermittent Titration Technique (GITT) has been studied to understand Li-ion insertion and extraction kinetics of electrodes and hence used to determine Li-ion diffusion coefficient [27]. Fig. 14 (a-c) shows GITT curves of SnSe, SnSe/C and SnSe/MWCNT between 0.01 and 2.5 V. For titration process, the cells were first discharged for 10 min at a current density of $40mAg^{-1}$ and were put on rest for 30 min. The obtained GITT curves shows plateaus which is in accordance with obtained CV curves for the same electrode. The Li-ion diffusion coefficient is calculated using the given equation which is based on Fick's second law [28]:

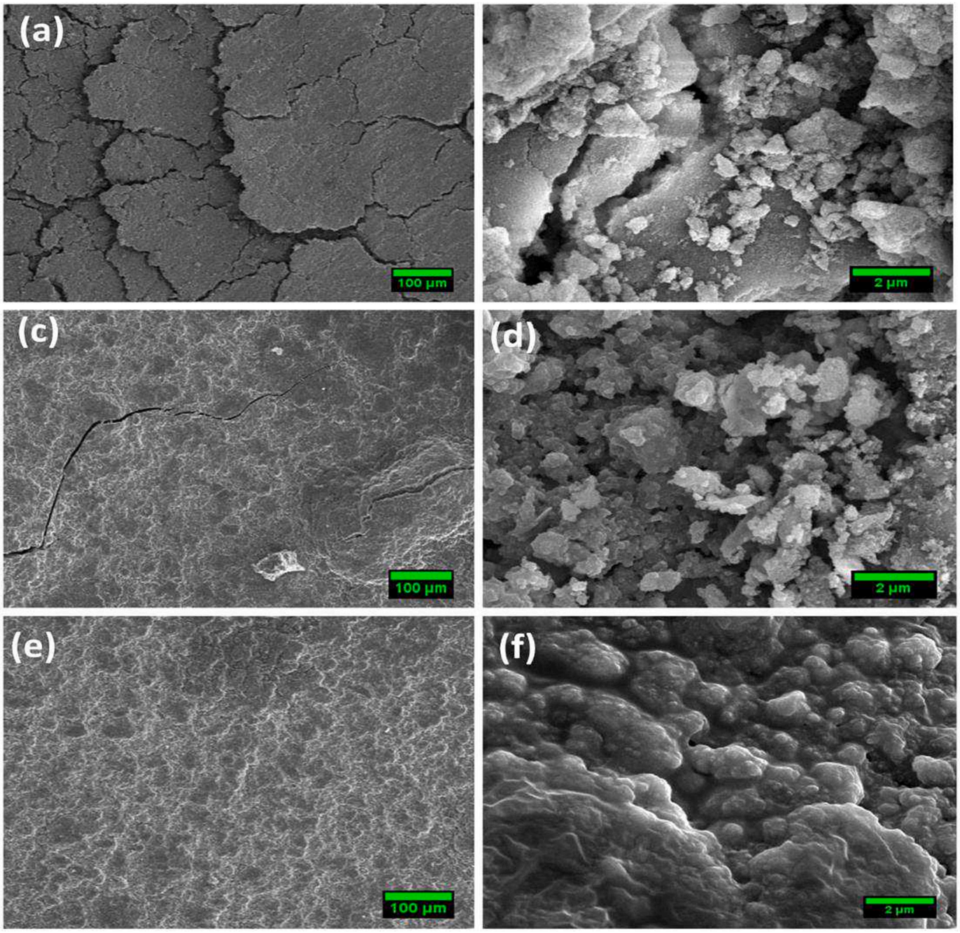


Fig. 13. FESEM micrographs of (a-b) SnSe, (c-d) SnSe/C, (e-f) SnSe/MWCNT after 100 cycles.

$$D_{Li^+} = \frac{4}{\pi}\left(\frac{m_B V_M}{M_B A}\right)^2 \left(\frac{\Delta E_S}{\tau\left(\frac{dE_\tau}{d\sqrt{\tau}}\right)}\right)^2 \left(\tau \ll \frac{L^2}{D_{Li^+}}\right) \tag{13}$$

where τ, L, A, m_B, V_M and M_B is the duration in which electrode stays at rest, thickness of the electrode, surface area of the electrode, mass, volume and molecular weight of the active material, respectively. When the change in cell potential exhibit a linear relationship when plotted against square root of duration time (τ) as shown in Fig. 14(g-i), then eq. (13) can be re-written as:

$$D_{Li^+} = \frac{4}{\pi\tau}\left(\frac{m_B V_M}{M_B A}\right)^2 \left(\frac{\Delta E_S}{\Delta E_\tau}\right)^2 \left(\tau \ll \frac{L^2}{D_{Li^+}}\right) \tag{14}$$

where, ΔE_S represents the steady-state voltage change during the current pulse, after eliminating the iR drop occurring from the electrical internal resistance and ΔE_τ is the change in potential for the charge/discharge during the constant current pulse, and can be directly acquired from GITT curves [22,29].

The Li-ion diffusion coefficients are calculated from the GITT curve at all steps for discharging except for 1st discharge due to large variation and plot of log (D_{Li^+}) as a function of Voltage (V) as shown in Fig. 14(d-f). As can be seen from Fig. 14(d-f), the three electrodes showed three minimum Li-ion diffusion coefficient points at voltage shown in graph. These obtained minimums are due to a phase transition from strong attractive interaction between the host matrix and the intercalation species. Compared to the Cyclic Voltammetry, these potentials are the redox potentials. The lithium-ion diffusion coefficient from these three points for SnSe, SnSe/C and SnSe/MWCNT are summarized in Table 6. The obtained lithium diffusion coefficient may not be completely accurate considering the calculation error in dE/dx and deviation in parameter V_M. Therefore, the calculated D_{Li^+} values showed slight variation when compared to the results obtained from EIS and CV.

4. Conclusion

Synthesis of SnSe, SnSe/C and SnSe/MWCNT composites have been successfully carried out via high energy ball milling route as an anode material for Li-ion batteries. SnSe/C and SnSe/MWCNT composite exhibited better reversible capacity, cycle life and rate performance than SnSe. The initial cycle discharge capacity of SnSe/C and SnSe/MWCNT has been observed as 1069 mAh/g and 1107mAh/g, respectively. However, after 100 cycles, the specific capacity of SnSe/C and SnSe/MWCNT electrodes is maintained at 385 mAh/g (capacity retention of 36%) and 495 mAh/g (capacity retention of 43.8%). The enhanced electrochemical performance of SnSe/C and SnSe/MWCNT was assigned to the good dispersion of SnSe within Super P and the conducting, buffering effect of Super P and to the tubular structure of MWCNTs which can alleviate volume change during cycling and reduce the pulverization of SnSe particles. Thus, preserving the connectivity of SnSe particles and also, due to the reversible decomposition of SnSe and lithiation/de-lithiation of Super P and MWCNT. CV, EIS and GITT measurements were used to determine Li-ion diffusion coefficient of prepared electrodes. The Li + diffusion coefficient from EIS calculations for SnSe, SnSe/C and SnSe/MWCNT were in the range of 10^{-15} cm^{2}s^{-1}. The Li + diffusion coefficient from CV calculations was found to be in the range of 10^{-14} to 10^{-15} cm^{2}s^{-1}and from GITT calculations it was

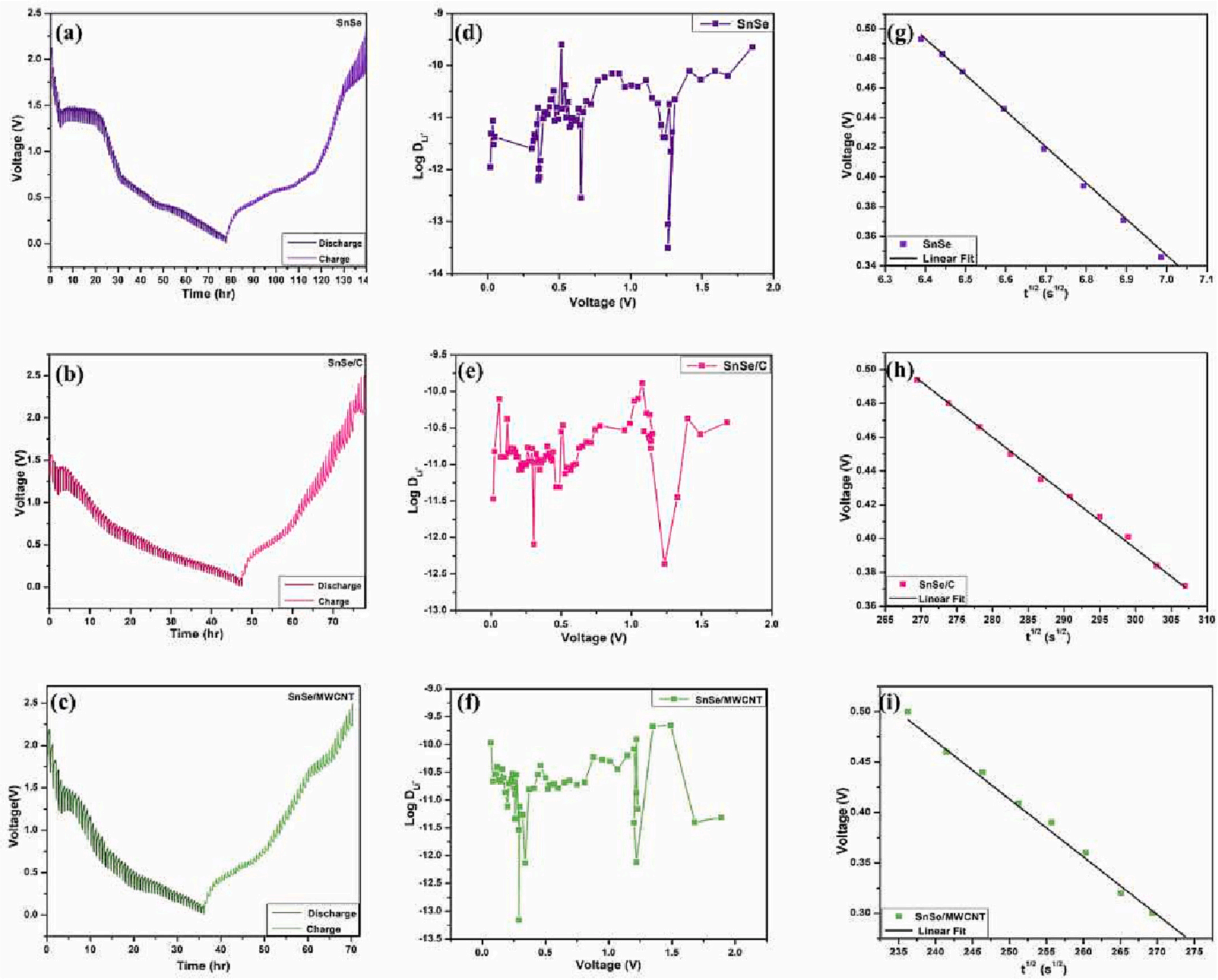


Fig. 14. Discharge/charge GITT curve of (a)SnSe (d) SnSe/C and (g) SnSe/MWCNT, lithium diffusion coefficient as a function of voltage for (b) SnSe (e) SnSe/C and (h) SnSe/MWCNT, the plot of V vs $\tau^{0.5}$ for (c) SnSe (f) SnSe/C and (i) SnSe/MWCNT.

Table 6
Calculated D_{Li^+}values of (a) SnSe (b) SnSe/C (c) SnSe/MWCNT from GITT.

Potential (V) ↓	$D_{Li^+}(cm^2s^{-1})$		
	SnSe	SnSe/C	SnSe/MWCNT
1.68	–	–	3.98×10^{-12}
1.2–1.3	3.12×10^{-14}	4.73×10^{-13}	7.57×10^{-13}
0.652	2.82×10^{-13}	–	–
0.3–0.5	7.3×10^{-14}	3.44×10^{-13}	7.33×10^{-13}
0.28	–	–	2.9×10^{-12}

found to be in the range of 10^{-12} to 10^{-14} cm^2s^{-1}. The Li + diffusion coefficient value obtained from GITT is higher than the values obtained from CV and EIS. This difference in value is due to the different equilibrium conditions. The CV and EIS were performed under more equilibrium conditions than GITT and electrode has more relaxation time. Among bare SnSe, SnSe/C and SnSe/MWCNT samples, SnSe/MWCNT showed higher values of Li-ion diffusion coefficient, indicating that the SnSe/MWCNT electrode has superior kinetics over SnSe/C and SnSe. Hence, SnSe/MWCNT may be good alternate as a probable anode material for lithium-ion batteries.

Authors statement

The corresponding Author of the manuscript entitled: *'Study of Lithium Diffusion Properties and Electrochemical Performance of SnSe/C and SnSe/MWCNT Composite Anode for Li-ion Batteries'* by Shivangi Rajput, Amrish K. Panwar and Amit Gupta, certify that the contributors' and conflicts of Interest statements included in this paper are correct and have approved by all co-authors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

The authors gratefully acknowledge Delhi Technological University, Delhi for the financial support through project Grant No.: F. No. DTU/IRD/619/2019/2114 to carry out this research work.

References

[1] E.K. Heidari, A. Kamyabi-gol, M.H. Sohi, A. Ataie, Electrode Materials for Lithium Ion Batteries : A Review 51, 2018, pp. 1–12, https://doi.org/10.22059/JUFGNSM.2018.01.01.

[2] A. Mishra, A. Mehta, S. Basu, S.J. Malode, N.P. Shetti, S.S. Shukla, M. N. Nadagouda, T.M. Aminabhavi, Electrode materials for lithium-ion batteries, Mater. Sci. Energy Technol. 1 (2018) 182–187, https://doi.org/10.1016/j.mset.2018.08.001.

[3] W. Qi, J.G. Shapter, Q. Wu, T. Yin, G. Gao, D. Cui, Nanostructured anode materials for lithium-ion batteries: principle, recent progress and future perspectives, J. Mater. Chem. A 5 (2017) 19521–19540, https://doi.org/10.1039/c7ta05283a.

[4] B. Scrosati, J. Garche, Lithium batteries: status, prospects and future, J. Power Sources 195 (2010) 2419–2430, https://doi.org/10.1016/j.jpowsour.2009.11.048.

[5] D. Deng, Li-ion batteries: basics, progress, and challenges, Energy Sci. Eng. 3 (2015) 385–418, https://doi.org/10.1002/ese3.95.

[6] W. Luo, J.J. Gaumet, L.Q. Mai, Antimony-based intermetallic compounds for lithium-ion and sodium-ion batteries: synthesis, construction and application, Rare Metals 36 (2017) 321–338, https://doi.org/10.1007/s12598-017-0899-4.

[7] M.G. Park, D.H. Lee, H. Jung, J.H. Choi, C.M. Park, Sn-based nanocomposite for Li-ion battery anode with high energy density, rate capability, and reversibility, ACS Nano 12 (2018) 2955–2967, https://doi.org/10.1021/acsnano.8b00586.

[8] R. Zhang, S. Upreti, M. Stanley Whittingham, Tin-iron based nano-materials as anodes for li-ion batteries, J. Electrochem. Soc. 158 (2011) A1498, https://doi.org/10.1149/2.108112jes.

[9] G. Xiao, Y. Wang, J. Ning, Y. Wei, B. Liu, W.W. Yu, G. Zou, B. Zou, Recent advances in IV-VI semiconductor nanocrystals: Synthesis, mechanism, and applications, RSC Adv. 3 (2013) 8104–8130, https://doi.org/10.1039/c3ra23209c.

[10] H. Tian, F. Xin, X. Wang, W. He, W. Han, High capacity group-IV elements (Si, Ge, Sn) based anodes for lithium-ion batteries, J. Mater. 1 (2015) 153–169, https://doi.org/10.1016/j.jmat.2015.06.002.

[11] L. Liu, F. Xie, J. Lyu, T. Zhao, T. Li, B.G. Choi, Tin-based anode materials with well-designed architectures for next-generation lithium-ion batteries, J. Power Sources 321 (2016) 11–35, https://doi.org/10.1016/j.jpowsour.2016.04.105.

[12] A. Mauger, H. Xie, C.M. Julien, Composite anodes for lithium-ion batteries: status and trends, AIMS Mater. Sci. 3 (2016) 1054–1106, https://doi.org/10.3934/matersci.2016.3.1054.

[13] Y. Xu, Q. Liu, Y. Zhu, Y. Liu, A. Langrock, M.R. Zachariah, C. Wang, Uniform nano-Sn/C composite anodes for lithium ion batteries, Nano Lett. 13 (2013) 470–474, https://doi.org/10.1021/nl303823k.

[14] L. Yu, L. Qin, X. Xu, K. Kim, J. Liu, J. Kang, K. Ho Kim, $SnSe_x$ (x=1,2) nanoparticles encapsulated in carbon nanospheres with reversible electrochemical behaviors for lithium-ion half/full cells, Chem. Eng. J. 431 (2022), 133463, https://doi.org/10.1016/j.cej.2021.133463.

[15] Z. Zhang, X. Zhao, J. Li, SnSe/Carbon Nanocomposite Synthesized by High Energy Ball Milling as an Anode Material for Sodium-Ion and Lithium-Ion Batteries, Electrochim. Acta 176 (2015) 1296–1301, https://doi.org/10.1016/j.electacta.2015.07.140.

[16] Y. Kim, Y. Kim, Y. Park, Y.N. Jo, Y.J. Kim, N.S. Choi, K.T. Lee, SnSe alloy as a promising anode material for Na-ion batteries, Chem. Commun. 51 (2015) 50–53, https://doi.org/10.1039/c4cc06106c.

[17] R. Saroha, A.K. Panwar, Effect of in situ pyrolysis of acetylene (C_2H_2) gas as a carbon source on the electrochemical performance of $LiFePO_4$ for rechargeable lithium-ion batteries, J. Phy. D: Appli. Phys. 50 (2017) 255501–255512, https://doi.org/10.1088/1361-6463/aa708c.

[18] A. Gurung, R. Naderi, B. Vaagensmith, G. Varnekar, Z. Zhou, H. Elbohy, Q. Qiao, Electrochimica acta tin selenide – multi-walled carbon nanotubes hybrid anodes for high performance lithium-ion batteries, Electrochim. Acta 211 (2016) 720–725, https://doi.org/10.1016/j.electacta.2016.06.065.

[19] J. Lu, B. Jia, X. Lu, Y. Guo, R. Hu, R. Khatoon, L. Jiao, L. Zhang, Two-dimensional $SnSe_2$/CNTs hybrid nanostructures as anode materials for high-performance lithium-ion batteries, Chem. Eur. J. 25 (2019) 9973–9983, https://doi.org/10.1002/chem.201901487.

[20] T. Moon, C. Kim, S.T. Hwang, B. Park, Electrochemical properties of disordered-carbon-coated SnO_2 nanoparticles for Li rechargeable batteries, Electrochem. Solid-State Lett. 9 (2006) 408–411, https://doi.org/10.1149/1.2214332.

[21] D. Lakshmi, B. Nalini, Performance of SnSb: Ce,Co alloy as anode for lithium-ion batteries, J. Solid State Electrochem. 21 (4) (2016) 1027–1034, https://doi.org/10.1007/s10008-016-3456-4.

[22] Y.S. Lee, K.S. Ryu, Study of the lithium diffusion properties and high rate performance of $TiNb_6O_{17}$ as an anode in lithium secondary battery, Sci. Rep. 7 (2017) 1–14, https://doi.org/10.1038/s41598-017-16711-9.

[23] Q. Yu, B. Wang, J. Wang, S. Hu, J. Hu, Y. Li, Flowerlike Tin Diselenide Hexagonal Nanosheets for High-Performance Lithium-Ion Batteries, F. Chem. 8 (2020) 1–6, https://doi.org/10.3389/fchem.2020.00590.

[24] Y. Cheng, J. Huang, L. Cao, Y. Wang, Y. Ma, S. Xi. Investigation of Electrochemical Performance on $SnSe_2$ and SnSe Nanocrystals as Anodes for Lithium Ions Batteries, Nano 14, 2019, pp. 1–8, https://doi.org/10.1142/S1793292019501558.

[25] R. Saroha, A. Gupta, A.K. Panwar, Electrochemical performances of Li-rich layered-layered Li_2MnO_3 - $LiMnO_2$ solid solutions as cathode material for lithium-ion batteries, J. Alloys Compd. 696 (2017) 580–589, https://doi.org/10.1016/j.jallcom.2016.11.199.

[26] X. Cao, A. Li, Y. Yang, J. Chen, ZnSe nanoparticles dispersed in reduced graphene oxides with enhanced electrochemical properties in lithium/sodium ion batteries, RSC Adv. 8 (2018) 25734–25744, https://doi.org/10.1039/c8ra03479f.

[27] N. Ding, J. Xu, Y.X. Yao, G. Wegner, X. Fang, C.H. Chen, I. Lieberwirth, Determination of the diffusion coefficient of lithium ions in nano-Si, Solid State Ionics 180 (2009) 222–225, https://doi.org/10.1016/j.ssi.2008.12.015.

[28] W. Weppner, R.A. Huggins, Determination of the kinetic parameters of mixed-conducting electrodes and application to the system Li_3Sb, J. Electrochem. Soc. 124 (1977) 1569–1578, https://doi.org/10.1149/1.2133112.

[29] N. Böckenfeld, A. Balducci, Determination of sodium ion diffusion coefficients in sodium vanadium phosphate, J. Solid State Electrochem. 18 (2014) 959–964, https://doi.org/10.1007/s10008-013-2342-6.

Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

p-ISSN: 0972-6268
(Print copies up to 2016)
e-ISSN: 2395-3454

Vol. 22 | No. 1 | pp. 477-487 | 2023

Original Research Paper

https://doi.org/10.46488/NEPT.2023.v22i01.046

Open Access Journal

Sustainable Green Approach of Silica Nanoparticle Synthesis Using an Agro-waste Rice Husk

Mikhlesh Kumari*, Kulbir Singh*(**), Paramjeet Dhull*, Rajesh Kumar Lohchab*† and A. K. Haritash***
*Department of Environmental Science & Engineering, Guru Jambheshwar University of Science & Technology, Hisar, Haryana, India
**Department of Civil Engineering, M.M. Engineering College, Maharishi Markandeshwar (Deemed to be University), Mullana-Ambala-133207, India
***Department of Environmental Engineering, Delhi Technological University, Shahbad, Daulatpur, Delhi-110042, India
†Corresponding author: Rajesh Kumar Lohchab; rajeshlohchab@gmail.com

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 10-08-2022
Revised: 03-10-2022
Accepted: 04-10-2022

Key Words:
Silica nanoparticle
Biogenic
Agricultural waste
Rice husk ash

ABSTRACT

Agro-waste can provide a non-metallic, environmentally friendly bio-precursor for the production of green silica nanoparticles. To manufacture silica nanoparticles from rice husk, biogenic silica nanoparticles were generated using an alkaline precipitation approach. Rice husk as a source of silica nanoparticles is environmentally and economically valuable because it is a plentiful lower price agricultural derivative that can be used to help with waste management. During the synthesis process, the dose of rice husk ash was used at 5 g at pH 7, alkali dose concentration of 0.5 M, reaction period of 3.5 h, and temperature of 90°C that produced maximum silica nanoparticles with a yield of 88.5%. To optimize the silica nanoparticle production from rice husk ash Box Behnken Design (BBD) a subcategory of the response surface methodology (RSM) was accomplished. BBD model was successfully matched, as evidenced by the high correlation values of adjusted R^2 0.9989 and predicted R^2 0.9977. Silica nanoparticles' amorphous form generated from rice husk ash is indicated by XRD analysis 2Θ peak at 22.12° and UV-Vis Spectroscopy absorbance peak at 312 nm. The amorphous shape of silica is amorphous and crystalline defined through XRD. nanoparticles generated from rice husk ash is indicated by FESEM analysis and EDX analysis, confirming that the SiO_2 elemental configuration comprises the highest concentration of Si and O. The existence of a siloxane group in the produced compound was revealed by FTIR spectra stretching vibrations at 803.69 and 1089.05 cm^{-1}

INTRODUCTION

Rice is among the greatest crops cultivated around the world since it is one of the most important food varieties and a supplier of nourishment for individuals. In contrast to 2020, the estimated global rice production has increased from 513.2 million tons to 519.7 million tons in 2021 (FAO 2021). As a result of generation and refining in the industry of agriculture, rice husk accounts for the major residue. Due to its characteristics like lignin, cellulose, hemicellulose content, hard surface, a small quantity of proteins, and high silicon quantity, rice husk can't be bacterial decomposed easily and is also insoluble in water (Soltani et al. 2015). In many countries, mostly rice husk has pointlessly burned and it is responsible for the air pollution problem. Additionally, composting of rice husk generates a huge quantity of methane. Rice husk contains approximately 20% of rice in weight ratio and incineration of these rice husks under 500-700°C of controlled temperature can generate ash and also amorphous silica (Kang et al. 2019). As compared to other crops, silica act as a unique crop residue in rice (Setiawan & Chiang 2021). The silica-separated rice husk is eco-accommodating as a result of its procurement from natural items and affordable because of the low-cost raw substance value.

The development of silica dioxide nanoparticles has drawn gigantic consideration in the world of science and technology in light of their broad use in different fields like biomedical fields, drug delivery systems, pesticides degradation, thermal insulators, humidity sensors, and electronic devices (Bharti et al. 2015, Bapat et al. 2016, Nazeran & Moghaddas 2017, Chong et al. 2018, Kano et al. 2019). According to previous studies, SiO_2 nanoparticles have been synthesized using natural resources like rice husk, marine sponges and diatom, coal fly ash, sand, and sugarcane bagasse (Sankar et al. 2018, Falk et al. 2019, Aphane et al. 2020, Ismail et al. 2021). One of these natural resources that are easily accessible in huge amounts is a byproduct of rice. There are different methods of silica nanoparticles synthesis like biotransformation method, microwave synthesis, thermal decomposition technique, laser ablation, chemical precipitation, plasma-assisted aerosol precipitation, sol-gel method, vapor-phase reaction, sonochemical synthesis, mechano-chemical method, hydrothermal synthesis, precipitation method, and pyrolysis, etc. (Krishna et al. 2021). Due to some environmental and technical problems like harsh experimental conditions, costly chemicals, and long-time and complex technologies in all these above syntheses, they are not environment friendly and economical to produce silica nanoparticles, thus, it is intriguing to build up flexible and elective techniques for acquiring nano silica from such biomass. There is a solution approach to a vital field of nanotechnology's potential applications for the present innovation. Furthermore, because the reuse of biomass assets has the potential to be extremely beneficial for eco-companion nanotechnology and nanoscience, the production of SiO_2 nanoparticles from rice husk has been extensively researched utilizing several exploratory methodologies. Chemical reaction techniques such as hydrothermal synthesis, acid-alkali leaching, microwave, combustion synthesis, precipitation, sonochemical, pyrolysis, and sol-gel are popular methods for producing SiO_2 nanoparticles from rice husk (Dubey et al. 2015, Sankar et al. 2016, Gao et al. 2017, Peres et al. 2018, Sankar et al. 2018, Almeida et al. 2019, Bui et al. 2020). When involving the biomass asset, rice husk is used to create silica as a siliceous substance at an unimaginable pace of gig tons/year, since rice husk contains an amorphous form of silica 90% (Hossain et al. 2018). After reviewing the literature, we found a persuasive and straightforward procedure for producing bio-created silica from a variety of rice husk sources. We used a rapid sonochemical method to demonstrate the properties of amorphous silica nanoparticles obtained from brown rice husk, which is one of the simplest methods for obtaining excellent silica of high quality from siliceous biomass reserves. The surface-to-volume ratio and microstructural size of silica oxide nanoparticles can be easily regulated by adjusting the sonication period throughout the sonochemical process. The optical, textural, morphological, and structural aspects of rice husk-derived silica oxide nanoparticles were investigated, as well as the impacts of sonochemical procedure duration on the objective properties (Sankar et al.2018). Because of the foregoing, rice husk is regarded as the greatest economically significant silica source.

Subsequently, we present a simple technique for synthesizing biogenic silica nanoparticles of high purity from rice husk and evaluated their biocompatible qualities. The silica nanoparticle was well characterized by different methods like X-Ray Diffraction, UV-VIS spectroscopy, Fourier Transform Infrared, Field Emission Scanning Electron Microscopy, and Energy Dispersive X-Ray. Silica nanoparticles have been made synthetically by various methods involving substance reductants.

MATERIALS AND METHODS

Material

For this study, rice husk (RH) was taken as a raw material from a rice mill. An analytical-grade chemical was used to make the nanoparticles.

Process of Silica Nanoparticles Synthesis

To remove dust and soil, rice husk was cleaned thoroughly with tap water adhering to it until the water was clear. The pH was then neutralized by washing it with distilled water. The rice husk was then rinsed and dried out in the sunlight for two days before being dried for 3 h at 90°C. The dried rice husk was ground into flour. The dried rice husk was then ignited at 600°C for 4 h to generate a grey powder of rice husk ash (RHA). A 500 mL NaOH (0.5M) solution was used to disperse the rice husk ash. For dissolving silica, stirring was done with a magnetic stirrer at 200 rpm for 3.5 h at a particular temperature of 90°C to form a solution of sodium silicate. Whatman no. 41 filter paper was used to filter the resultant solution. The filtrate from the sodium silicate solution was permitted to cool to room temperature. To generate silica precipitation, with regular stirring the sodium silicate solution was titrated with H_2SO_4 acid solution to pH 7. The solution was then agitated for one day before being aged for two days to let the silica gel to develop. Finally, using distilled water, the gel-containing solution was filtered, fragmented, and washed, yielding a fresh silica gel that was lyophilized overnight to get rid of water. For further characterization, the produced SiO_2 nanoparticles are stored in vacuum desiccators. Fig. 1 represents an alkali-based silica extraction process.

$$SiO_2 + 2NaOH \rightarrow Na_2SO_3 + H_2O$$
$$Na_2SiO_3 + H_2SO_4 \rightarrow SiO_2.H_2O + Na_2SO_4$$

Optimization and Characterization of Silica Nanoparticles

For value addition, silica production from rice husk using alkali digestion was precisely carried out at optimum pH, alkali dose concentration, digestion time, temperature, and adsorbent concentration.

Optimization has a lengthy history of research, notably in the subject of operational analysis, which has resulted in

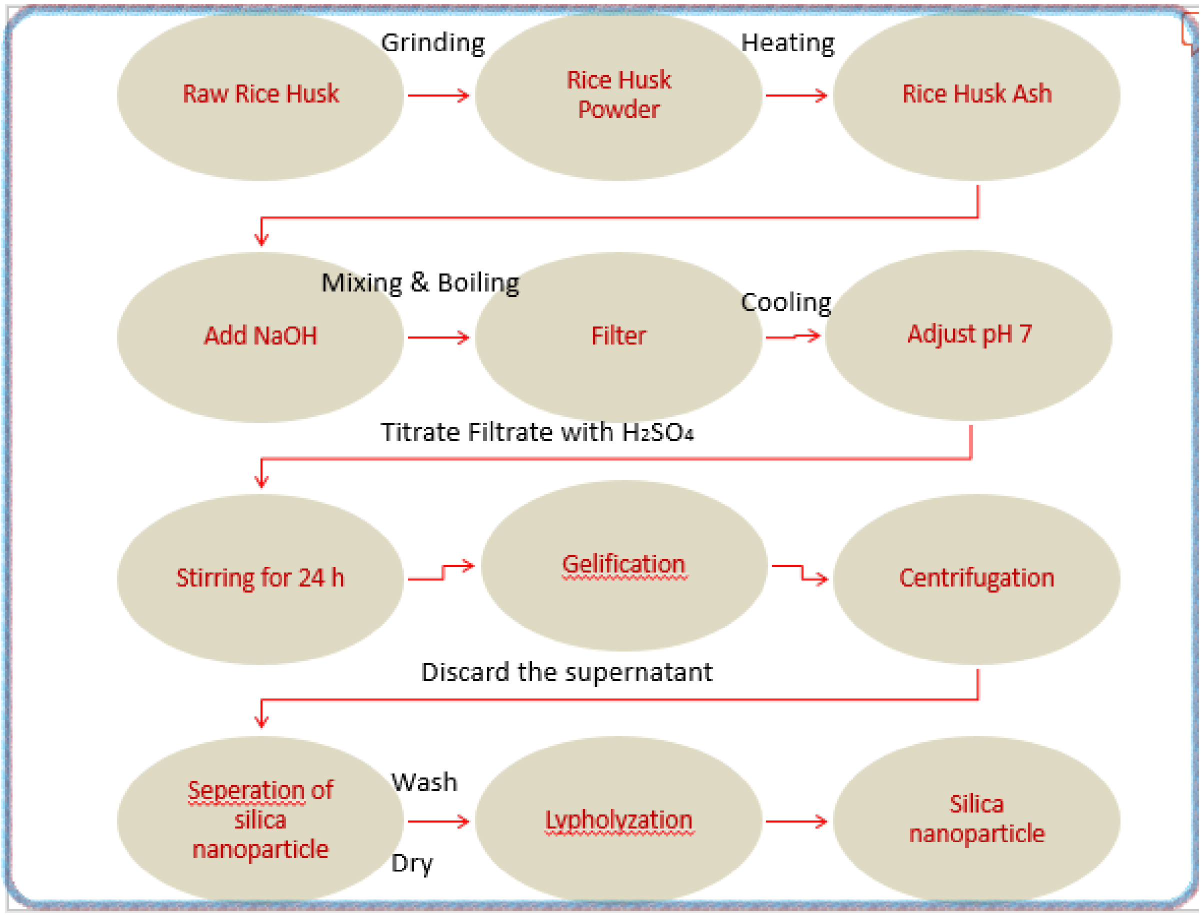


Fig. 1: Extraction process of silica nanoparticles from rice husk.

a plethora of methodologies. The influence of interaction among the elements is overlooked in traditional single-factor time testing because the experimenter modifies a single factor while keeping the other factors constant. Response surface methodology is a systematic analytical strategy for investigating the correlations between design features and responses to gain better overall knowledge with the fewest possible experiments (Cheng et al.2015). The Box Behnken generates designs that have favorable statistical features, allowing the quadratic model to be used. Response surface methodology optimizes processes and products by using quantitative data in an experimental design to discover and simultaneously solve multivariate equations. (Li et al. 2019).

The standard RSM, the Box–Behnken design model, for optimization, was built using Design Expert software 13. The optimum pH, NaOH concentration, temperature, time, and raw material dose were determined using a five-variable Box Behnken design. The design was chosen because it meets the majority of the requirements for silica nanoparticle production optimization. The fundamental goal of response surface methodology is to find the process's optimum practical conditions that meet the operating criteria. There were 46 experiments in the quadratic model's Box Behnken design (BBD). The design variables of BBD for silica nanoparticle production include pH (2-12), NaOH concentration (0.3-0.8 M), Temperature (40-130°C), time (2-5 h), and raw material dose (3-8 gm). To characterize the connection between independent variables and their observed responses, the equation employed was a quadratic polynomial (Cheng et al. 2015). The following is the model equation:

$$\boldsymbol{Y_{Pred}} \; = \; \boldsymbol{\beta_0} + \sum_{i=1}^{n} \boldsymbol{\beta_i x_i} + \sum_{i=1}^{n-1} \sum_{j=1}^{n} \boldsymbol{\beta_{ij} x_i x_j} + \sum_{i=1}^{n} \boldsymbol{\beta_{ii} x_i^2} + \mathbf{e} \qquad \ldots(1)$$

Where, Y_{pred} is the approximation value of the reactive variable, β_0 is constant, β_i, β_{ii}, β_{ij} are the linear, quadratic, and interaction constants of the regression coefficients, and x_i and x_j are the independent variables in the form of coded values and e is the residual error. The consequences of the silica nanoparticle manufacturing were analyzed using model graphs and analysis of variance (ANOVA) like three-dimensional graphs, expected vs actual value plots, and contour plots.

The properties of the Silica nanoparticles like functional groups, phase identification, surface topography, size analysis, determining elemental composition and absorbance were determined using Scanning Electron Microscopy (JSM-7610F Plus, JEOL Japan), X-Ray Diffraction (Rigaku Miniflex-II Diffractometer, Japan), Energy Dispersive X-Ray Analysis (EDX), UV-Vis Spectroscopy (Shimadzu) and Fourier Transform Infrared (Spectrum Two, Perkin Elmer).

Optimization of Silica Nanoparticles Synthesis

Effect of pH: The influence of pH on silica extraction was studied from pH 2 to pH 10 using 5 gm rice husk ash dissolving in 0.5 M NaOH solution at 100°C for 4 hours. Silica extraction productivity improved when the pH was raised from 2 to 7, which was 29.56% to 88.5% (Fig. 2(a)). Further increases in pH did not lead to a significant increase in silica nanoparticle yield. The results show that the optimum pH of 7 results in a dose for getting the maximum output of silica nanoparticles from rice husk ash. This optimum pH for silica synthesis was 7. Similar findings were made by Yang et al. (2019) of China in their ingvestigation on mesoporous silica aerogels.

Effect of alkali dose: To study the NaOH concentration effect on the digestion of 5 g rice husk ash to manufacture silica nanoparticles varying dose of alkali was studied at pH 7 for 4 hours at 100°C. When the NaOH concentration was elevated from 0.3 M to 0.5 M, the silica elimination rose from 40.6 % to 88.5 %. (Fig. 2(b)). Further rise in the alkali dose did not result in a substantial increase in the production of silica nanoparticles. As a result, in the given digestion conditions, the optimum dose of NaOH concentration for silica extraction was found as 0.5 M. In their research, Ghorbani et al. (2015), of Iran study on silica nanoparticles, employed a similar NaOH concentration and observed similar results.

Effect of Digestion Time

The influence of processing time on silica extraction was studied for 5 grams of rice husk ash at pH 7, 0.5 M NaOH, and 100°C digesting temperature over time intervals ranging from 0.5 to 6 hours. It has been detected that the silica production was elevated from 16.91 % to 88.45 % when the digestion period was increased from 0.5 to 3.5 hours [Fig. 2(c)]. However, increasing the duration beyond 3.5 hours there was no considerable increase in silica nanoparticle extraction, hence 3.5 hours was considered to be the ideal reaction time for rice husk ash working dose aforementioned specified digestion states for silica extraction. According to Yang et al. (2019), in a China study on mesoporous silica aerogels, the optimal time was 4 hours since there was no significant extraction after that.

Effect of temperature: To study the influence of digestion temperature ranging from 30°C to 150°C on silica extraction from rice husk ash (5 gm) in 0.5 M NaOH solution at pH 7, with a digestion time of 3.5 hours. The silica extraction efficiency escalated from 14.8 % to 88.45% when the digestion temperature was elevated from 30°C to 90°C (Fig. 2(d)). However, an increase in the temperature did not result in a substantial increase in nanoparticle extraction. The optimum temperature for extraction of silica nanoparticles was found 90°C. Manaa (2015) from Egypt's study on silica products also noted comparable outcomes.

Raw material dose: A variable dose of 2 gm to 10 gm of rice husk ash was used to extract silica nanoparticles while keeping the alkali dose constant at 0.5 M NaOH and a digestion period of 3.5 hours at 90°C. As shown in Fig. 2(e), the raw dose concentration increasing from 2 to 5 gm increased the silica extraction from 54.8% to 88.5 %. Though, raising it after 5 gm did not yield a remarkable rise in silica nanoparticle extraction. As a consequence, a 5 g dose of rice husk ash was discovered to be ideal for the largest yield of silica nanoparticles. Ghorbani et al. (2015) revealed a 5.0 g optimum dose of rice husk ash in their research.

RSM-BBD Analysis

Design Expert Software 13 was used to analyze the data performance for regression analysis. The encoded versions of a second-order polynomial equation, Eq. (2), that reveals silica nanoparticle production is shown below:

Silica Production (%)

$$Y_{Pred} = +89.17 + 26\,A + 6.38\,B + 10.63\,C + 2.56\,D + 2.56\,E + 2.00\,AB + 2.00\,AC + 2.50\,AD + 0.0000\,AE - 4.25\,BC - 1.75\,BD - 1.50\,B - +1.25\,CD - 1.0000\,CE + 2.75\,DE - 31.29\,A^2 - 4.29\,B^2 - 6.79C^2 - 2.37\,D^2 - 1.04E^2 \quad \ldots(2)$$

Fisher's F-test was accustomed to determining the arithmetical significance of the polynomial equation. Table 1 shows the study of data variability for the response surface quadratic model. Table 1 also includes the regression coefficients for the quadratic, linear, intercept, and interaction factors of the model. The p-values of the model terms were used to determine their significance. An F-test revealed that the model was extremely effective, with a Fvalue of 2122.78 and a p-value < 0.0001. The "lack-of-fit" F-value of 3.81

and p-value of 0.0716 suggested that the "lack-of-fit" was insignificant in comparison to the pure error.

In Predicted vs. Actual (Fig. 3), Contour (Fig. 4), and 3D response-surface (Fig. 5) plots show the kind of interactions between the five studied variables, as well as the relationship between experimental levels and responses of each variable.

Characterization of Silica Nanoparticles

Fourier transform infra-red (FTIR) spectroscopy: Fig. 6 shows the FTIR spectrum of silica dioxide nanoparticles. Specifically, the Si-O-Si vibration peak could be seen. The bending vibration, asymmetric stretching vibration, and symmetric stretching vibration are assigned to the transmittance peaks of Si-O-Si at 469.49 cm^{-1}, 1089.05 cm^{-1}, and 803.69 cm^{-1}, correspondingly (Mohd et al. 2017, Nandiyanto et al. 2016, Wibowo et al. 2017). The silica surfaces produced a broad peak at 3149.79 cm^{-1} due to the hydroxyl stretching vibration produced by the remaining adsorbed water and the silanol group vibration. (Chen et al. 2014). The H-O-H bond in molecular water is liable for the bending vibration, the band saw at roughly 1626.26 cm^{-1} (Chen et al. 2014). The carboxyl side groups show a symmetric stretching peak at 1406.01 cm^{-1} (Sarkar et al.2014). As in charged amines ($C=NH^+$), the 2359.78 cm^{-1} band exhibits NH^+ stretching (Lade et al. 2015). The findings are consistent with previous research on silica nanoparticles (Ghorbani et al. 2015, Manna 2015).

X-Ray powder diffraction: The SiO_2 nanoparticles XRD patterns made from rice husk ash shown in Fig. 7. The broad peak in the XRD pattern of rice husk ash at 22.12° established the amorphous nature of silica (Wibowo et al. 2017), which is appropriate for the formation of sodium silicate solution, however sharp peaks at 31.37°, 45.16°, 55.99°, and 75.02° specify the silica nanoparticles crystalline nature (Wahab et al. 2019). As a result of these XRD peaks, it was concluded

Table 1: Analysis of variance for response surface quadratic model.

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	22996.83	20	1149.84	2122.78	< 0.0001	significant
A-pH	10816.00	1	10816.00	19968.00	< 0.0001	
B-NaOH Concentration	650.25	1	650.25	1200.46	< 0.0001	
C-Temperature	1806.25	1	1806.25	3334.62	< 0.0001	
D-Time	105.06	1	105.06	193.96	< 0.0001	
E-Raw Material Dose	105.06	1	105.06	193.96	< 0.0001	
AB	16.00	1	16.00	29.54	< 0.0001	
AC	16.00	1	16.00	29.54	< 0.0001	
AD	25.00	1	25.00	46.15	< 0.0001	
AE	0.0000	1	0.0000	0.0000	1.0000	
BC	72.25	1	72.25	133.38	< 0.0001	
BD	12.25	1	12.25	22.62	< 0.0001	
BE	9.00	1	9.00	16.62	0.0004	
CD	6.25	1	6.25	11.54	0.0023	
CE	4.00	1	4.00	7.38	0.0118	
DE	30.25	1	30.25	55.85	< 0.0001	
A^2	8545.47	1	8545.47	15776.25	< 0.0001	
B^2	160.74	1	160.74	296.76	< 0.0001	
C^2	402.56	1	402.56	743.19	< 0.0001	
D^2	49.23	1	49.23	90.88	< 0.0001	
E^2	9.47	1	9.47	17.48	0.0003	
Residual	13.54	25	0.5417			
Lack of Fit	12.71	20	0.6354	3.81	0.0716	not significant
Pure Error	0.8333	5	0.1667			
Cor Total	23010.37	45				

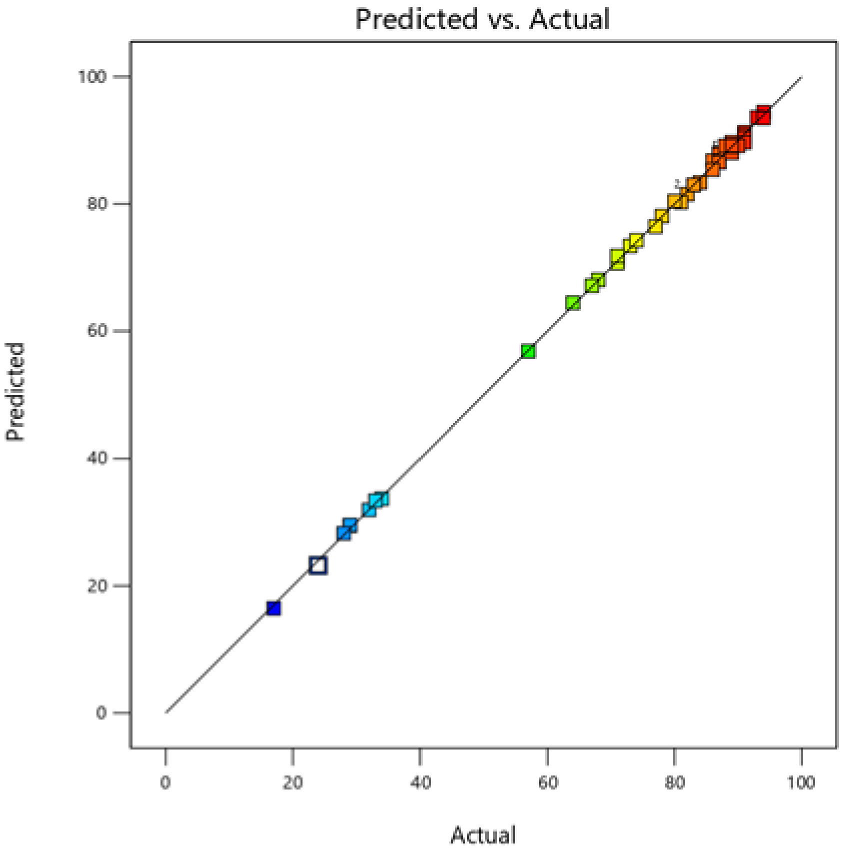


Fig. 3: Predicted vs. Actual plots.

that the rice husk contains a mixture of crystalline and amorphous silica phases. It was obvious that no other material impurities were present based on the peak positions of the observed spectra (Wahab et al. 2019). The creation of an amorphous form of silica nanoparticles has varied applications in our daily lives and adds to the beneficial effect (Raut & Panthi 2019). Peaks in the XRD pattern of silica nanoparticles generated by alkaline precipitation from rice husk ash are amorphous and contain a small proportion of crystalline silica (Wahab et al. 2019).

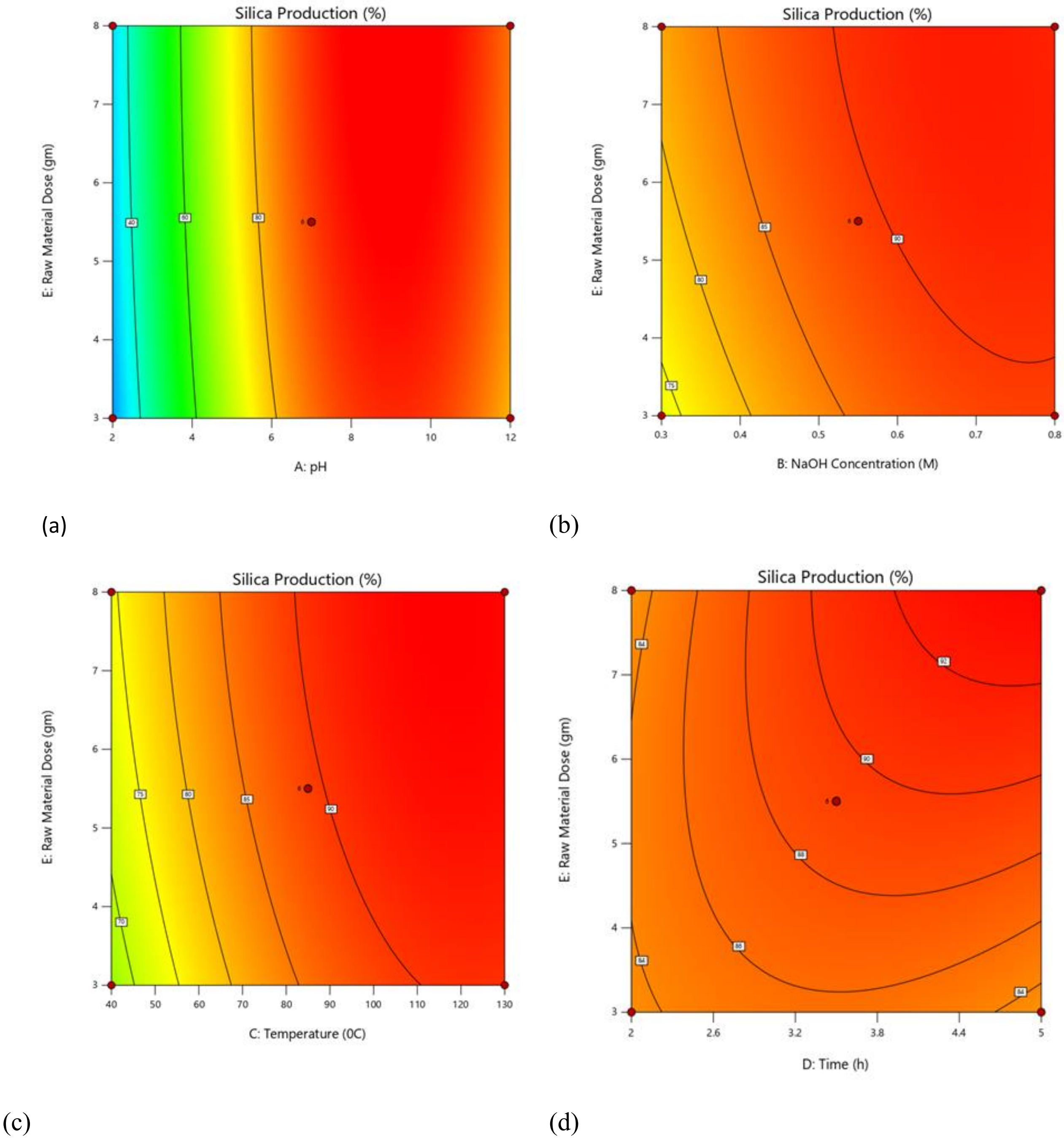
Scanning electron microscope (SEM) and EDX analysis: Scanning electron microscope images reveal the spherical shape structure of silica nanoparticles (Fig. 8). The silica nanoparticles produced were approximately 61.87 nm in size. A small portion of the generated SiO_2 particles formed an aggregation of SiO_2 nanoparticles, according to the findings. As a result, amorphous silica nanoparticles were discovered in nature, as seen in Fig. 8. Ahmad et al. (2017) also made similar observations.

EDX confirmed the chemical configuration of silica nanoparticles. The strongest peaks are displayed by oxygen and silica existent in silica nanoparticles, indicating that SiO_2 nanoparticles are generated. The non-appearance of other elements indicated that extensive washing with water had eliminated most of the soluble ions. During the thermal decomposition of rice husk, metal contaminants have also been from the rice husk transported along with the volatiles. Akhayere et al. (2019), reported similar results for synthetic silica in their elemental analysis.

UV-visible spectroscopy: The absorption band edge of silica nanoparticles by UV-visible spectroscopy was analyzed between 200 and 700 nm and a major adsorption band has been discovered at 312 nm with an absorbance of 1.91 (Fig. 9). The absorbance of nanoparticles is strongly influenced by wavelength and sample amount. The presence of silica nanoparticles is indicated by these observations, which lead to a Si-O-Si link. Patil et al. (2018) reported comparable findings.

CONCLUSIONS

The percentage yield of nano silica produced from rice husk ash at 600°C was 88.5%. The surface response approach using



(a)

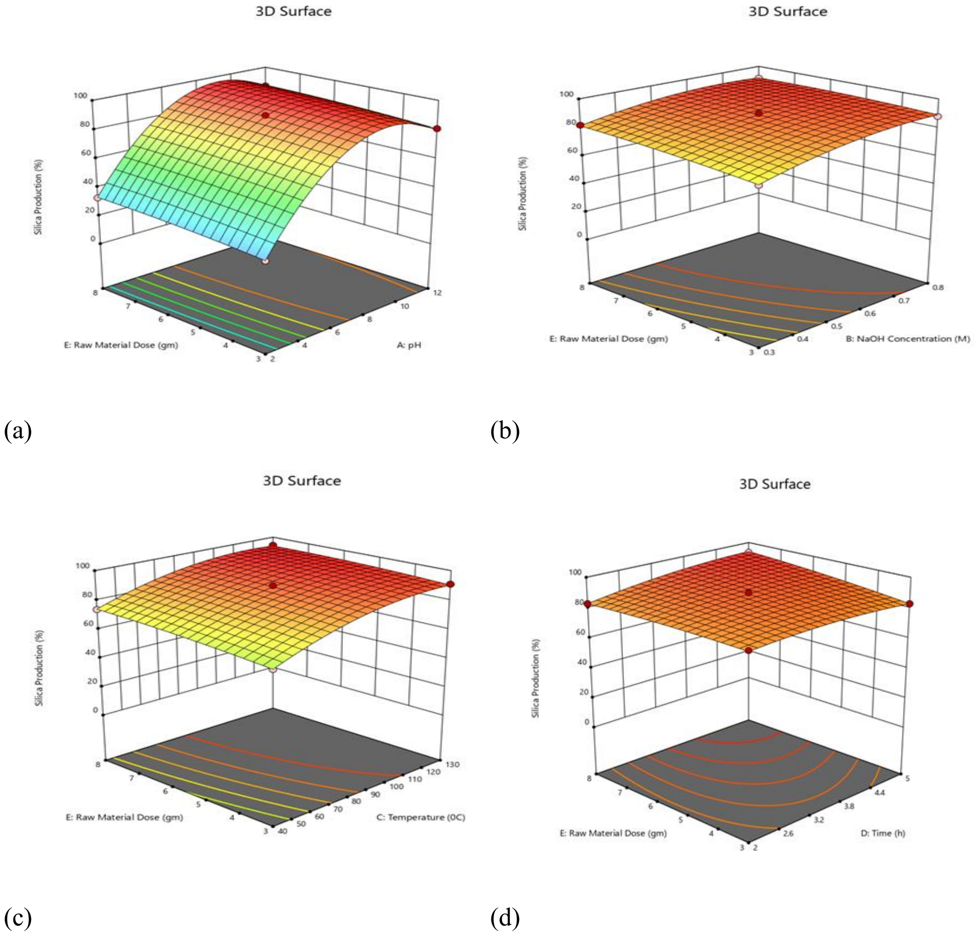
(b)

(c)

(d)

Fig. 4: Contour plots showing the interaction between (a) pH vs. raw material dose (b) NaOH concentration vs. raw material dose (c) Temperature vs raw material dose (d) Time vs. raw material dose.

the BBD model was successfully matched, as evidenced by the high correlation values of Adjusted R^2 0.9989 and Predicted R^2 0.9977. FESEM analysis of nano-silica particles from rice husk ash has shown its agglomeration form, with a particle diameter of 61.87 nm. The form of the particles was observed to be consistent. The presence of a significant broad peak at 22.12 on the XRD spectrum suggested that the nano-silica made from rice husk ash was mainly amorphous. The existence of hydrogen-linked groups siloxane and silanol in silica was established by FTIR data. The occurrence of O and Si in the ultimate formation, SiO_2 nanoparticles, was confirmed by EDX analysis. Biosynthesis



(a) (b)

(c) (d)

Fig. 5: 3D Response surface plots showing the interaction between (a) pH vs. raw material dose (b) NaOH concentration vs. raw material dose (c) Temperature vs. raw material dose (d) Time vs. raw material dose.

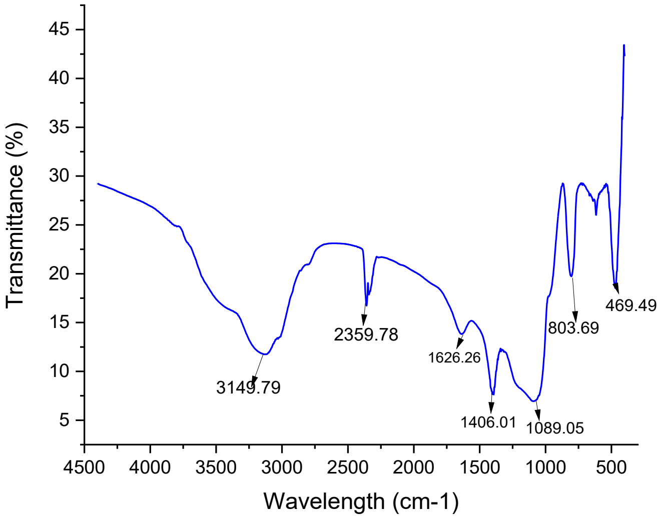


Fig. 6: FT-IR spectra SiO_2 nanoparticles prepared from rice husk.

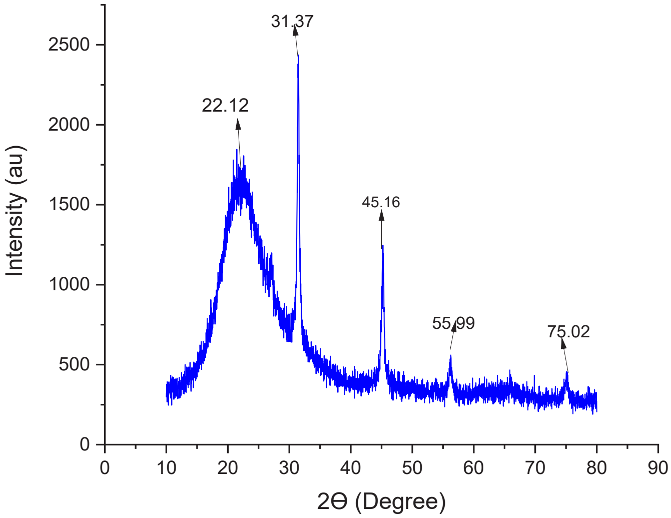
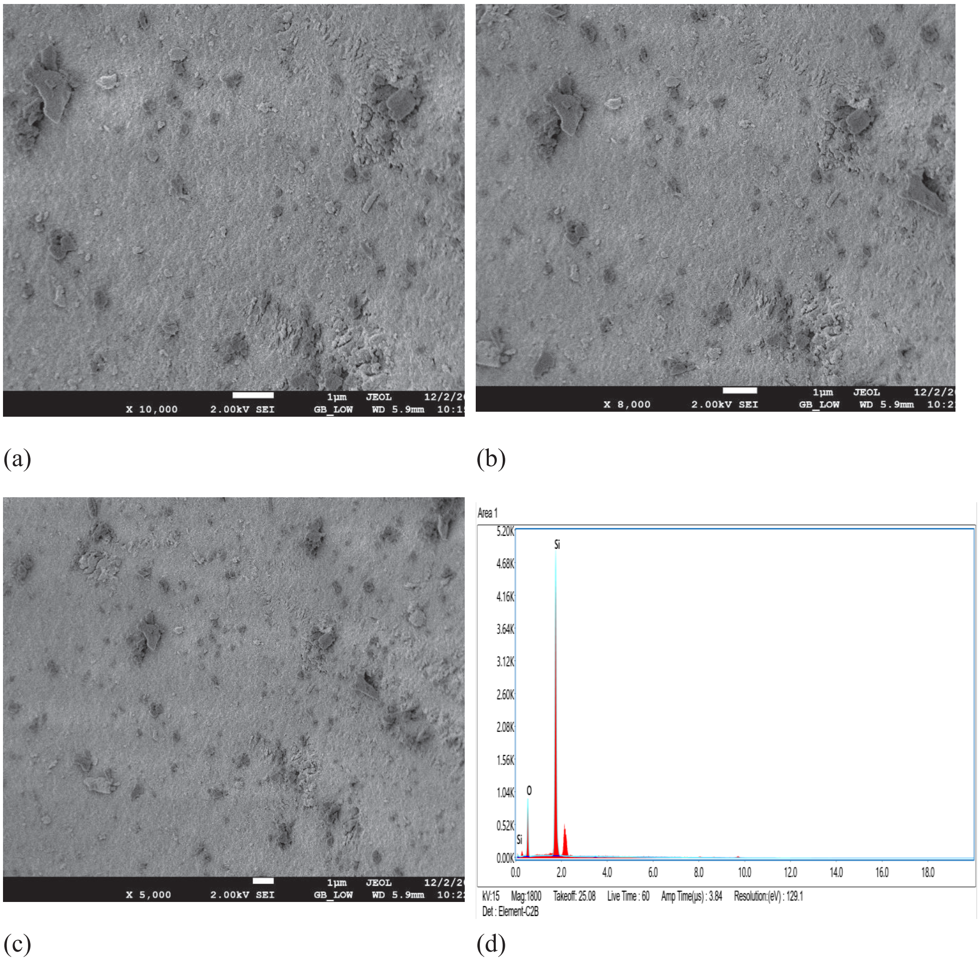


Fig. 7: XRD of SiO_2 nanoparticles prepared from rice husk.



(a)

(b)

(c)

(d)

Fig. 8: (a) SEM images taken with 1 **μm** (10,000) index (b) SEM images taken with 1 **μm** (8,000) index (c) SEM images taken with 1 **μ**m (5,000) index (d) EDX of SiO_2 nanoparticles.

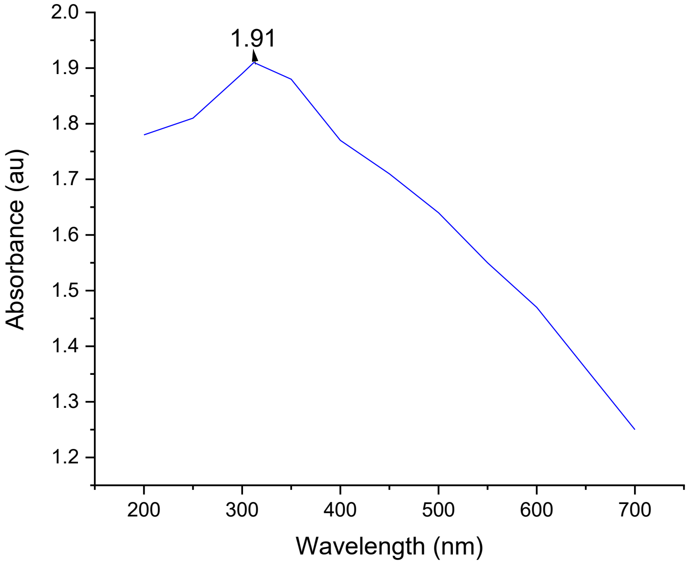


Fig. 9: UV-Visible Spectroscopy of SiO_2 nanoparticles procured from rice husk.

of silica nanoparticles using rice husk is an environmentally favorable, cost-effective green synthesis method. Rice husk as an origin of silica nanoparticles has a beneficial economic and environmental influence because it is a plentiful lower valuable agronomic by-product that can help with agro-waste clearance. In industry and agriculture, biosynthesized silica nanoparticles can be employed for a variety of applications.

ACKNOWLEDGMENTS

The authors express their gratitude to the Department of Environmental Science & Engineering and Central Instrumentation Laboratory, Guru Jambheswar University of Science & Technology, Hisar, for their support and productivity grants.

REFERENCES

Ahmad, I., Siddiqui, W. A. and Ahmad, T. 2017. Synthesis, characterization of silica nanoparticles and adsorption removal of Cu^{2+} ions in aqueous solution. Int. J. Emerg. Technol. Adv. Eng., 7(8): 439-445.

Akhayere, E., Kavaz, D. and Vaseashta, A. 2019. Synthesizing nano silica nanoparticles from barley grain waste: effect of temperature on mechanical properties. Pol. J. Environ. Stud., 28(4): 2513-2521.

Almeida, S. R., Elicker, C., Vieira, B. M., Cabral, T. H., Silva, A. F., Sanches Filho, P. J., Raubach, C. M., Hartwig, C. A., Mesko, M. F., Moreira, M. L. and Cava, S. 2019. Black SiO_2 nanoparticles obtained by pyrolysis of rice husk. Dyes Pigm., 164: 272-278.

Aphane, M. E., Doucet, F. J., Kruger, R. A., Petrik, L., and van der Merwe, E. M. 2020. Preparation of sodium silicate solutions and silica nanoparticles from south african coal fly ash. Waste Biomass Valorization, 8)11): 4403-4417.

Bapat, G., Labade, C., Chaudhari, A. and Zinjarde, S. 2016. Silica nanoparticle based techniques for extraction, detection, and degradation of pesticides. Adv. Colloid Interface Sci., 237: 1-14.

Bharti, C., Nagaich, U., Pal, A. K. and Gulati, N. 2015. Mesoporous silica nanoparticles in target drug delivery system: A review. Int. J. Pharm. Investig., 5: 124-133.

Bui, T. M. A., Nguyen, T. V., Nguyen, T. M., Hoang, T. H., Nguyen, T. T. H., Lai, T. H., Tran, T. N., Nguyen, V. H., Hoang, V. H., Le, T. L., Tran, D. L., Dang, T. C., Vu, T. Q. and Nguyen-T. P. 2020. Investigation of crosslinking, mechanical properties and weathering stability of acrylic polyurethane coating reinforced by SiO_2 nanoparticles issued from rice husk ash. Mater. Chem. Phys., 241: 122445.

Chen, X., Jiang, J., Yan, F., Tian, S. and Li, K. 2014. A novel low temperature vapor phase hydrolysis method for the production of nano-structured silica materials using silicon tetrachloride. RSC Adv., 4: 8703-8710.

Cheng, Z., Zhang, L., Guo, X., Jiang, X. and Liu, R. 2015. Removal of lissamine rhodamine B and acid orange 10 from aqueous solution using activated carbon/surfactant: process optimization, kinetics and equilibrium. J. Taiwan Inst. Chem. Eng., 47: 149-159.

Chong, M. Y., Numan, A., Liew, C. W., Ng, H. M., Ramesh, K. and Ramesh, S. 2018. Enhancing the performance of green solid-state electric double-layer capacitor incorporated with fumed silica nanoparticles. J. Phys. Chem. Solids., 117: 194-203.

Dubey, R. S., Rajesh, Y. B. R. D. and More, M. A. 2015. Synthesis and characterization of SiO_2 nanoparticles via sol-gel method for industrial applications. Mater. Today: Proc., 3579-3575 :2.

Falk, G., Shinhe, G. P., Teixeira, L. B., Moraes, E. G. and de Oliveira, A. N. 2019. Synthesis of silica nanoparticles from sugarcane bagasse ash and nano-silicon via magnesiothermic reactions. Ceram. Int., 45(17): 21618-21624.

FAO Statistics on rice production 2021. https://www.fao.org/worldfoodsituation/csdb/en/

Gao, M., Ma, Q., Lin, Q., Chang, J. and Ma, H. 2017. A novel approach to extract SiO_2 from fly ash and its considerable adsorption properties. Mater. Des., 116: 666-675.

Ghorbani, F., Sanati, A. M. and Maleki, M. 2015. Production of silica nanoparticles from rice husk as agricultural waste by environmental friendly technique. Environ. Stud. Persian Gulf., 2: 56-65.

Hossain, S. S., Mathur, L. and Roy, P. K. 2018. Rice husk/rice husk ash as an alternative source of silica in ceramics: A review. J. Asian Ceram. Soc., 6: 299-313.

Ismail, A., Saputri, L. N. M. Z., Dwiatmoko, A. A., Susanto, B. H. and Nasikin, M. 2021. A facile approach to synthesis of silica nanoparticles from silica sand and their application as superhydrophobic material. J. Asian Ceram. Soc., 9(2): 665-672.

Kang, S. H., Hong, S. G. and Moon, J. 2019. The use of rice husk ash as reactive filler in ultra-high performance concrete. Cem. Concr. Res., 115: 389-400.

Kano, S., Yamamoto, A., Ishikawa, A. and Fujii, M. 2019. Respiratory rate on exercise measured by nanoparticle-based humidity sensor. In: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3567-3570.

Krishna, J., Perumal, A. S., Khan, I., Chelliah, R., Wei, S., Swamidoss, C. M. A., Oh, D. H. and Bharathiraja, B. 2021. Synthesis of nanomaterials for biofuel and bioenergy applications. In: Nanomaterials: Application in Biofuels and Bioenergy Production Systems, Academic Press Publishers, pp. 97-165.

Lade, H., Govindwar, S. and Paul, D. 2015. Mineralization and detoxification of the carcinogenic azo dye Congo red and real textile effluent by a polyurethane foam immobilized microbial consortium in an upflow column bioreactor. Int. J. Environ. Res. Public Health., 12: 6894-6918.

Li, H., van den Driesche, S., Bunge, F., Yang, B. and Vellekoop, M. J. 2019. Optimization of on-chip bacterial culture conditions using the Box-Behnken design response surface methodology for faster drug susceptibility screening. Talanta, 194: 627-633.

Manaa, A. 2015. Extraction of highly pure silica from local rice straw and activation on the of the left carbon for chromium (VI) adsorption. Indian J. Chem. Technol., 10: 242-251.

Mohd, N. K., Wee, N. N. A. N. and Azmi, A. A. 2017. Green synthesis of silica nanoparticles using sugarcane bagasse. AIP Conference Proceedings, **1885**(1): 020123. AIP Publishing LLC.

Nandiyanto, A. B. D., Rahman, T., Fadhlulloh, M. A., Abdullah, A. G., Hamidah, I. and Mulyanti, B. 2016. Synthesis of silica particles from rice straw waste using a simple extraction method. IOP Conference Series: Materials Science and Engineering, **128**(1): 012040.

Nazeran, N. and Moghaddas, J. 2017. Synthesis and characterization of silica aerogel reinforced rigid polyurethane foam for thermal insulation application. J. Non-Cryst. Solids., 461: 1-11.

Patil, N. B., Sharanagouda, H., Doddagoudar, S. R., Ramachandra, C. T. and Ramappa, K. T. 2018. Biosynthesis and Characterization of Silica Nanoparticles from Rice (*Oryza sativa* L.) Husk. Int. J. Curr. Microbiol. Appl. Sci., 7: 2298-2306.

Peres, E. C., Slaviero, J. C., Cunha, A. M., Hosseini–Bandegharaei, A. and Dotto, G. L. 2018. Microwave synthesis of silica nanoparticles and its application for methylene blue adsorption. J. Environ. Chem. Eng., 6: 649-659.

Priya, T. R., Nelson, A. R. L. E., Ravichandran, K. and Antony, U. 2019. Nutritional and functional properties of coloured rice varieties of South India: a review. J. Ethn. Foods., 6: 1-11.

Raut, B. K. and Panthi, K. P. 2019. Extraction of silica nanoparticles from rice husk ash (RHA) and study of its application in making composites. J. Nepal Chem. Soc., 40: 67-72.

Sankar, S., Kaur, N., Lee, S. and Kim, D. Y. 2018. Rapid sonochemical synthesis of spherical silica nanoparticles derived from brown rice husk. Ceram. Int., 44: 8720-8724.

Sankar, S., Sharma, S. K. and Kim, D. Y. 2016. Synthesis and characterization of mesoporous SiO2 nanoparticles synthesized from biogenic rice husk ash for optoelectronic applications. Int. J. Eng. Sci., **8-353 :17.**

Sarkar, J., Ghosh, M., Mukherjee, A., Chattopadhyay, D. and Acharya, K. 2014. Biosynthesis and safety evaluation of ZnO nanoparticles. Bioprocess Biosyst. Eng., 37: 165-171.

Setiawan, W. K. and Chiang, K. Y. 2021. Crop residues as potential sustainable precursors for developing silica materials: a review. Waste Biomass Valorization., 12: 2207-2236.

Soltani, N., Bahrami, A., Pech-Canul, M. I. and González, L. A. 2015. Review on the physicochemical treatments of rice husk for production of advanced materials. Chem. Eng. J., 264, 899-935.

Wahab, R., Khan, F., Gupta, A., Wiggers, H., Saquib, Q., Faisal, M. and Ansari, S. M. 2019. Microwave plasma-assisted silicon nanoparticles: cytotoxic, molecular, and numerical responses against cancer cells. RSC Adv., 9: 13336-13347.

Wibowo, E. A. P., Arzanto, A. W., Maulana, K. D., Hardyanti, I. S., Septyaningsih, H. D. and Nuni W. N. 2017. Preparation and characterization of silica nanoparticles from rice straw ash and its application as fertilizer. J. Chem. Pharm. Res., 9: 193-199.

Yang, R., Wang, X., Zhang, Y., Mao, H., Lan, P. and Zhou, D. 2019. Facile synthesis of mesoporous silica aerogels from rice straw ash-based biosilica via freeze-drying. Bioresources., 14: 87-98.

Temporal Variation and Source Identification of Carbonaceous Aerosols in Monrovia, Liberia

Article *in* Scientific African · December 2022

DOI: 10.1016/j.sciaf.2022.e01540

Emmanuel Juah Dunbar
Delhi Technological University

Lovleen Gupta
Delhi Technological University

Scientific African 19 (2023) e01540

Contents lists available at ScienceDirect

Scientific African

journal homepage: www.elsevier.com/locate/sciaf

Temporal variation and source identification of carbonaceous aerosols in Monrovia, Liberia

Emmanuel Juah Dunbar, Lovleen Gupta[*]

Department of Environmental Engineering, Delhi Technological University, Shahbad-Daulatpur, Main Bawana Road, Delhi 110042, India

ARTICLE INFO

Article history:
Received 13 September 2022
Revised 26 December 2022
Accepted 27 December 2022

Editor: DR B Gyampoh

Keywords:
Black carbon
Organic carbon
Monrovia
Temporal variation
Cape Mesurado
Carbonaceous aerosols

ABSTRACT

This study examined the atmospheric Black Carbon (BC) and Organic Carbon (OC) data for two years from January 2019 to December 2020 in Monrovia, Liberia. This study is the first of its kind in Monrovia, Liberia and the Mano River Union region. The objective is to evaluate the temporal variation in BC and OC over Monrovia and locate the probable local and regional source locations contributing to BC and OC therein. We also presented here the relationship of BC and OC with meteorological parameters. The highest BC (~1.4 µg m^{-3}) and OC (~14 µg m^{-3}) concentrations were observed in January 2019 and 2020. A clear seasonal effect was found in both years, with high mean BC (~4.9 µg m^{-3}) and OC (~10 µg m^{-3}) during the dry season (November to April). Diurnal variation suggested high BC and OC during 5:00–7:00 AM, 12 noon, 2:00 PM, and 4:00–7:00 PM. Long-range transportation and local emission sources were studied using Conditional Bivariate Probability Function (CBPF), and Concentration Weighted Trajectory (CWT). CBPF indicated that higher BC and OC concentrations were experienced when the wind was blowing from the South (at >2 ms^{-1}) and North-West (1–2 ms^{-1}) where a lot of shipping activities are carried out. CWT for BC showed that the air mass passes over Senegal through Guinea Bissau via the Atlantic Ocean while for OC, on the Atlantic Ocean and Freeport of Monrovia, indicating shipping emissions as probable sources. The results of this study can help policymakers devise appropriate strategies to control the BC and OC emissions over Monrovia, Liberia.

Introduction

Black carbon (BC) and Organic carbon (OC) are some of the significant environmental pollutants that can affect air quality, change climate patterns, can enter deep into the human respiratory system leading to severe health problems, like cardiovascular diseases, respiratory diseases, cancer, etc. [9,37,40]. BC and OC originate from incomplete combustion operations of biomass or fossil fuels and are major contributors to ambient air fine particulate matter ($PM_{2.5}$) [9,37]. An investigation in Barcelona, Spain demonstrated that the BC-related respiratory mortality rate increased by 10% with an increment of 1.4 µg m^{-3} on the list of air pollutants [19]. According to Tefera et al. [34], carbonaceous constituents (BC and OC) comprise a considerable fraction of atmospheric $PM_{2.5}$; 69.8 ± 9.3% of 24-h averaging of $PM_{2.5}$, ranging from 46% to 94% in Addis Ababa,

[*] Corresponding author.
E-mail addresses: juahdunbar31@gmail.com (E.J. Dunbar), lgupta@dce.ac.in (L. Gupta).

https://doi.org/10.1016/j.sciaf.2022.e01540

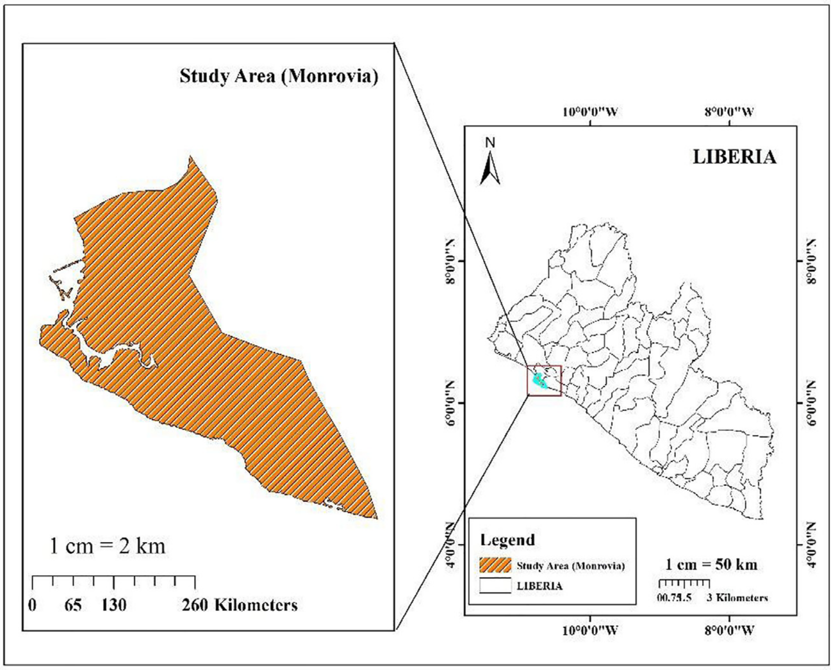


Fig. 1. Location of interested area (Monrovia).

Ethiopia, Africa. As claimed by the WHO report, BC and other toxic pollutants contribute not only to global warming but also to the increased premature deaths of 7 million [38].

Analysis of Africa indicated that air contamination already negatively affected African countries' financial strength and death rate. If potential and adequate measures are not taken, this effect could be stronger in the nearer future increasing diseases and death rates [13]. Liberia has been one of the underdeveloped countries in Africa and the world at large, the effect of air pollution and climate change is anticipated to be extreme since Liberia lacks a proper monitoring system and adequate support [11]. Stakeholders and citizens have low knowledge of Liberia's air quality. After the civil war in the country, it had been at a fast rate of development such as; infrastructure development, the increase of import and export of goods from the Freeport via the Atlantic Ocean, and also the increase of human population had led to the demand of charcoal for domestic and commercial use, leading to poor air quality. According to CEICdata.com, [6], Liberia had a mean annual $PM_{2.5}$ showing 18 μgm^{-3} in 2017, 17.3 μgm^{-3} in 2016, and 16.6 μgm^{-3} in 2015, showing a steady increase since 2015 clearly showing that $PM_{2.5}$ levels exceed 10 μgm^{-3} of WHO Interim Target-1 Value (% of Total) - Country Ranking, [22].

Many studies have been carried out on $PM_{2.5}$ and BC in Africa as well as the western region of Africa including Abidjan Ivory Coast, Dakar Senegal, Bamako Mali, and Nigeria which border and neighbor Liberia [10,18,26]. But to the best of our knowledge, no study on both OC and BC has been carried out so far in Liberia. Environmental Protection Agency in Liberia [12], the Liberian government body recognizes that reducing the concentration of BC and other pollutants over Monrovia will avoid crossing important thresholds such as 1.5°C temperature rise above pre-industrial levels and potential climate stumbling focus, which is in the capacity to impact the poor and vulnerable environment. This current investigation is the first study conducted on Monrovia, the capital city of Liberia, the Mano River Union, aiming to determine the air quality in Monrovia and to explore the potential source of air pollutants therein. The specific objectives of this study were to 1) Evaluate temporal variation in BC and OC over Monrovia; and 2) locate the probable local and regional source contribution to BC and OC concentration in Monrovia, Liberia.

Methodology

Study area

The observation site Monrovia (Fig. 1) which is the capital of Liberia located on Cape Mesurado on the Atlantic Ocean. Monrovia is the largest and the most populated city in Liberia with a total area of 194.25 km^2 and a total population of little above 939,524 according to the 2008 census, accounting for 28% of Liberia's population. Monrovia can be located within latitude 6.315 °N and longitude −10.8074 °W. The city of Monrovia is realizing rapid urbanization and growth, which is categorized with the increase in population. Due to the rapid developments in and around Monrovia, and a non-decentralization,

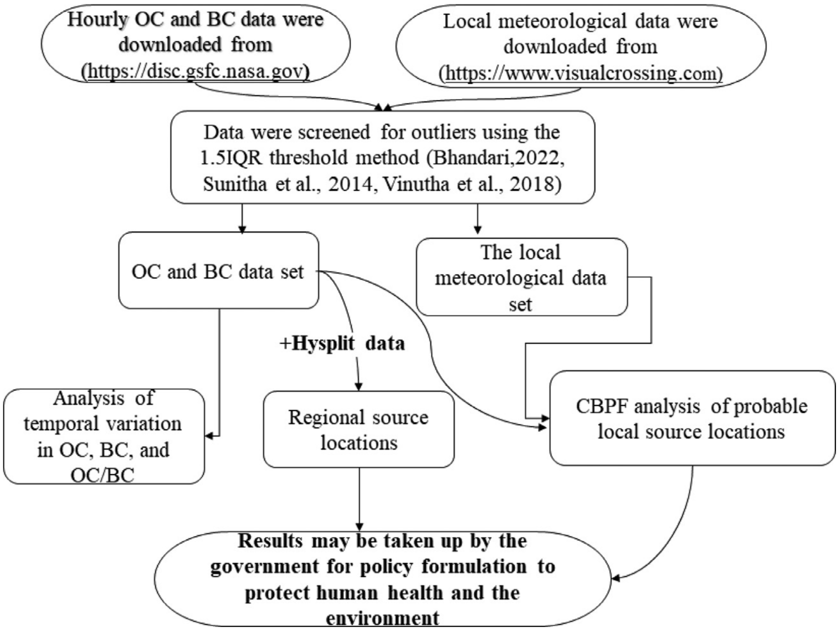

Hourly OC and BC data were downloaded from (https://disc.gsfc.nasa.gov)
Local meteorological data were downloaded from (https://www.visualcrossing.com)
Data were screened for outliers using the 1.5IQR threshold method (Bhandari,2022, Sunitha et al., 2014, Vinutha et al., 2018)
OC and BC data set
The local meteorological data set
+Hysplit data
Analysis of temporal variation in OC, BC, and OC/BC
Regional source locations
CBPF analysis of probable local source locations
Results may be taken up by the government for policy formulation to protect human health and the environment

Fig. 2. Flowchart of the methodology.

Monrovia is faced with serious anthropogenic emissions, including traffic emissions, shipping emissions, charcoal production emissions, etc.

Monrovia has two major seasons dry and wet seasons. The wet season runs from May to October and the dry season runs from November to April. The average maximum temperature of Liberia is around 30 °C and the minimum temperature is around 25.8 °C, during the day, humidity remains high. The monthly rainfall in Monrovia ranges from a minimum of 31 mm (January) to a maximum of 279.4 mm (September).

Study period and data

The study period chosen was from 1st January 2019 to 31 December 2020. Data of OC and BC was retrieved from satellite observation MERRA-2 (Modern-Era Retrospective analysis for Research and Application) (available online via https://disc.gsfc.nasa.gov) (MERRA-2, 2020). Correspondingly, meteorological data including temperature (°C), wind speed (m/s), wind directions (°), and relative humidity (%) was taken from (https://www.visualcrossing.com) (Corporation, 2020).

Reanalysis datasets have generated $PM_{2.5}$, OC, and BC fields at a high spatial-temporal resolution e.g., MACC and MERRA. Since these are evaluated inferred from the simulation of $PM_{2.5}$ employing a total aerosol life cycle in an art model, they are less prone to errors that are actuated in the case of satellite-based retrievals [20]. However, the accuracy of the model or chemical reanalysis depends primarily on their ability to stimulate the primary aerosols and their interaction [39]. In addition, Prabhu et al. [23] stated that MERRA-2 datasets and ground-based measurements had a good positive correlation of ($r = 0.80$), and the relationship gets more accurate when compared monthly. The detailed flowchart of the methodology is shown in Fig. 2.

Data analysis

The data were examined for outliers and maintenance periods as part of a quality control exercise. The 1.5 IQR threshold was used to screen outliers across the full data set, and those that did not meet its requirement were removed [4,32,36]. In addition, a few data points of OC that appeared to be outliers but were produced by extreme pollution events were not removed.

Data were analyzed using various software including SPSS (Version 22.0), Excel, and Openair package in R programme. Analysis was broken down into six (6) categories namely, seasonal trends, monthly trends, OC/BC ratio, meteorological parameters effect, and source analysis which includes Conditional Bivariate Probability Function (CBPF), Concentration-weighted Trajectory (CWT). The non-parametric Spearman rank correlation was adopted in this study to explore the relationship between BC, OC, and meteorological parameters [14]. This could be the most possible way of exploring the relationship between these pollutants and meteorological parameters since meteorological data are not normally distributed and are not linear.

Conditional bivariate probability function (CBPF)

CBPF is a method that is built on the probability of the concentration of the experimental pollutant that surpasses the threshold set of each range of wind direction and wind speed and also considers intervals of concentration [30,35]. The outcomes attained from the CBPF investigation can be used to compare the spatial map to determine the uniformity of the important source that impact the level of pollutants [30]. This study used CBPF (Bivariate polar plot) to identify probable local source locations. The bivariate polar-plot case offers additional evidence on the types of sources being identified by providing important depression characteristic evidence [35]. The below Eq. (1) can be used in calculating CBPF.

$$\mathrm{CBPF}_{\Delta\theta,\Delta v} = \frac{m_{\Delta\theta,\Delta v}|_{y\geq C\geq x}}{n_{\Delta\theta,\Delta v}} \tag{1}$$

where, $m\Delta\theta$, Δv is an amount of sampling data with a concentration between the given concentration interval x and y, and within the range of wind speed (Δv) and wind direction ($\Delta\theta$). $n\Delta\theta$, Δv is an amount of sampling data with any concentrations within the range of wind speed (Δv) and wind direction ($\Delta\theta$).

Air mass concentrated weight-trajectory (CWT)

An analysis of CWT was performed at the receptor location to indicate and explore the possible sources of air pollutants (BC and OC) over Monrovia. CWT was figured using R program 4.0.0 free version software along with its beneficial "Openair" package. Air mass trajectory data was attained from the Global Data Assimilation System (GDAS) with a resolution of 1° × 1° (ftp://arlftp.arlhq.noaa.gov/pub/archives/gdas1/) and was accessed on January 28, 2022. The 120 backward trajectories at (00, 06, 12, 18) hours intervals were analyzed at an altitude of 500 m during this investigation period. The concentration of pollutants was calculated using Eq. (2) below as given by [25].

$$\bar{\mathrm{C}}_{\mathrm{ij}} = \frac{1}{\sum_{\mathrm{k}=1}^{\mathrm{N}} \tau_{\mathrm{ijk}}} \sum_{\mathrm{k}=1}^{\mathrm{N}} \ln\left(\mathrm{ck}\right)\tau_{\mathrm{ijk}} \tag{2}$$

where i and j are the indices of the grid, k is the index of trajectory, N is the entire number of trajectories used in the investigation, $\mathrm{C_K}$ is the pollutant concentration measured upon arrival of trajectory k, and τijk the residence time of trajectory k in grid cell (i, j). A high value of $\mathrm{C_{ij}}$ means that air parcels passing over the cell (i, j) would, on average, cause high concentrations at the receptor site.

Results and discussion

Temporal variations of OC and BC

The variation of OC and BC was done on varying time scales viz. seasonal, monthly, and diurnal, which is discussed in the following paragraphs.

Seasonal variation

Fig. 3 shows the seasonal variation of BC and OC mass concentration of Monrovia, Liberia for two (2) years 2019 and 2020. It was observed that the highest BC and OC concentration for both years 2019 and 2020 was in the dry season. Results obtained for the dry season were recorded to be [4.90 ± 3.01 μgm^{-3} in 2019 and 3.42 ± 2.49 μgm^{-3} in 2020] and followed by the wet season [3.57 ± 2.03 μgm^{-3} in 2019 and 0.94 ± 1.06 μgm^{-3} in 2020] respectively for BC, while OC average concentration level was not in the same direction of extent for 2019 and 2020 when compared. From November to April which indicates the dry season, shows an average mean OC concentration of (10.00 ± 7.57 μgm^{-3}), and from May to October was recorded to be (3.04 ± 3.51 μgm^{-3}) in the wet season, 2019. Regarding 2020, the average OC concentration was obtained to be (6.54 ± 5.33 μgm^{-3}) in the dry season while (3.65 ± 3.43 μgm^{-3}) in the wet season, which indicated a slight increase in the wet season and a decrease dry season, shown in (Table S1 of SI). In addition, BC shows a decrease of (1.48 μgm^{-3}) in dry the season and (2.63 μg m^{-3}) in wet the season in 2020 as compared to 2019.

The high level of BC and OC concentration during the dry season in the atmosphere of Monrovia city could be due to biomass fires from land clearing during the farming period. Abbadie et al. [1] found that approximately 80% of savanna land is burnt in preparation for cultivation, thus strongly impacting BC concentration levels. Taking advantage of the high temperature in the dry season leads to the high burning of wood for the production of charcoal for domestic use (cooking) and the burning of garbage in and around Monrovia, which also leads to high concentration levels. Another major factor could be high traffic congestion in the dry season during the morning and evening hours along the Somalia Drive, Sinkor-Congo town road and the United Nation Drive Road due to poor road conditions, another key factor that could lead to the high concentration level in the dry season could be due to long-range transportation. During the wet season, there is low movement due to school closure and less garbage and wood burning, which could contribute to the low concentration of BC and OC in the city during the wet season. Rainfall during the wet seasons also aids in cleaning the atmosphere of pollutants like organic aerosol, BC, by coagulation, a natural phenomenon that improves air quality. In addition, during the wet season, anthropogenic activities reduce, for example, no cultivation and clearing of land by burning takes place during this season which leads to lesser emissions in the wet season in Monrovia. There were some extreme cases of OC concentration, the reasons for which cannot be ascertained at this time. More research is required to be carried to ascertain that.

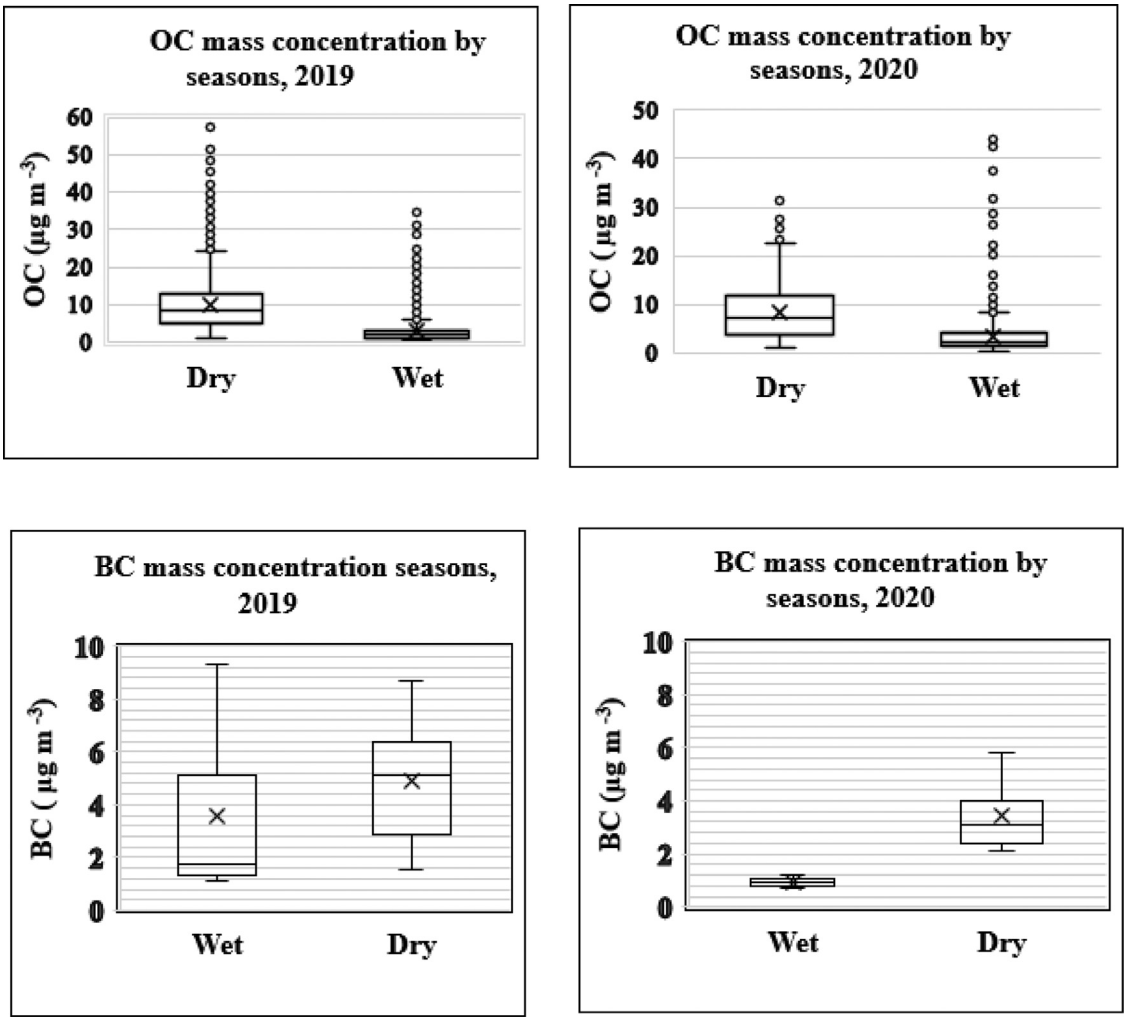
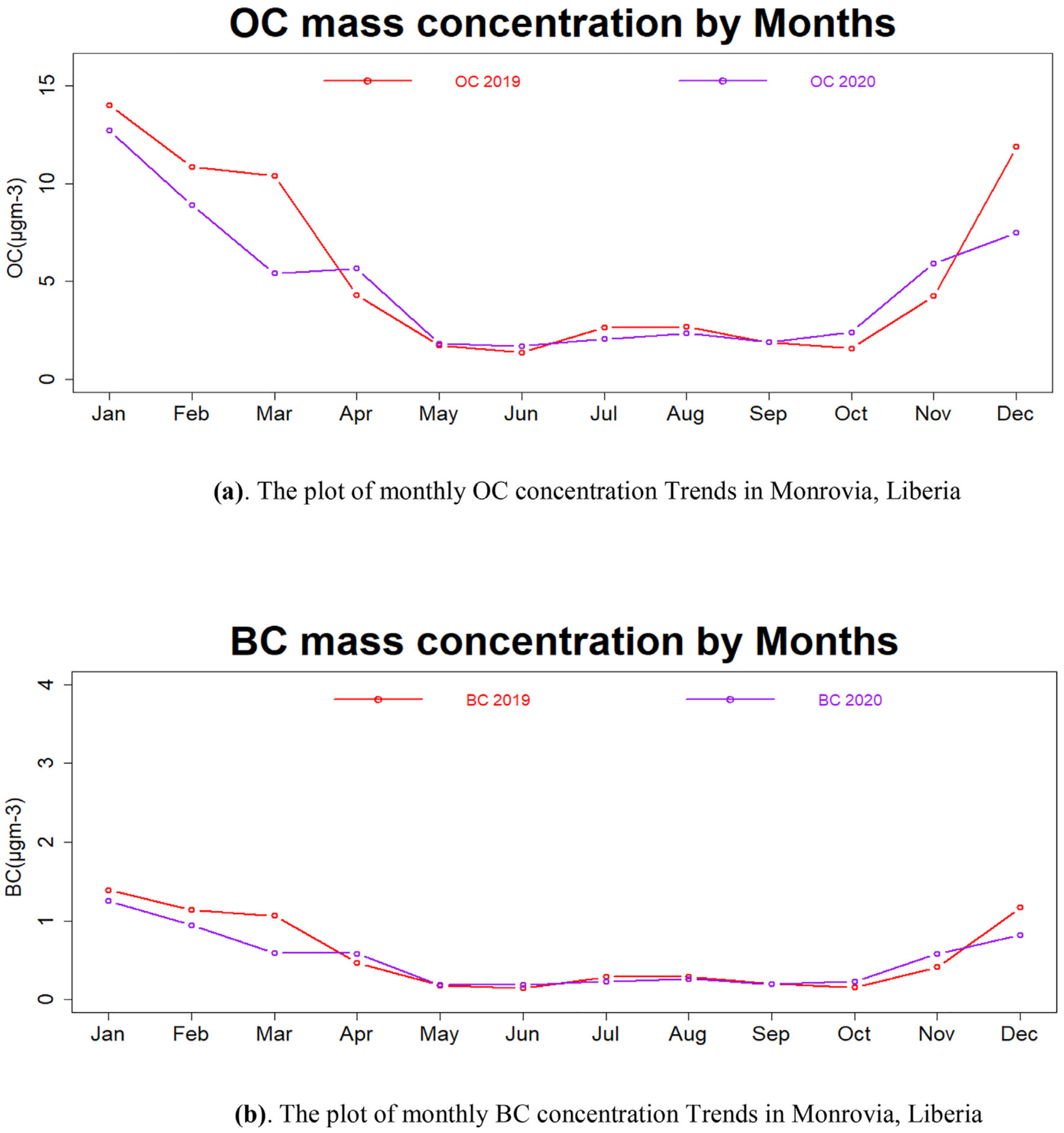


Fig. 3. Box plots for the BC and OC mass concentration for both 2019 and 2020 includes season. The upper and lower edge of the boxes indicates the 75% and 25% percentiles, respectively; the lines and stars in the boxes represent the medium and the mean values, respectively; the below and the upper box and whiskers denotes the 90% and 10% percentiles, respectively.

Monthly variation in OC is shown in Fig. 4(a). During 2019, high values of OC were recorded in the months of January, February, March, and December to be 14.0 µg m^{-3}, 10.8 µg m^{-3}, 10.4 µg m^{-3} & 11.9 µg m^{-3} respectively, however, the minimum value was observed in June to be 1.37 µg m^{-3}. For 2020, however, high values were recorded in January, February, and December to be 12.7 µg m^{-3}, 8.90 µg m^{-3} & 7.5 µg m^{-3} respectively while the minimum was observed in June to be 1.70 µg m^{-3}. For the other months, there was some decrease in OC concentration in 2020, which shows an improvement in OC concentration over Monrovia in 2020. For BC (Fig. 4b), the highest concentration was recorded in January (1.39 μgm^{-3} in 2019 and 1.25 μgm^{-3} in 2020), and the lowest concentration was recorded in June (0.14 μgm^{-3} in 2019 and 0.18 μgm^{-3} in 2020).

The high concentration of BC and OC in December, January, February, and March could strongly be influenced by the long-range transport of pollutants and secondary pollutants. During these months, the strong harmattan wind, which is characterized by the cold-dry wind blowing from the east or the northeast direction carries a high amount of dust and other unknown pollutants over west Africa in the western Sahara (desert) via the Atlantic Ocean with a mixture of sea aerosol arriving in Monrovia. Furthermore, the subtropical ridge of high pressure remained over the central region of the Sahara and the Gulf of Guinea with low pressure prevailing in the intertropical zone during these months. The above impact of harmattan wind may be the reason for high concentrations during these months. Doumbia et al. [10], stated that monthly outflow or emissions in Dakar, Senegal, range from 0.3 to 700 kg of BC concentration per month, with the highest concentration in January, which concord with the result in the study. Therefore, the source region of BC and OC in the West Africa region should be similar. Senegal is a West African nation that has a similar season as Liberia. The reduction of OC in 2020 could be justified by the COVID-19 lockdown policy in 2020 contributed to OC and BC reductions due to traffic mobility restrictions (COVID-19 - Response from Liberia | ITUC-AFRICA / CSI-AFRIQUE, 2020).



(a). The plot of monthly OC concentration Trends in Monrovia, Liberia

(b). The plot of monthly BC concentration Trends in Monrovia, Liberia

Fig. 4. (**a**). The plot of monthly OC concentration Trends in Monrovia, Liberia
(**b**). The plot of monthly BC concentration Trends in Monrovia, Liberia.

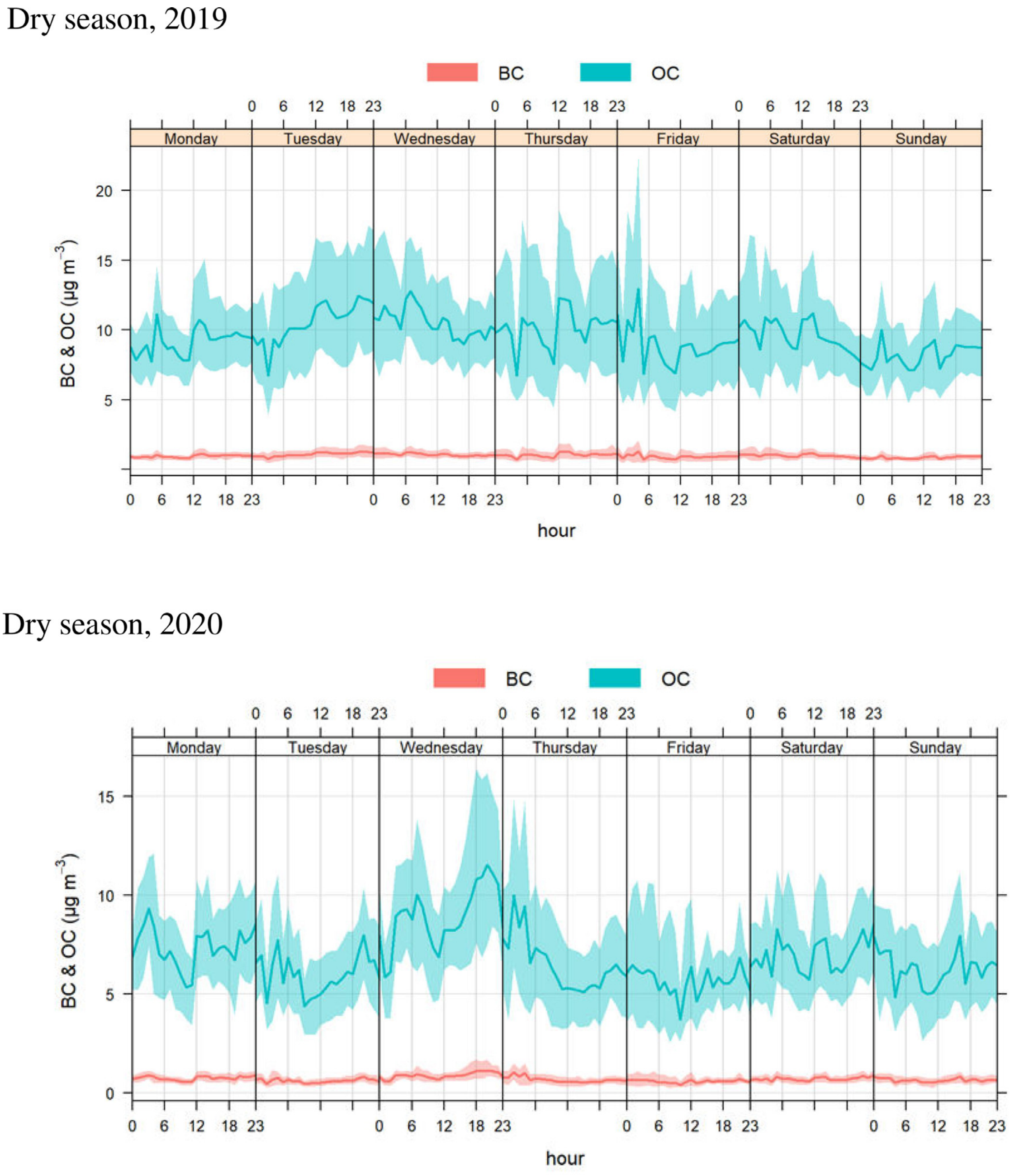
Diurnal analysis of dry and wet season

The diurnal plots during the dry and wet seasons during both years is shown in Fig. 5. Dry season high BC and OC concentrations were found associated with hours ranging from 5:00 AM to 7:00 AM, 12:00 PM to 2:00 PM, and 4:00 PM to 7:00 PM in 2019 Fig. 5(a). This high and moderate concentration could be justified by traffic emissions and other factors which will be discussed later in this section. Meanwhile, the afternoon to evening (4:00 PM to 7:00 PM) could strongly be influenced by the high burning of garbage in and around the city of Monrovia (e.g., Redlight market, Waterside market, and the Duwala market) where garbage is burnt in high quantity during the afternoon to the evening hours on a day-to-day basis. However, high concentration during the hours of 12:00 PM to 2:00 PM can be explained by three major contributing factors below.

- ➢ Junior and senior students have come to an end and school children leave to go home leading to high movement of vehicles and increasing traffic emissions.
- ➢ The second factor could be the high local charcoal production for commercial and domestic use during the dry seasons taking advantage of the high temperature at this time of the day (12:00 PM to 2:00 PM). During the production of local

charcoal using a traditional method, high deforestation and very high wood combustion take place leading to BC and OC emissions.

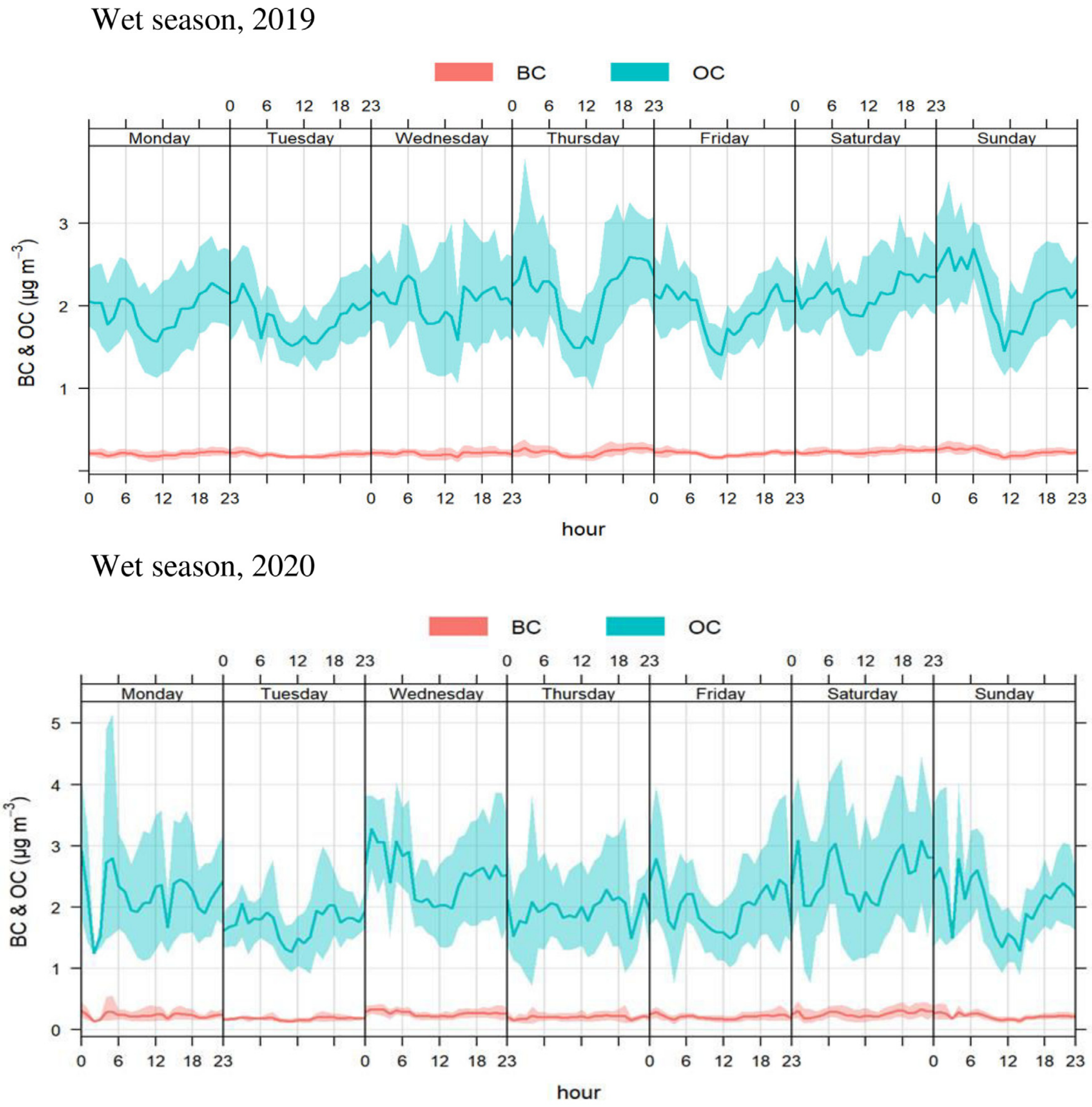
➢ The ocean breeze or ocean wind often takes place on hot and warm days during the dry and summer season when the temperature of the land is very high than the temperature of the water (Ocean, River, and Lakes). When Ocean-breeze tries to cool the air temperature by blowing cool air to the land, it also transports pollutants (BC, OC, and sea salt). To further explain this, Monrovia is bounded to the South by the Atlantic Ocean. Therefore, the above analysis suggests strongly that BC and OC concentration during the day and hours ranging from 12:00 PM to 2:00 PM are strongly associated with long-range transport via the Atlantic Ocean and its source could be due to high shipping activities.


Dry season, 2019

Dry season, 2020

(a). Diurnal concentration of BC and OC in dry season

Fig. 5. **(a)**. Diurnal concentration of BC and OC in dry season
(b) Diurnal concentration of BC and OC in wet season.



Wet season, 2019

Wet season, 2020

(b) Diurnal concentration of BC and OC in wet season

Fig. 5. Continued

In view of 2020 Fig. 5(a), the dry season only shows moderately high concentrations during the morning and evening indicating local traffic sources and wood combustion from cooking. The result of 2020, was influenced by the COVID-19 lockdown.

As shown in Fig. 5(b), analysis of diurnal variation suggested a similar trend for 2019 and 2020 respectively, in the wet season. A high concentration of BC and OC shows its peak from 5:00 AM to 7:00 AM and 3:00 PM to 8:00 PM with a slow decline from 7:00 AM and 9:30 pm. This increase in BC and OC concentration during the evening and morning strongly indicate the significant increase in traffic emission and wood combustion. Another factor that could lead to the high BC and OC concentrations during these hours could be due to local charcoal combustion from cooking since 95% of Liberia's population uses wood and coal for cooking. In addition, some moderate concentrations of BC and OC were found associated with hours between 1:00 AM and 2:00 AM which indicated long-range transportation, since anthropogenic activities are not taking place at that time. These results of diurnal variation in the wet season support the results and conclusion of CWT analysis below in this study, indicating that emissions during the wet seasons were meanly local with some small fraction being received at the receptor location from a long range. The lowest concentration of BC and OC in 2019 and 2020 respectively during the wet season was identified to be correlated with hours between 10:00 AM and 1:00 PM which indicates non-traffic hours. Therefore, this low concentration between 10:00 AM and 1:00 PM reflects the closure of schools during vacations.

Impact of meteorological parameters on BC and OC

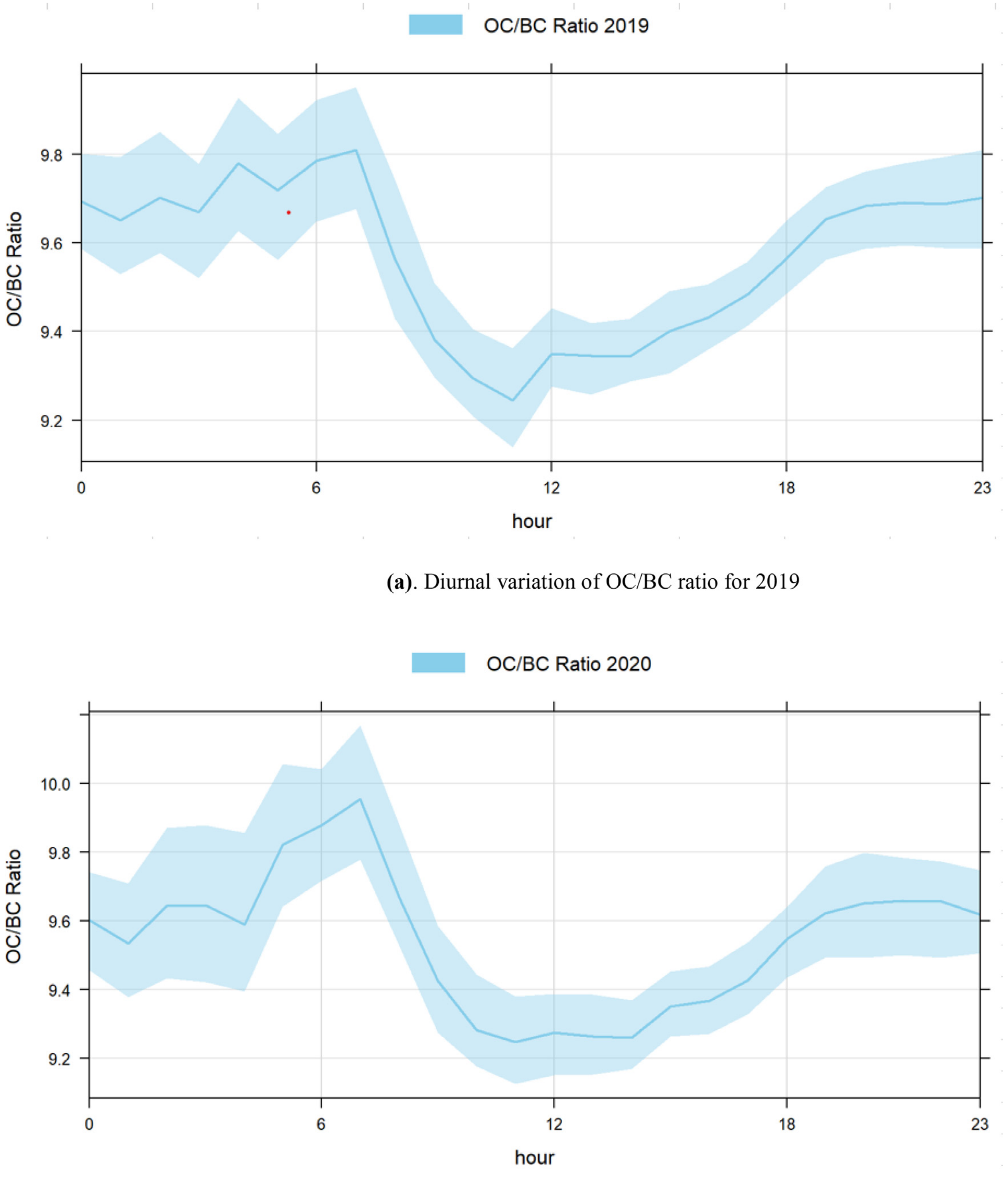
The relationship between wind speed (WS) and wind direction (WD) on BC and OC was investigated using the non-parametric Spearman correlation. The results (Table S3 of SI), indicated the relationship between BC, WS, and WD marked on a yearly basis. The results for both 2019 and 2020 indicate a negative correlation of WS and WD on BC. Wind speed is one significant component that influences BC concentration, higher wind speed denotes stronger BC concentration and dispersion [7]. In addition, if the wind speed is equal to or amounts to 1 ms^{-1}, BC concentration literally increased, if wind speed increases to a higher value of 2 ms^{-1} to 3 ms^{-1}, BC concentration decreases [17]. In view of Botsa et al. [5], BC concentration is unequivocally impacted by wind speed and wind direction, wind speed displays a better relationship with BC concentration. Analysis of the OC, WS, and WD relationship is shown (Table S3 of SI). It is clearly shown that meteorological parameters as a great effect on air pollutants (OC). In view of WS, WD, and OC, the analysis revealed that WS has a correlation coefficient of (−0.038) and p-value (∼ 0.004) in 2019 and a correlation coefficient of (−0.037) and p-value (∼ 0.025) in 2020. In view of WD, 2019 had a correlation coefficient of (−0.013) and p-value (∼ 0.3.40), whereas a correlation coefficient of (−0.054) and p-value (∼ 0.001) was recorded for 2020. This indicates a negative correlation between OC, WS, and WD.

Table S3 of SI, investigation of spearman correlation on RH and TEMP, results were found to be 2019-TEMP p-value (∼ 0.043) and correlation coefficient to be (0.027) and RH *p*-value (∼ 0.000) correlation coefficient (−0.103) while 2020-TEMP *p*-value (∼ 0.000) and correlation coefficient to be (−0.126) and RH *p*-value (∼ 0.390) and correlations coefficient of (−0.014) on BC. Furthermore, TEMP and RH have the following results on OC, TEMP as a correlation coefficient of (0.009) and *p*-value (∼ 0.487) in 2019 and a correlation coefficient of (−0.145) and *p*-value (∼ 0.000) in 2020. Given RH, 2019 had a correlation coefficient of (−0.086) and *p*-value (∼ 0.000), whereas a correlation coefficient of (0.002) and *p*-value (∼ 0.910) was recorded for 2020. The above outcomes indicate that TEMP had a positive correlation with both BC and OC in 2019, while in 2020 TEMP had a negative correlation. However, RH displays a negative correlation in 2019 between BC and OC whereas a negative correlation between BC was observed in 2020 and a positive correlation with OC in 2020. Therefore, the results of TEMP and RH on both BC and OC in 2019 and 2020 over Monrovia indicate that there is a need for further analysis in other years to establish the true relationship of TEMP and RH on air pollutants (OC and BC). In regard to Takemura and Suzuki, [33], it appears that a reduction in surface TEMP with the decrease in BC outflow is not strong as anticipated. According to Baker et al. [3], a recent investigation has indicated that the affectability of surface temperature on BC tends to be lower than anticipated. It is accepted that the expanding displayed day concentrations or outflows due to BC by indeed a calculated factor of ten (10) would denote little changes in surface air temperature, due to the prevailing fast alteration of the climate framework [31]. Changes in air temperature are conventionally evaluated from immediate constraining which relate to climate sensitivity. These findings suggest that decreasing the climatic TEMP may not be successful for a reduction in the local BC concentration of Monrovia, Liberia. In addition, 2019 BC concentration and TEMP in Monrovia had a positive correlation while RH shows a negative correlation with BC. This conforms to the result of Chen et al. [7] stating that analysis indicates that BC and TEM indicate a positive correlation during winter in Ahmedabad, BC concentration gets higher with the increase of TEMP, mostly during the dry season. The year 2020 in Monrovia shows the opposite of 2019, denoting that BC, TEMP, and RH had a negative correlation. This result is in line with Botsa et al. [5], investigation expressed that BC contained a negative correlation with RH and TEMP.

Carbonaceous aerosol ratios analysis

Diurnal variation of OC/BC ratio

Diurnal variation of the OC/BC ratio for 2019 and 2020 is presented in (Fig. 6). Diurnal variation of OC/BC-2019 Fig. 6(a) shows that hours from 12:00 AM to 5:00 AM and 9:00 PM to 11:00 PM were associated with ratios ranging between 9.7 to 9.79 and 9.66 to 9.7 with a peak taking place around 3:00 AM. Hours ranging between 12:00 AM to 5:00 AM and 9:00 AM to 11:00 PM with a high ratio indicates long-range transport of pollutants since anthropogenic activities are not taking place during these hours. Accounting for the peak around 3:00 AM, the peak around 3:00 AM could be justified by the high wind which influences the transportation of pollutants. It was also observed that the hours of 6:00 AM to 8:00 AM and 4:00 PM to 8:00 PM were linked with the ratio of (9.8 and 9.7), these hours indicate traffic rush hours, high OC/BC ratio at 6:00 AM to 8:00 AM and 4:00 PM to 8:00 PM denote the presence of secondary organic carbon. These secondary organic carbon sources could be from waste combustion, which disposed of some chemical substances such as Volatile organic compounds (VOCs) into the atmosphere leading to a mixture of existing pollutants (such as primary organic carbon) and those substances forming those secondary organic aerosols. The low ratio of 9.34 around the hour between 12:00 PM to 3:00 PM indicates local emissions such as charcoal production and traffic emission. In 2020, the OC/BC-2020 of diurnal variation Fig. 6(b) indicated a very high ratio of 9.91 during the hours of 6:00 AM to 7:30 AM which shows that traffic emission and fossil fuel burning were dominant at these hours. Meanwhile, a low ratio was observed between 10:00 AM to 2:00 PM denoting biomass burning. Further analysis of the OC/BC ratio on seasonal variation is found in the supplementary text (1.0) section.



(a). Diurnal variation of OC/BC ratio for 2019

(b): Diurnal variation of OC/BC ratio for 2020

Fig. 6. **(a)**. Diurnal variation of OC/BC ratio for 2019
(b): Diurnal variation of OC/BC ratio for 2020.

Relationship between OC and BC

In research by Singh and Srivastava [28], the most probable emission source of OC and BC concentration can be suggested or revealed by their interrelation with each other. In (Fig. 7) a significant positive linear correlation (R^2 = 0.98 in 2020 and R^2 = 0.99 in 2019) was observed between OC and BC suggesting that the emission sources of both OC and BC sources was similar. The high correlation denotes the existence of a primary source of OC and BC and these primary sources could be biomass combustion, vehicle emissions, and shipping emission. In research by Adon et al. [2], there was a positive linear

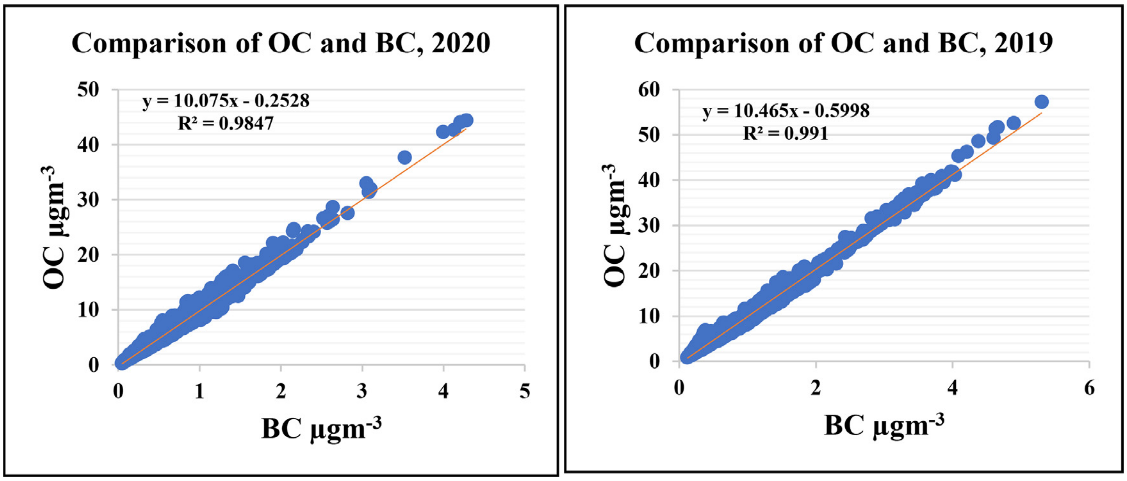


Fig. 7. Correlation between OC and BC concentration for both 2019 and 2020.

correlation recorded (R^2 = 0.9, 0.9, 0.8, and 0.8) in Ivory Coast (Abidjan) and Benin (Cotonou) of which these values of R^2 correspond to the value in this study. Abidjan and Cotonou are both west African cities with Ivory Coast (Abidjan) sharing a common border with Liberia. In addition, the OC/BC ratio from (Table S6 of SI) was ranging between 6.09 to 14.9 for 2020 and 6.75 to 17.9 for 2019, an annual average was obtained to be 9.55 µg m^{-3} in 2019 and 9.51 μgm^{-3} in 2020. The high OC/BC ratio for both years indicated the existence of secondary organic carbon (SOC), long-range transport of pollutants, and biomass burning.

This secondary organic carbon could be aerosol that originates from a nearby region and passes over the Atlantic Ocean with some mixed marine aerosol arriving at the receptor location Monrovia, it can also be local anthropogenic secondary organic aerosol. Samples of BC and OC concentration collected on and around the shore of the Atlantic Ocean indicates a high OC/BC ratio of 10, which indicate biomass fire to be the source of Atlantic Ocean OC and BC (Maritz et al., 2015). According to Saarikoski et al. [27], a larger OC/BC ratio suggests the long-range transport of pollutants to a receptor location. In view of Chow et al. [8], the ratio of OC/BC that is greater than two (>2) denotes the existence of secondary organic carbon and the lower OC/BC ratio denotes biomass burning influence. Literature from other studies including the closest regions to this current study location (Table S7 of SI) reported OC/BC ratios ranging from 2 to 29.5, observing the highest in Asia (Kosan) in 2005 and Ivory Coast (Abidjan) in 2020 to be (29.5 and 25), comparing the ratio observed in our study (9.56 and 9.52) for 2019 and 2020 to literature (Table S7 of SI) ratios, our OC/BC ratio is lower than Asia (Kosan) and Ivory Coast (Abidjan) but in the same range of Abidjan and Cotonou (2 to 10) in 2015–2017 but slightly lower than S.Pietre ratio (16) in 1998 to 1999. Therefore, indications are BC concentration in this study was higher than in Asia (Kosan) 2005, Ivory Coast (Abidjan) 2020, and S. Pietre 1998–1999. By Novakov et al. [21], the lower the OC/BC ratio is the higher the concentration of BC and the higher the OC/BC ratio is the lower the concentration of BC.

Source analysis of black carbon and organic carbon

Bivariate polar plot for BC and OC concentration with the function of wind speed and wind direction in hours was created at the 90th percentile and shown in (Fig. 8), for 2019 and 2020. Due to the presence of major point sources of emissions close to the study site, the higher percentile was chosen, and this can indicate the most effective way to evaluate the feasible geographical origins of BC and OC concentration in Monrovia. The higher concentration of BC and OC is presented mainly with a wind speed of (≈ 2 m s^{-1} – 5 m s^{-1}) in the South direction which is the Atlantic Ocean and the lowest BC and OC concentration was observed with wind speed t be (≈ 3 m s^{-1} - 5 m s^{-1}) in the North-west direction indicating the freeport of Monrovia, it was also recorded that moderate BC and OC concentration is associated with a very low wind speed of (≈ 1 m s^{-1} - 2 m s^{-1}) in 2019. This result of 2019 indicated that BC originated both locally and was transported, it can further be considered that 80% of BC and OC were transported due to its high wind speed association. In consideration of 2020, low BC concentration was partially distributed in all regions (West, South, East, and North) and was observed to correlate with wind speed to be (≈ 2 m s^{-1} - 5 m s^{-1}). In addition, a very high OC concentration was denoted by wind speed (≈ 4 m s^{-1} - 5 m s^{-1}) in the direction of the north-west and low concentration in the south-east direction with wind speed ranging (≈ 2 m s^{-1} - 4 m s^{-1}) in 2020.

Therefore, the result of both BC and OC in 2020 denotes that majority of emissions were not locally generated. The result of 2020 could strongly be affected by the COVID-19 lockdown when industrial activities and movement were restricted. Black carbon mass concentration reduced significantly between earlier and throughout the lockdown period by > 40% within the atmosphere of Bhubaneshwar, Chongqing, Hangzhou, Milan, and Suzhou. 35% was recorded over Ahmedabad as a reduction

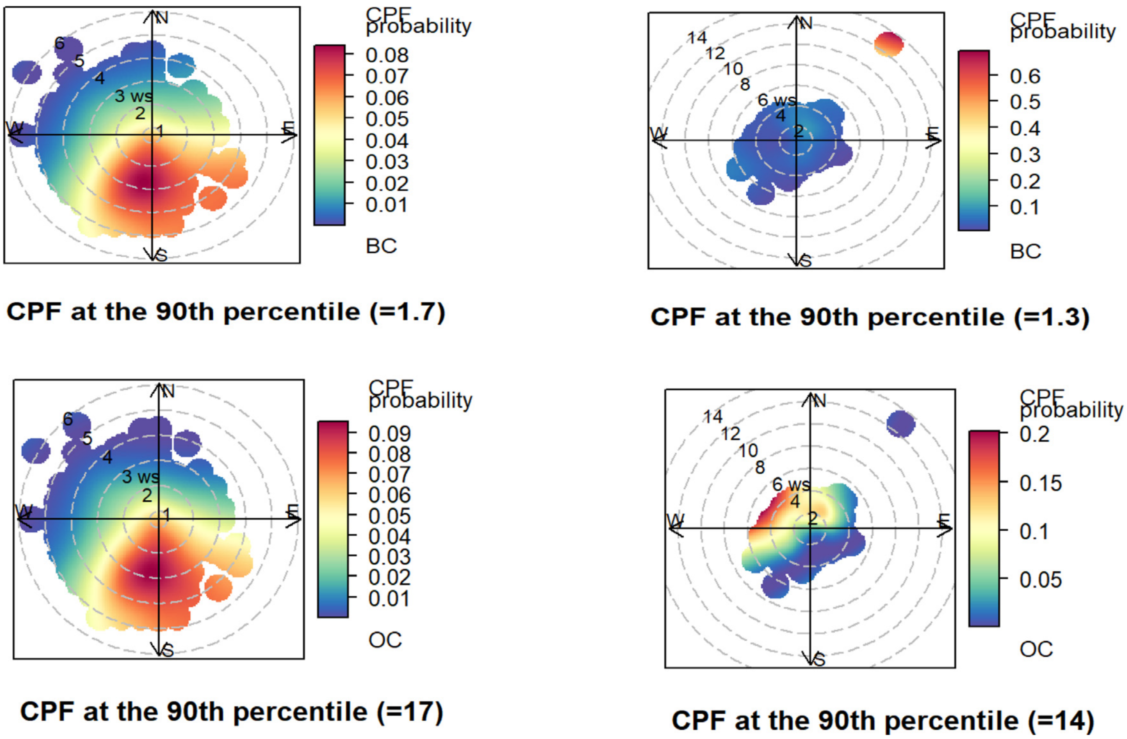


Fig. 8. CBPF of BC and OC for 2019 and 2020 respectively, to the upper left is 2019 BC, upper right is 2020 BC, lower left is OC 2019 and the lower right is 2020 OC, respectively.

in the black carbon concentration throughout the lockdown period when linked to the normal period (2017–2019), therefore, on average, the BC mass concentrations have reduced by more than 35% across the globe [24]. Results of both years denote that BC is more or less directional dependent and emissions are less local. Since Monrovia was built along the shores of the Atlantic Ocean, on the Mesurado Peninsula, this result could be influenced by the high shipping activities. Liberia's major route for export and import is the Free Port of Monrovia. Another significant factor could be the long-range transport of air pollutants indicating that the harmattan wind plays an important role in transporting pollutants via the Atlantic Ocean arriving over Monrovia. Guo et al. [15], also found that in reality ship emissions were, in addition, a crucial contributing factor to the BC concentration of Qingdao in past investigations.

CBPF plots revealed that the probable sources affecting Monrovia were prevalent at different wind speeds and were in different directions. Although the CBPF gave a decent representation of local source directions, this analysis's ability to depict the likely source directions for transboundary pollution-related episodic episodes has limits. Consequently, CWT was run, and the results are discussed in the next section.

Concentration-Weighted trajectory analysis

Further analysis in inspecting and evaluating the proper region and direction of pollutants (BC and OC), BC and OC were linked with air-mass backward trajectories forming the analysis of CWT. (Fig. 9), indicated the seasonal distribution outline of CWT-BC and CWT-OC concentration at the receptor location in Monrovia, Liberia. As denoted in Fig. 9, the high CWT-BC concentration of 2019, the dry season in the southwest direction, passes over Senegal through Guinea Bissau via the Atlantic Ocean possibly a mix of marine aerosol arriving at the receptor location Monrovia, Liberia. The high CWT-OC 2019 during the dry season, was associated with two unique regions, the south and the north-west region which indicate marine aerosol because Monrovia is bounded to the south by the Atlantic Ocean and north-west by the freeport of Monrovia where high shipping activities go on daily. It should also be noted or taken into consideration that some moderately high CWT-OC concentration was associated with the central region of the receptor location, Monrovia which indicates a local source. However, CWT-BC and CWT-OC in 2020 show a unique directional dependence. CWT- OC, and CWT-BC were linked to the south region and central region of the receptor location. Moderate CWT concentration of (0.4 μg m^{-3} – 0.6 μg m^{-3}), BC and (4 μg m^{-3} - 6 μg m^{-3}), OC on the scale of CWT, was observed in the eastern region arriving from Burkina Faso and Côte *d*'Ivoire. In addition, from Fig. 8, Wet season CWT-BC and CWT-OC concentrations for both 2019 and 2020 respectively, were observed that most of the aerosol concentrations originated from the south direction of Monrovia which is the region of the Atlantic Ocean. It can further be explained that a very high CWT-BC and CWT-OC concentration arrived from the

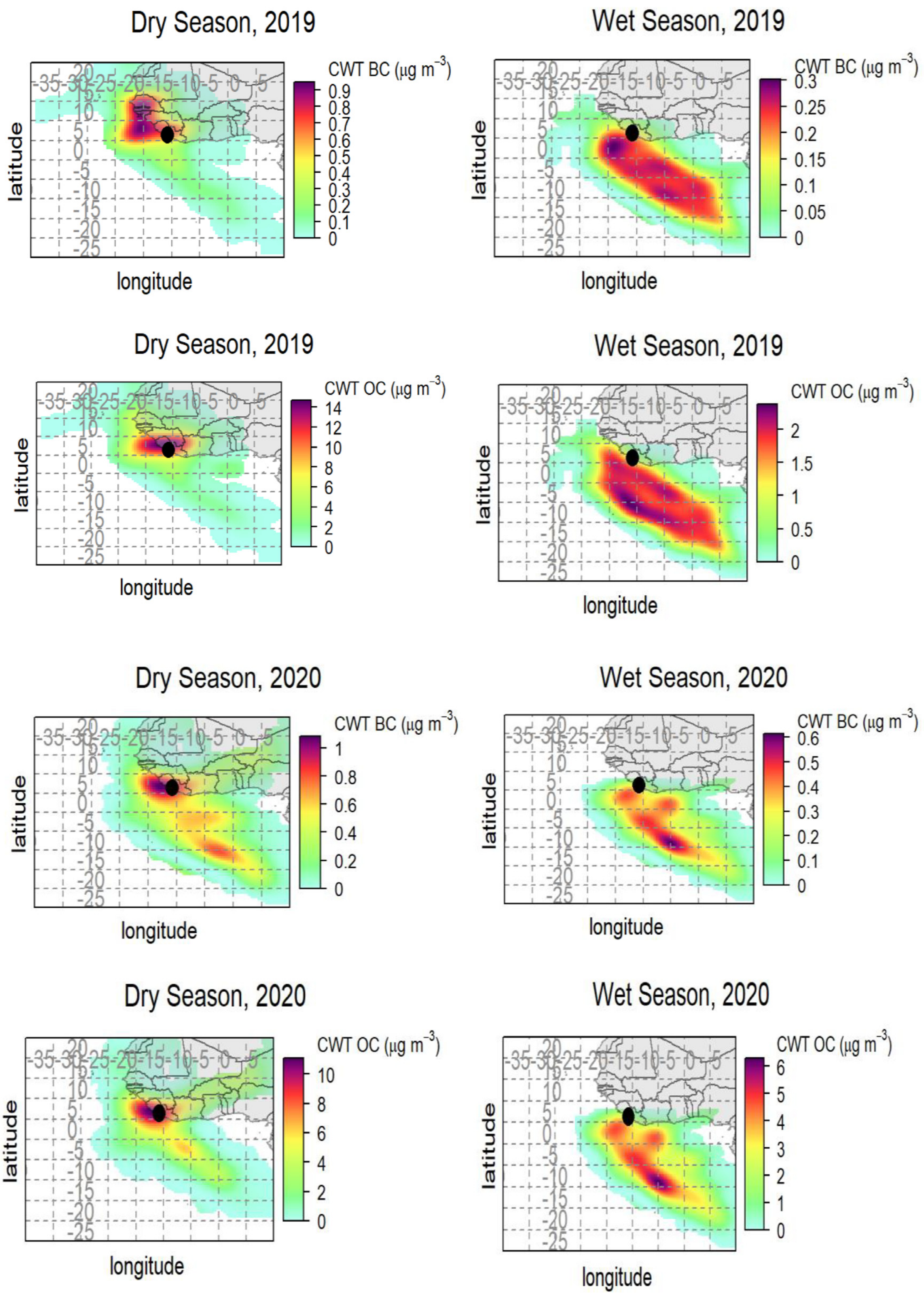


Fig. 9. CWT of BC and OC for two years in seasons dry and wet, respectively. The color block on the right denotes the concentration of CWT BC and OC in µgm^{-3}.

south-east to the south region but did not reach the receptor location, only a moderate amount ranging from BC (0.1 µg m^{-3} - 0.15 µg m^{-3}) and OC (1 µg m^{-3} – 1.5 µg m^{-3}) reached the receptor location in 2019.

These results from both CWT dry season and CWT wet season indicate long-range transportation and marine aerosol contribution to BC and OC in Monrovia. It could further be concluded that an increase in shipping activities could affect the concentration of BC and OC greatly over Monrovia. Black carbon (BC) contributes greatly after CO_2, it is the most climate effect of shipping, representing 7% of total shipping CO_2-eq emissions on a 100-year time scale and 21% of CO_2-eq emissions on a 20-year time scale. Since BC is a short-lived climate pollutant, reducing BC emissions from ships would instantly decrease shipping's climate impacts [16]. In view of Smith et al. [29], roughly 1 billion tonnes of GHS emission (CO_2) during the period of 2007–2012 were accounted for by ships. Guo et al. [15], also found that with the reality that ship emissions were in addition a crucial contributing factor to the BC concentration of Qingdao in past investigations.

Comparison of this study's results with literature

As shown in (Table S4 of SI), the results of BC and OC from this current study compared to literature for Europe, Asia, and West African countries (bordering and neighboring) this study location, results were based on (Mean ± Std. Dev) and (Min-Max), Min -Max was used in the case of Standard deviation is greater than the means. It can be seen from (Table S4 of SI) that results from this study as related to BC concentration for 2019 and 2020 remain below the results of ([10,18,26] and [2]), except for Lamto Jan 91 that this current study in 2019 maximum BC concentration (5.3 µg m^{-3}) was greater compared to that of Lamto (3.9 µg m^{-3}). It can be noted that the above results were influenced by the high industrial activities in those cities as compared to Monrovia in this current study. Regarding OC, it can be indicated that OC concentration is situated below the results of [2] Benin (Cotonou), Ivory Coast (Abidjan), and Pakistan, (Table S4 of SI).

In addition, our analysis revealed that Monrovia was low to moderate polluted in OC and low polluted in BC during the study period. The R^2 value was in the same order of magnitude for BC and OC recorded as (0.98 for 2020 and 0.99 for 2019), suggesting that the emission sources of BC and OC at Monrovia are similar. An investigation by [2], also recorded a linear correlation of (R^2 = 0.9, 0.9, 0.8, and 0.8) in Ivory Coast (Abidjan) and Benin (Cotonou) indicating that emission sources are similar for Ivory Coast (Abidjan), Benin (Cotonou), and this study area.

Conclusion

This study examined the temporal variation, and local, and regional sources that contribute to the enhancement of BC and OC over the atmosphere of Monrovia. BC and OC concentrations show a unique monthly and seasonal trend over Monrovia in both 2019 and 2020. High BC concentration was observed in January for both years, which denotes the dry season, and the minimum BC concentration in June respectively, the wet season. OC concentration varies from high to low obtaining the maximum value in January, February, March, and December in 2019 and January, February, and December in 2020, during the dry season, respectively.

Meanwhile, the minimum value was observed in June during the wet season. The high concentration during these months was due to the harmattan wind season which facilitated the movement of pollutants from the east or the northeast arriving at Monrovia. Diurnal variation suggested a similar trend for 2019 and 2020 respectively, in the wet season. A high concentration of BC and OC shows its peak from 5:00 AM to 7:00 AM and 3:00 PM to 8:00 PM with a slow decline from 7:00 AM and 9:30 PM. This increase in BC and OC concentration during the evening and morning hours strongly indicate a significant increase in traffic emission. Dry season high BC and OC concentration was found associated with hours ranging from 5:00 AM to 7:00 AM, 12:00 PM to 2:00 PM, and 4:00 PM to 7:00 PM in 2019. This high and moderate concentration was justified by traffic emissions and other factors which were discussed in this study.

Spearman correlation analysis shows that meteorological parameters play an important role in BC and OC concentration at Monrovia. OC/BC ratio shows an indication of a local source, long-range source, and secondary aerosols. CBPF plots show that a higher concentration of BC and OC is presented mainly with a wind speed of (∼ 2 m s^{-1} – 5 m s^{-1}) in the South direction which indicated the region of the Atlantic Ocean and the lowest BC concentration was observed with a wind speed to be (∼ 3 m s^{-1} - 5 m s^{-1}) in the North-west direction and moderate BC and OC concentration was associated with a very low wind speed of (∼ 1 m s^{-1} - 2 m s^{-1}) in 2019.

The result of 2019 indicated that BC originated both locally and long-range transported. Low BC concentration was partially distributed in all regions (West, South, East, and North) and was observed to correlate with wind speed to be (∼ 2 ms^{-1} - 5 ms^{-1}) in 2020. In addition, a very high OC concentration was denoted by wind speed (∼ 4 ms^{-1} - 5 ms^{-1}) in the direction of the north-west (freeport of Monrovia) and low concentration in the south-east direction with wind speed ranging (∼ 2 ms^{-1} - 4 ms^{-1}). CWT analysis reveals that the south direction was dominant with a high concentration of BC and OC for the dry season and wet season, some moderate high OC concentration was observed from the northwest direction indicating marine aerosol and ship emission. 2020-CWT results indicated that BC and OC were having strong local emission sources with some long-range source contribution. In a comparison of this study's results with other studies in west Africa and the world at large, it was concluded that Monrovia was low to moderate polluted in OC and low polluted in BC during the study period.

These findings have significant implications for the air quality in Monrovia, as this study is the first of its kind in Monrovia, Liberia. The results of this study will guide the Government in making appropriate policy decisions in order to reduce the OC and BC concentration and thus safeguard human health and the environment. Hence, this study has implications for policy intervention than the engineering application. Subsequent to the policy intervention, engineering devices may be required to effectively implement the policy, which may be explored as a future area of research. This study will also benefit the entire of Liberia and West Africa at large as few studies in the west Africa region studied BC only not OC as in this study. Additionally, implementing proper policy, to help reduce or mitigate pollutants emissions from sources will safeguard achieving one of Africa's Union Agenda for 2063 "*Environmentally sustainable and climate resilient economies and communities*" and also implement the United Nations Sustainable Development Goals (SDGs) for 2030, by "*Integrating climate change measures into national policies, strategies, and planning*".

Author's contribution

Emmanuel Juah Dunbar: Data organizing, Formal analysis, Investigation, Writing-Original Draft. **Lovleen Gupta:** Conceptualization, Formal analysis, Investigation, Supervision, Writing-Review and Editing.

Funding

There is no extend funding attached to this project.

Data availability

The data used in this study can be made available based on a valid request.

Declaration of Competing Interest

There is no conflict of interest declared by any of the authors.

Acknowledgments

The authors also like to acknowledge NASA's Modern-Era Retrospective Analysis, Version 2, and Visual Crossing for their efforts to make this data available for public use.

Web References

Corporation, V. C. (2020). Weather Data & Weather API | Visual Crossing. Visual crossing (2019–2020). Retrieved 9 September 2021, from https://www.visualcrossing.com.

COVID-19 - Response from Liberia | ITUC-AFRICA / CSI-AFRIQUE. (2020). https://www.ituc-africa.org/COVID-19-Response-from-Liberia.html.

MERRA-2. (2020) (Modern-Era Retrospective analysis for Research and Applications, Version 2) (https://disc.gsfc.nasa.gov), xxxxMERRA-2 (Modern-Era Retrospective analysis for Research and Applications, Version 2) (https://disc.gsfc.nasa.gov) [Accessed 13th 08 2021].

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.sciaf.2022.e01540.

References

[1] L. Abbadie, J. Gignoux, M. Lepage, X. le Roux, Environmental constraints on living organisms, Ecol. Stud. 179 (2006) 45–61, doi:10.1007/0-387-33857-8_4.

[2] A.J. Adon, C. Liousse, E.T. Doumbia, A. Baeza-Squiban, H. Cachier, J.F. Léon, V. Yoboué, A.B. Akpo, C. Galy-Lacaux, B. Guinot, C. Zouiten, H. Xu, E. Gardrat, S. Keita, Physico-chemical characterization of urban aerosols from specific combustion sources in West Africa at Abidjan in Côte d'Ivoire and Cotonou in Benin in the frame of the DACCIWA program, Atmos. Chem. Phys. 20 (9) (2020) 5327–5354, doi:10.5194/acp-20-5327-2020.

[3] L.H. Baker, W.J. Collins, D.J.L. Olivié, R. Cherian, A. Hodnebrog, G. Myhre, J. Quaas, Climate responses to anthropogenic emissions of short-lived climate pollutants, Atmos. Chem. Phys. 15 (14) (2015) 8201–8216, doi:10.5194/acp-15-8201-2015.

[4] Bhandari, P. (2022, May 20). How to find and remove outliers. Scribbr. https://www.scribbr.com/statistics/outliers/

[5] S.M. Botsa, T. DLLM, N.S. Magesh, A.K. Tiwari, Characterization of black carbon aerosols over Indian Antarctic station, Maitri and identification of potential source areas, Environ. Sci.: Atmos. 1 (6) (2021) 416–422, doi:10.1039/d1ea00024a.

[6] CEICdata.com. (2018, June 6). Liberia LR: PM2.5 air pollution: mean annual exposure: micrograms per cubic meter. Economic Indicators CEIC. https://www.ceicdata.com/en/liberia/environment-pollution/lr-pm25-air-pollution-mean-annual-exposure-micrograms-per-cubic-meter

[7] X. Chen, Z. Zhang, G. Engling, R. Zhang, J. Tao, M. Lin, X. Sang, C. Chan, S. Li, Y. Li, Characterization of fine particulate black carbon in Guangzhou, a megacity of South China, Atmos. Pollut. Res. 5 (3) (2014) 361–370, doi:10.5094/apr.2014.042.

[8] J.C. Chow, J.G. Watson, Z. Lu, D.H. Lowenthal, C.A. Frazier, P.A. Solomon, R.H. Thuillier, K. Magliano, Descriptive analysis of PM2.5 and PM10 at regionally representative locations during SJVAQS/AUSPEX, Atmos. Environ. 30 (12) (1996) 2079–2112, doi:10.1016/1352-2310(95)00402-5.

[9] S. Cui, J. Xian, F. Shem, L. Zhang, B. Deng, Y. Zhang, X. Ge, One-year real-time measurement of black carbon in the rural area of Qingdao, northeastern China: seasonal variations, meteorological effects, and the COVID-19 case analysis, Atmosphere (Basel) 12 (3) (2021), doi:10.3390/atmos12030394.

[10] E.H.T. Doumbia, C. Liousse, C. Galy-Lacaux, S.A. Ndiaye, B. Diop, M. Ouafo, E.M. Assamoi, E. Gardrat, P. Castera, R. Rosset, A. Akpo, L. Sigha, Real time black carbon measurements in West and Central Africa urban sites, Atmos. Environ. 54 (2012) 529–537, doi:10.1016/j.atmosenv.2012.02.005.

[11] Environmental Protection AgencyLiberia's First Biennial Update Report (October 2020) October, UNFCCC, 2020 https://unfccc.int/sites/default/files/resource/BUR1.pdf.

[12] Environmental Protection Agency. (2020, July). Liberia's NDC Review - July 2020. UNFCCC. https://www.conservation.org/docs/default-source/gef-documents/liberia-ndc-review—july-2020.pdf?sfvrsn=f88fe04_2

[13] S. Fisher, D.C. Bellinger, M.L. Cropper, P. Kumar, A. Binagwaho, J.B. Koudenoukpo, Y. Park, G. Taghian, P.J. Landrigan, Air pollution and development in Africa: impacts on health, the economy, and human capital, Lancet Planet. Health 5 (10) (2021) e681–e688, doi:10.1016/s2542-5196(21)00201-1.

[14] L. Gupta, R. Dev, K. Zaidi, R. Sunder Raman, G. Habib, B. Ghosh, Assessment of PM10 and PM2.5 over Ghaziabad, an industrial city in the Indo-Gangetic Plain: spatio-temporal variability and associated health effects, Environ. Monit. Assess. 193 (11) (2021), doi:10.1007/s10661-021-09411-5.

[15] Q. Guo, M. Hu, S. Guo, Z. Wu, J. Peng, Y. Wu, The variability in the relationship between black carbon and carbon monoxide over the eastern coast of China: BC aging during transport, Atmos. Chem. Phys. 17 (17) (2017) 10395–10403, doi:10.5194/acp-17-10395-2017.

[16] International Council on Clean Transportation. (2017, October). Greenhouse gas emissions from global shipping, 2013–2015. https://theicct.org/sites/default/files/publications/Global-shipping-GHG-emissions-2013-2015_ICCT-Report_17102017_vF.pdf

[17] B. Jereb, B. Gajšek, G. Šlpek, P. Kovše, M. Obrecht, Traffic density-related black carbon distribution: impact of wind in a basin town, Int. J. Environ. Res. Public Health 18 (12) (2021) 6490, doi:10.3390/ijerph18126490.

[18] A.A. Kouassi, M. Doumbia, S. Silue, E.M. Yao, A. Dajuma, M. Adon, N.E. Touré, V. Yoboue, Measurement of atmospheric black carbon concentration in rural and urban environments: cases of Lamto and Abidjan, J. Environ. Prot. (Irvine, Calif) 12 (11) (2021) 855–872, doi:10.4236/jep.2021.1211050.

[19] S.L. Kuzu, E. Yavuz, E. Akyüz, A. Saral, B.O. Akkoyunlu, H. ÖZdemir, G. Demir, A. ÜNal, Black carbon and size-segregated elemental carbon, organic carbon compositions in a megacity: a case study for Istanbul, Air Qual., Atmos. Health 13 (7) (2020) 827–837, doi:10.1007/s11869-020-00839-1.

[20] C.D. Navinya, V. Vinoj, S.K. Pandey, Evaluation of PM2.5 surface concentrations simulated by NASA's MERRA Version 2 aerosol reanalysis over India and its relation to the air quality index, Aerosol. Air Qual. Res. 20 (6) (2020) 1329–1339, doi:10.4209/aaqr.2019.12.0615.

[21] T. Novakov, S. Menon, T.W. Kirchstetter, D. Koch, J.E. Hansen, Aerosol organic carbon to black carbon ratios: analysis of published data and implications for climate forcing, J. Geophys. Res. 110 (D21) (2005), doi:10.1029/2005jd005977.

[22] PM2.5 pollution, population exposed to levels exceeding WHO Interim Target-1 value (% of total) - Country Ranking. (2017). Index Mundi. Retrieved 4 March 2022, from https://www.indexmundi.com/facts/indicators/EN.ATM.PM25.MC.M3/rankings

[23] V. Prabhu, A. Soni, S. Madhwal, A. Gupta, S. Sundriyal, V. Shridhar, V. Sreekanth, P.S. Mahapatra, Black carbon and biomass burning associated high pollution episodes observed at Doon valley in the foothills of the Himalayas, Atmos. Res. 243 (2020) 105001, doi:10.1016/j.atmosres.2020.105001.

[24] T. Rajesh, S. Ramachandran, Assessment of the coronavirus disease 2019 (COVID-19) pandemic-imposed lockdown and unlock effects on black carbon aerosol, its source apportionment, and aerosol radiative forcing over an urban city in India, Atmos. Res. 267 (2022) 105924, doi:10.1016/j.atmosres.2021.105924.

[25] K. Ropkins, D. Carslaw, openair - data analysis tools for the air quality community, R J 4 (1) (2012) 20, doi:10.32614/rj-2012-003.

[26] S. Ruellan, H. Cachier, Characterisation of fresh particulate vehicular exhausts near a Paris high flow road, Atmos. Environ. 35 (2) (2001) 453–468, doi:10.1016/s1352-2310(00)00110-2.

[27] S. Saarikoski, H. Timonen, K. Saarnio, M. Aurela, L. Järvi, P. Keronen, V.M. Kerminen, R. Hillamo, Sources of organic carbon in fine particulate matter in northern European urban air, Atmos. Chem. Phys. 8 (20) (2008) 6281–6295, doi:10.5194/acp-8-6281-2008.

[28] A.K. Singh, A. Srivastava, Seasonal variation of carbonaceous species in PM1 measured over residential area of Delhi, India, SN Appl. Sci. 2 (12) (2020), doi:10.1007/s42452-020-03854-0.

[29] Smith, T.W.P., Jalkanen, J.P., Anderson, B.A., Corbett, J.J., Faber, J., Hanayama, S., … & Pandey, A. (2015). Third IMO Greenhouse Gas Study 2014. Retrieved from http://www.imo.org/en/OurWork/Environment/PollutionPrevention/AirPollution/Documents/Third%20Greenhouse%20Gas%20Study/GHG3%20Executive%20 Summary%20and%20Report.pdf

[30] S. Sooktawee, T. Kanabkaew, S. Boonyapitak, A. Patpai, N. Piemyai, Characterising particulate matter source contributions in the pollution control zone of mining and related industries using bivariate statistical techniques, Sci. Rep. 10 (1) (2020), doi:10.1038/s41598-020-78445-5.

[31] C.W. Stjern, B.H. Samset, G. Myhre, P.M. Forster, I. Hodnebrog, T. Andrews, O. Boucher, G. Faluvegi, T. Iversen, M. Kasoar, V. Kharin, A. Kirkevåg, J.F. Lamarque, D. Olivié, T. Richardson, D. Shawki, D. Shindell, C.J. Smith, T. Takemura, A. Voulgarakis, Rapid adjustments cause weak surface temperature response to increased black carbon concentrations, J. Geophys. Res.: Atmos. 122 (21) (2017) 11,462-11,481, doi:10.1002/2017jd027326.

[32] Sunitha, L., BalRaju, M., Sasikiran, J., & Ramana, E.V. (2014, June). Automatic outlier identification in data mining using IQR in real-time data. https://ijarcce.com/wp-content/uploads/2012/03/IJARCCE8E-a-sunitha-Automatic-Outlier-Identification.pdf

[33] T. Takemura, K. Suzuki, Weak global warming mitigation by reducing black carbon emissions, Sci. Rep. 9 (1) (2019), doi:10.1038/s41598-019-41181-6.

[34] W. Tefera, A. Kumie, K. Berhane, F. Gilliland, A. Lai, P. Sricharoenvech, J. Samet, J. Patz, J.J. Schauer, Chemical characterization and seasonality of ambient particles (PM2.5) in the City Centre of Addis Ababa, Int. J. Environ. Res. Public Health 17 (19) (2020) 6998, doi:10.3390/ijerph17196998.

[35] I. Uria-Tellaetxe, D.C. Carslaw, Conditional bivariate probability function for source identification, Environ. Modell. Softw. 59 (2014) 1–9, doi:10.1016/j.envsoft.2014.05.002.

[36] H.P. Vinutha, B. Poornima, B.M. Sagar, Detection of outliers using interquartile range technique from intrusion dataset, Adv. Intell. Syst. Comput. (2018) 511–518, doi:10.1007/978-981-10-7563-6_53.

[37] J. Wang, A. Yu, L. Yang, C. Fang, Research on organic carbon and elemental carbon distribution characteristics and their influence on fine particulate matter (PM2.5) in Changchun City, Environments 6 (2) (2019) 21, doi:10.3390/environments6020021.

[38] World Health OrganizationReducing Global Health Risks Through Mitigation of Short-Lived Climate Pollutants. Scoping Report for Policy-Makers, World Health Organization, 2015 https://apps.who.int/iris/handle/10665/189524.

[39] Y. Zhang, J. He, S. Zhu, B. Gantt, Sensitivity of simulated chemical concentrations and aerosol-meteorology interactions to aerosol treatments and biogenic organic emissions in WRF/Chem, J. Geophys. Res.: Atmos. 121 (10) (2016) 6014–6048, doi:10.1002/2016jd024882.

[40] N. Zioła, B. Błaszczak, K. Klejnowski, Temporal variability of equivalent black carbon components in atmospheric air in southern Poland, Atmosphere (Basel) 12 (1) (2021) 119, doi:10.3390/atmos12010119.

The Awkward World of Python and C++

Manasvi Goyal[1], **Ianna Osborne**[2], **Jim Pivarski**[2]

[1] Delhi Technological University, Delhi, India

[2] Princeton University, Princeton, NJ 08544, USA

E-mail: `mg.manasvi@gmail.com`

Abstract. There are undeniable benefits of binding Python and C++ to take advantage of the best features of both languages. This is especially relevant to the HEP and other scientific communities that have invested heavily in the C++ frameworks and are rapidly moving their data analyses to Python. Version 2 of Awkward Array, a Scikit-HEP Python library, introduces a set of header-only C++ libraries that do not depend on any application binary interface. Users can directly include these libraries in their compilation rather than linking against platform-specific libraries. This new development makes the integration of Awkward Arrays into other projects easier and more portable as the implementation is easily separable from the rest of the Awkward Array codebase. The code is minimal, it does not include all of the code needed to use Awkward Arrays in Python, nor does it include references to Python or pybind11. The C++ users can use it to make arrays and then copy them to Python without any specialized data types - only raw buffers, strings, and integers. This C++ code also simplifies the process of just-in-time (JIT) compilation in ROOT. This implementation approach solves some of the drawbacks, like packaging projects where native dependencies can be challenging. In this paper, we demonstrate the technique to integrate C++ and Python by using a header-only approach. We also describe the implementation of a new LayoutBuilder and a GrowableBuffer. Furthermore, examples of wrapping the C++ data into Awkward Arrays and exposing Awkward Arrays to C++ without copying them are discussed.

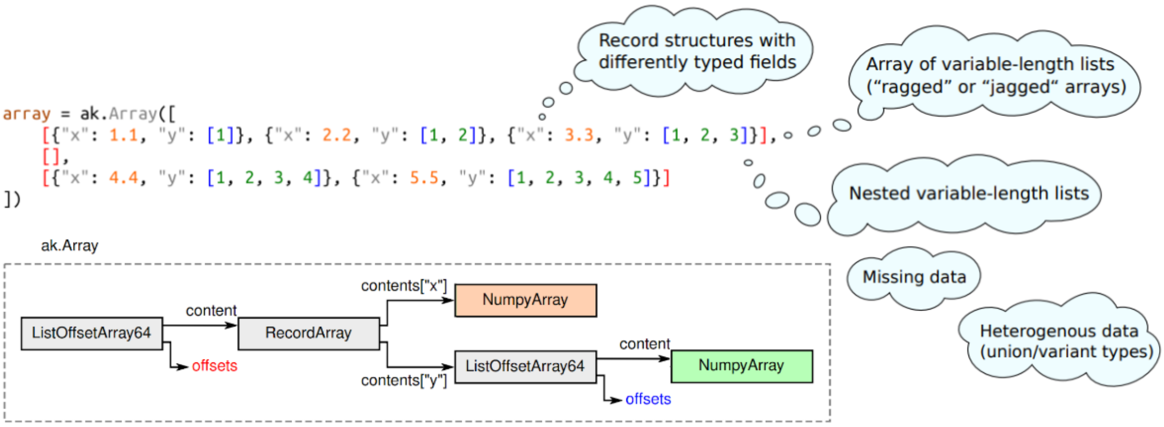
1. Introduction

Awkward Array [1] is an important tool for physics analysis in Python for High Energy Physics (HEP) community. It is a part of the Scikit-HEP [2] ecosystem. Nested, variable-length lists (”ragged” or ”jagged” arrays), records with differently typed fields, missing data, and other heterogeneous data (union/variant types) can be defined as a set of primitives using NumPy-like [3] phrases in Python [4]. In Awkward arrays, a single, user-facing `ak.Array` consists of one small tree with large, contiguous data buffers attached to each node [5], as shown in Figure 1. Compiled operations are performed on these data buffers, not the objects they represent.

In this work, we present new tools for creating Awkward Arrays in C++. Previously, the main codebase was written in C++ [6] with the idea that downstream code would link to libawkward.so, but that route is full of hidden issues [7]. The method of a small, header-only library that only fills array buffers for downstream code to pass from C++ to Python using C-types only has considerably more promise.

2. Python-C++ Integration

Nowadays, more front-end users use Python [8], but large-scale processing still needs to have high performance of C++ [7]. That is why we combine Python and C++ to take advantage



```
array = ak.Array([
    [{"x": 1.1, "y": [1]}, {"x": 2.2, "y": [1, 2]}, {"x": 3.3, "y": [1, 2, 3]}],
    [],
    [{"x": 4.4, "y": [1, 2, 3, 4]}, {"x": 5.5, "y": [1, 2, 3, 4, 5]}]
])
```

Record structures with differently typed fields
Array of variable-length lists ("ragged" or "jagged" arrays)
Nested variable-length lists
Missing data
Heterogenous data (union/variant types)
ak.Array
ListOffsetArray64
content
offsets
RecordArray
contents["x"]
NumpyArray
contents["y"]
ListOffsetArray64
content
NumpyArray
offsets

Figure 1. Structure of an Awkward Array with nested variable-length lists and records, color-coded with an array example.

of the best features of both languages so that we can have a Python user interface and, at the same time, take advantage of the performance and memory management of C++. HEP and other scientific communities have extensively invested in the C++ frameworks and are swiftly migrating their data analyses to Python. These communities are particularly interested in bridging the gap between the two languages [8]. This raises an important question: *'How to do Python-C++ integration the right way?'*, which is addressed in the following sections.

3. The 'Header-Only' Approach

A set of header-only C++ libraries has been introduced to address the issues in the Python-C++ integration in Awkward Arrays [7]. These templated C++ libraries are not dependent on any application binary interface (ABI). They can be directly included in a project's compilation without the need to link against platform-specific libraries. This 'header-only' approach not only simplifies the production of Awkward Arrays in a project but also enhances the portability of the Awkward Arrays. The code is minimal and does not constitute all of the code required to use Awkward Arrays in Python. It contains no references to Python or Python bindings. The header files can be used by C++ users to create Awkward Arrays, which can then be copied into Python without any specialized data types - only raw buffers, strings, and integers. This approach addresses the issue of packaging projects with native dependencies.

4. LayoutBuilder

A 'layout' consists of composable elements that determine how an array is structured. It can only build a specific view determined by the layout Form. LayoutBuilder [9] is a set of compile time, templated static C++ classes implemented entirely in a header-only library. It uses a header-only GrowableBuffer (Figure 2), which is implemented as a linked list with smart pointers. `awkward::LayoutBuilder` specializes an Awkward data structure using C++ templates, which can be filled and converted to a Python Awkward Array through `ak.from_buffers`. The data comes out of LayoutBuilder as a set of named buffers and a JSON [10] Form. The Form is a unique description of an Awkward Array and returns a `std::string` that tells Awkward Array how to put everything together. LayoutBuilder is part of an awkward-cpp package that is separate from an awkward package. Both packages are individually pip-installable. The code doesn't have helper methods to pass the data to Python, so different projects can use different binding generators. The code relies on generalized lambda expressions to deduce parameter type during compile time that is available from the C++14 standard.

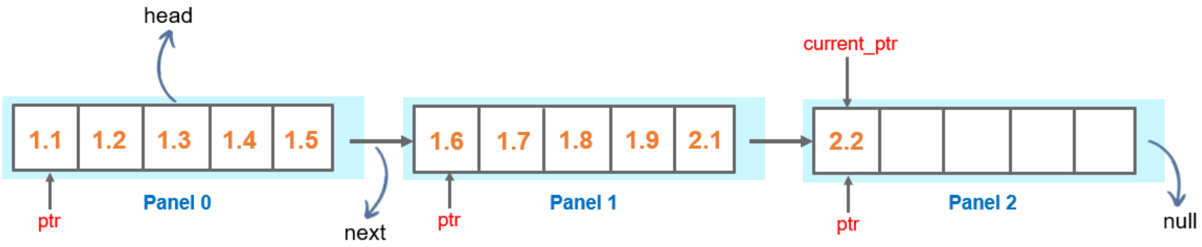


Figure 2. Awkward Array GrowableBuffer implemented as a linked list with multiple panels, each of size = 5, that are allocated as needed, i.e., when the GrowableBuffer runs out of space.

ArrayBuilder [11] and LayoutBuilder are both used to create Awkward Arrays. The main difference between a LayoutBuilder and an ArrayBuilder is that the data types that can be appended to the LayoutBuilder are defined in advance, while any data types can be appended to an ArrayBuilder. LayoutBuilder is designed to build Awkward Arrays faster. The flexibility of ArrayBuilder comes with performance limitations since it needs to discover the data type, while LayoutBuilder knows it in advance.

5. User Interface of LayoutBuilder

This section explains the user interface of LayoutBuilder with the help of an example of an Awkward Array with nested records and variable-length lists.

5.1. Phases of LayoutBuilder

There are three phases of using LayoutBuilder:

(i) **Constructing a LayoutBuilder:** from variadic templates (It is an implicit template instantiation).

(ii) **Filling the LayoutBuilder:** while repeatedly walking over the raw pointers.

(iii) **Taking the data out to user-allocated buffers:** then, the user can pass them to Python.

5.2. Illustrative Example

An example of RecordBuilder is illustrated in Listing 1. The first step is to include the LayoutBuilder header file (see [9] for the installation instructions). Next, the RecordBuilder is constructed with variadic templates. The contents of a RecordBuilder are heterogeneous type containers (`std::tuple`) that take the other Builders as the template parameters. The field names are non-type template parameters defined by the user. Currently, it is not possible to template on strings as this functionality comes only from C++20 and onwards. Therefore, for passing the field names as template parameters to the RecordBuilder, a user-defined `field_map`, with enumerated type field ID as keys and the field names as value, has to be provided. In the case of multiple RecordBuilder, a user-defined map has to be specified for each of the RecordBuilder used.

After that, the LayoutBuilder buffers are filled with the required data as shown in Listing 1. To make sure there are no errors while filling these buffers, the user can check their validity by using the `is_valid()` method, which can be called on every entry if they want to trade safety for speed. The example translates into the following Awkward Array in Python:

```
[{"x": 1.1, "y": [1]}, {"x": 2.2, "y": []}, {"x": 3.3, "y": [1, 2]},]
```

```cpp
#include "awkward/LayoutBuilder.h"

enum Field : std::size_t {x, y};
UserDefinedMap fields_map({
    {Field::x, "x"},
    {Field::y, "y"}});

// Constructing a LayoutBuilder from variadic templates!
RecordBuilder<
    RecordField<Field::x, NumpyBuilder<double>>,
    RecordField<Field::y, ListOffsetBuilder<int64_t, NumpyBuilder<int32_t>>>
> builder(fields_map);

auto& x_builder = builder.field<Field::x>();
auto& y_builder = builder.field<Field::y>();

// Filling the LayoutBuilder
x_builder.append(1.1);
auto& y_subbuilder = y_builder.begin_list();
y_subbuilder.append(1);
y_builder.end_list();

x_builder.append(2.2);
y_builder.begin_list();
y_builder.end_list();

x_builder.append(3.3);
y_builder.begin_list();
y_subbuilder.append(1);
y_subbuilder.append(2);
y_builder.end_list();
```

Listing 1: Example of a LayoutBuilder with nested records and variable-length lists.

We want NumPy to own the array buffers so that they get deleted when the Awkward Array goes out of Python scope, not when the LayoutBuilder goes out of C++ scope. The hand-off, therefore, needs a few steps:

(i) Retrieve the set of buffer names and their sizes (as a number of bytes).

```cpp
std::map<std::string, size_t> names_nbytes = {};
builder.buffer_nbytes(names_nbytes);
```

(ii) Allocate memory for these buffers in Python with `np.empty(nbytes, type = np.uint8)` and get `void*` pointers to these buffers by casting the output of `numpy_array.ctypes.data`.

(iii) Let the LayoutBuilder fill these buffers.

```cpp
std::map<std::string, void*> buffers;
builder.to_buffers(buffers);
```

(iv) Finally, JSON Form is generated with:

```cpp
std::string form = builder.form();
```

The Form generated for the example in Listing 1 is shown in Listing 2. Now, everything can be passed over the border from C++ to Python using pybind11's [12] `py::buffer_protocol` for the buffers, as well as an integer for the length and a string for the Form. If the user ever needs to make a change in the format of the records (add, remove, rename, or change the field type), there is no need to change anything in the Python-C++ interface. All of that is contained in the specialization of the C++ template and the filling procedure, which are both in the C++ code.

```
{"class": "RecordArray",
 "contents": {
     "x": {"class": "NumpyArray",
           "primitive": "float64",
           "form_key": "node1"},
     "y": {"class": "ListOffsetArray",
           "offsets": "i64",
           "content": {
               "class": "NumpyArray",
               "primitive": "int32",
               "form_key": "node3"},
          "form_key": "node2"}" },
 "form_key": "node0"}
```

Listing 2: Awkward Array Form for the example in Listing 1.

6. Applications

The header-only approach allows for multiple applications in both static and dynamic projects. Awkward RDataFrame [13] uses the C++ header-only libraries to simplify the process of just-in-time (JIT) compilation in ROOT [14]. The `ak.from_rdataframe` [15] function converts the selected ROOT RDataFrame [16] columns as native Awkward Arrays. The templated header-only implementation constructs the Form from the primitive data types [17]. The generation of all the types via templates makes it easier to dynamically generate LayoutBuilder from strings in Python and then compile it with Cling [18].

Another application of header-only LayoutBuilder could be in the ctapipe [19] project, which is currently in the planning stage. ctapipe is a framework for prototyping the low-level data processing algorithms for the Cherenkov Telescope Array [20]. It does some processing on structured ("awkward") event data, and the developers want to refactor their implementation to use Awkward Array. They already have C++ code that iterates over the custom file format, which has array types that are known at compile-time. The easiest way to Awkward Arrays in this project is to use a LayoutBuilder to fill the buffers and then send them to Python through pybind11.

7. Conclusion

The header-only approach presented in this paper facilitates Awkward Arrays Python-C++ integration and enhances their portability. A set of templated header-only libraries use only C-types (integers, strings, and raw buffers) to build Awkward Arrays and send them to Python by generating a JSON Form. A standalone awkward header-only C++ package opens up the doors for users to analyze their data in Python. Awkward Arrays can be seamlessly integrated with external projects without linking against platform-specific libraries or worrying about native dependencies. This new development also allows the extension of the use cases of Awkward Arrays to scientific communities beyond HEP.

8. Acknowledgment

This work is supported by NSF cooperative agreement OAC-1836650 (IRIS-HEP) and NSF cooperative agreement PHY-2121686 (US-CMS LHC Ops).

References

[1] Pivarski J, Osborne I, Ifrim I, Schreiner H, Hollands A, Biswas A, Das P, Roy Choudhury S, Smith N and Goyal M 2018 Awkward Array [Computer software] *Zenodo* `https://doi.org/10.5281/zenodo.4341376`

[2] Rodrigues E 2019 The Scikit-HEP Project *EPJ Web Conf.* **214** 06005 `DOI:10.1051/epjconf/201921406005`

[3] Harris C R, Millman K J, van der Walt S J et al. 2020 Array programming with NumPy *Nature* **585** 357–362. `DOI: 10.1038/s41586-020-2649-2`

[4] Pivarski J, Nandi J, Lange D and Elmer P, 2019 Columnar data processing for HEP analysis *EPJ Web Conf.* **214** 06026 `DOI: 10.1051/epjconf/201921406026`

[5] Pivarski J, Elmer P and Lange D 2020 Awkward Arrays in Python, C++, and Numba *EPJ Web Conf.* **245** 05023 `DOI: 10.1051/epjconf/202024505023`

[6] Pivarski J, Osborne I, Das P, Biswas A and Elmer P 2020 Awkward Array: JSON-like data, NumPy-like idioms *Proc. of the 19th Python in Science Conf. (SCIPY 2020)* 78-84

[7] Pivarski J 2021 Lessons learned in Python-C++ integration *20th International Workshop on Advanced Computing and Analysis Techniques in Physics Research* `https://indi.to/N69ds`

[8] Pivarski J, Rodrigues E, Pedro K, Shadura O, Krikler B and Stewart G A 2022 HL-LHC Computing Review Stage 2, Common Software Projects: Data Science Tools for Analysis *[arXiv 2202.02194]* `DOI: 10.48550/arXiv.2202.02194`

[9] User Guide: How to Use Header-Only LayoutBuilder in C++ *Awkward Array Documentation* `https://awkward-array.org/doc/main/user-guide/how-to-use-header-only-layoutbuilder.html`

[10] JSON `https://www.json.org/`

[11] ArrayBuilder `https://awkward-array.org/doc/main/reference/generated/ak.ArrayBuilder.html`

[12] Jakob W, Rhinelander J and Moldovan D 2016 pybind11 - Seamless operability between C++11 and Python. `https://github.com/pybind/pybind11`

[13] Osborne I and Pivarski J 2022 Awkward RDataFrame Tutorial *PyHEP 2022 (virtual) Workshop* `https://doi.org/10.5281/zenodo.7081586`

[14] Brun R, Rademakers F, Canal P, Naumann A, Couet O, Moneta L, Vassilev V, Linev S, Piparo D, GANIS G, Bellenot B, Guiraud E, Amadio G, wverkerke, Mato P, TimurP, Tadel M, wlav, Tejedor E, Blomer J, Gheata A, Hageboeck S, Roiser S, marsupial, Wunsch S, Shadura O, Bose A, CristinaCristescu, Valls X and Isemann R 2019 root-project/root: v6.18/02 (v6-18-02) *Zenodo* `https://doi.org/10.5281/zenodo.3895860`

[15] from_rdataframe `https://awkward-array.org/doc/main/reference/generated/ak.from_rdataframe.html`

[16] Piparo D, Canal P, Guiraud E, Pla X V, Ganis G, Amadio G, Naumann A and Tejedor E 2018 RDataFrame: Easy Parallel ROOT Analysis at 100 Threads *EPJ Web Conf.* **214** 06029 `DOI: 10.1051/epjconf/201921406029`

[17] Osborne I and Pivarski J 2022 Awkward to RDataFrame and back *[arXiv 2302.09860]* `DOI: 10.48550/arXiv.2302.09860`

[18] Cling [Online]. `https://root.cern.ch/cling`

[19] ctapipe documentation `https://ctapipe.readthedocs.io/en/latest/`

[20] ctapipe observatory `https://www.cta-observatory.org/`

Thermal-hydraulic Behaviour of Corrugated Pipe Configurations

Article *in* Journal of Polymer & Composites · November 2022

Gaurav Kumar
Delhi Technological University

Raj Kumar Singh
Delhi Technological University

STM JOURNALS

Journal of
Polymer & Composites

ISSN: 2321-2810 (Online)
ISSN: 2321-8525 (Print)
Volume 10, Special Issue 1, 2022
DOI (Journal): 10.37591/JoPC

http://engineeringjournals.stmjournals.in/index.php/JoPC/index

Research JoPC

Thermal-hydraulic Behaviour of Corrugated Pipe Configurations

Aryan Tyagi[1], Gaurav Kumar[2,*], Raj Kumar Singh[3]

Abstract

Heat exchangers have a wide range of applications and can be found in a variety of different industries, including chemical reactors, nuclear power plants, and solar energy generation. It is required to do thermal optimization on heat exchangers in order to decrease the size, cost, and energy requirements of the heat exchangers while simultaneously improving their capacity to transmit heat. In these types of businesses, heat exchangers frequently make use of corrugated pipes as piping material. They are put to use in the process of transferring heat across fluids that are maintained at varying temperatures. The flow of corrugated pipe was investigated by using different corrugated ring diameters, maintaining a corrugated ring angle of 360°, and keeping the spacing between the corrugated rings at 7.5 mm. At a Reynolds number of 5000, the focus of the study was on determining how the diameter of the ring effects the thermo-hydraulic flow of water. Utilizing various forms of numerical simulation, this study determined the thermo-hydraulic flow behaviour and the enhancement of heat transfer of fluid that was moving through a corrugated pipe while a constant heat flux was present. An enhancement in heat transfer coefficient can be confirmed because increase in turbulent kinetic energy was seen. There has been discussion over the impact that the diameter of the corrugated ring has on the behaviour of variables such as velocity, pressure, radial velocity, axial velocity, temperature distribution, and turbulent kinetic energy (TKE).

Keywords: Heat transfer, corrugated pipe, CFD, heat exchanger, thermal enhancement

INTRODUCTION

Heat exchangers are used in a number of different industries, such as solar energy, nuclear power production and chemical reactors [1–5]. Thermal optimization of heat exchangers is necessary to reduce the size, cost, and energy requirements, while at the same time enhancing their heat transfer capabilities. The use of corrugated pipes in heat exchangers is quite common in these industries. They are used to transfer heat between fluids kept at different temperatures [6–9]. Thermal enhancement of heat exchangers can be done in two ways, namely, active and passive techniques. Active methods require external power sources to produce surface vibration or to drive a mechanical device. Passive methods, as the name suggests, do not require any external power to enhance the performance of the heat exchanger. Passive techniques use extended or rough surfaces, or fluid additives to improve the heat transfer between the fluids. The use of corrugated pipes comes under the passive type of thermal enhancement. Corrugated pipes have been used in many applications to improve the performance of heat exchangers [10–12].

***Author for Correspondence**
Gaurav Kumar

[1]Student, Department of Mechanical Engineering, Delhi Technological University, Delhi, India
[2]Scholar, Department of Mechanical Engineering, Delhi Technological University, Delhi, India
[3]Professor, Department of Mechanical Engineering, Delhi Technological University, Delhi, India

Received Date: August 27, 2022
Accepted Date: November 24, 2022
Published Date: November 30, 2022

Citation: Aryan Tyagi, Gaurav Kumar, Raj Kumar Singh. Thermal-hydraulic Behavior of Corrugated Pipe Configurations. Journal of Polymer & Composites. 2022; 10(Special Issue 1): S108–S114.

Many numerical and experimental studies on corrugated pipes show their capabilities of enhancing heat transfer. Some Researchers performed experiments using active and passive techniques to improve heat exchanger performance. In their study, corrugated copper pipes with diameter 10 mm and length 3200 mm were used. They studied the thermal performance for different experimental conditions and the drop in pressure for different flow rates. They concluded that rotation of the tubes resulted in a greater pressure drop across the pipes [13]. Some have performed studies that showed a reduction in energy consumption and cost, while improving performance ratings and equipment life. They examined the effects of corrugated rib roughened pipes on frictional losses and heat transfer. The Reynolds number in their study ranged between 7500 and 50,000. The study showed a direct correlation between the number of gaps in the pipe and the Reynolds number of the flow. Increasing the number of gaps with increasing Reynolds number resulted in the increase of friction loss and heat transfer. They also observed an increase in Nusselt number, and the friction factor [14].

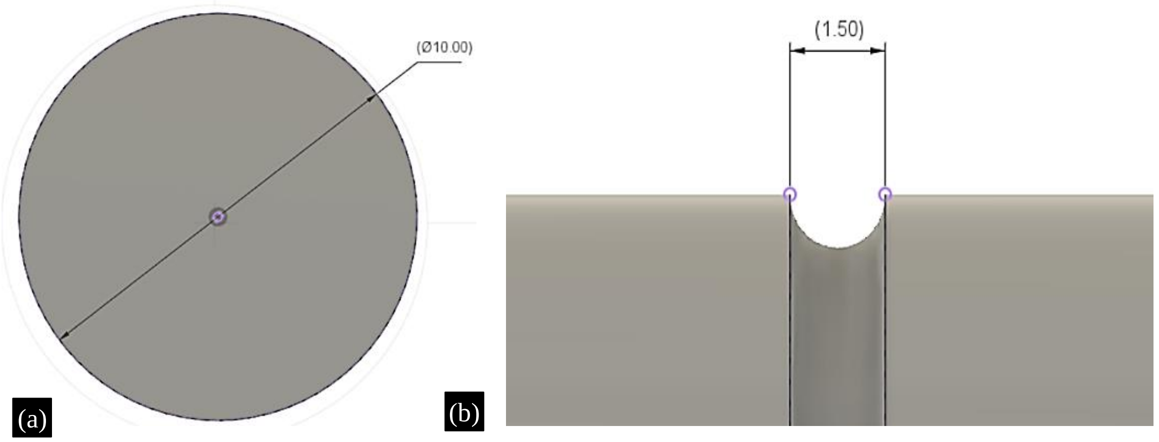
Numerical and experimental studies showed the corrugation angles and their effect on thermal performance of wavy heat exchangers. Their studies were performed in two-dimensional flows. They achieved a performance factor of 1.8 at a corrugation angle of 100°, and a Reynolds number of 1000. They observed a change from transition flow to turbulent flow in the channel, with an increase in the corrugation angle [15]. A study developed a three-dimensional model on ANSYS, studying the heat transfer in helically corrugated pipes. Their simulations resulted in an improvement in heat transfer performance as well as Nusselt number. The rotational flow showed little effect on the Nusselt number and heat transfer.

In this study, a three-dimensional Finite Volume Method (FVM) model was developed on ANSYS Fluent. Corrugated pipe flows were studied, with varying corrugated ring diameters (1.5, 2.5, and 3.5 mm), with corrugated ring angle of 360° and the distance between the corrugated rings being constant at 7.5 mm. The study was focused on how the ring diameter influences the thermo-hydraulic flow of water at the Reynolds number of 5000.

In the results, the effect of the corrugated ring diameter on the behaviour of velocity, pressure, radial velocity, axial velocity, temperature distribution, and turbulent kinetic energy (TKE) has been discussed. The results of these simulations can be used for further study and design for the enhancement of heat transfer corrugated pipe flows.

MODEL DESCRIPTION

The three-dimensional models for the corrugated pipes with different corrugated ring diameters were created in Fusion 360. The CAD models with the parameters are shown in Figure 1. The pipe length and diameter were 1000 and 10 mm, respectively.



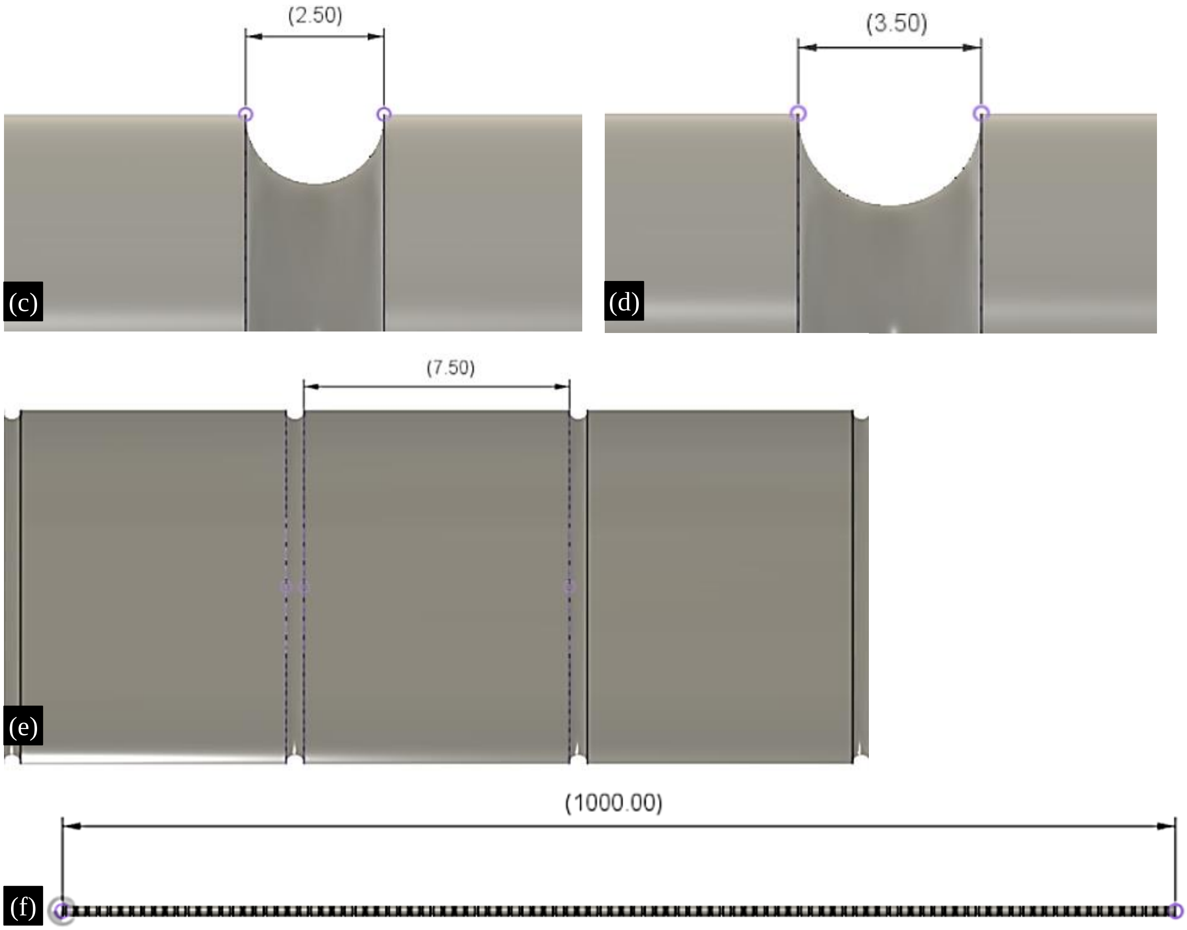


Figure 1. Geometrical Parameters of Corrugated Pipes. (a) Pipe diameter = 10 mm; (b) Corrugated Ring Diameter = 1.5 mm; (c) Corrugated Ring Diameter = 2.5 mm; (d) Corrugated Ring Diameter = 3.5 mm; (e) Gap between each ring is 7.5 mm; (f) Length of the entire pipe is 1000 mm.

There was a total of four geometries for the corrugated pipes used in this study. Each geometry had a corrugated ring angle of 360°, distance between rings is 7.5 mm and the ring diameter ranges from 1.5 to 3.5 mm, with increments of 1 mm. The last geometry was the smooth pipe, without any corrugations, for the purpose of comparison.

Numerical Model

For the numerical analysis, Computational Fluid Dynamics (CFD) software ANSYS Fluent was used. Fluent is a Finite Volume Method (FVM) solver, which solves for pressure, flow velocity, and heat transfer in this study. The k-ϵ turbulence model is applied to this simulation, as achieving convergence is relatively simple using this model. This turbulence model is extensively used in academia and industry for research and development purposes. The SIMPLE algorithm is applied as the pressure-velocity coupling scheme in the momentum equation. The residuals are kept lower than 10–4 for pressure, velocity, energy, and continuity.

Grid Generation

The discretization software ANSYS Mechanical was used to generate the mesh for the flow simulation. The mesh generation is a crucial step as it determines the computational cost as well as the accuracy of the solution. Usually, an ordered mesh containing hexahedral elements are preferred as they provide a more accurate solution, but the corrugated pipe is a complex geometry. For such complex geometries, a grid containing tetrahedral elements, as shown in Figure 2(a), is easier to generate, and also requires lesser time.

The disturbances in the boundary layer are amplified throughout the flow, and these small disturbances add up to larger effects on velocity, TKE, and pressure distributions. Hence, it is important to capture these disturbances in detail. Inflation layers are added to capture the crucial boundary layer effects through the pipe as shown in Figure 2(b).

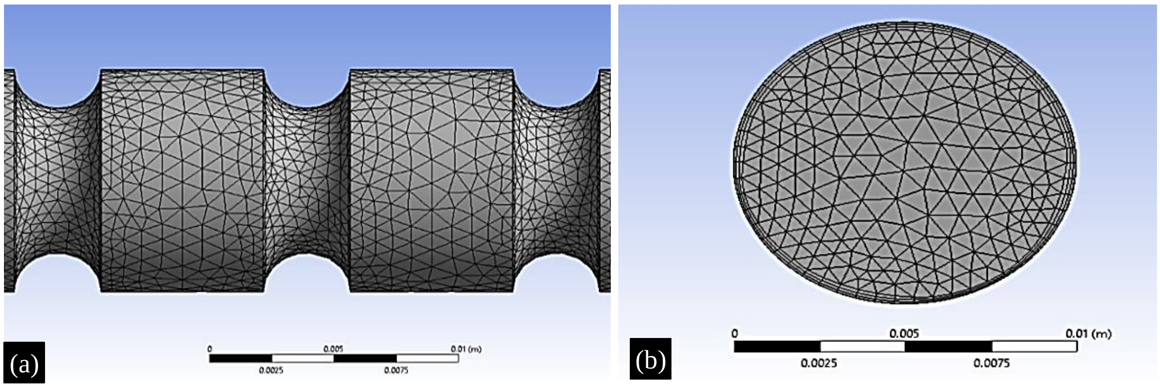


Figure 2. Final Mesh Generated. (a)Tetrahedral elements on pipe mesh. (b) Inflation layers.

Governing Equations

The conservation equations for the fluid flow through a corrugated pipe can be written as shown:

Continuity equation:

$$\frac{\partial}{\partial x_i}(\rho u_i) = 0 \tag{1}$$

Momentum equation:

$$\frac{\partial}{\partial x_i}(\rho u_i u_j) = \frac{-\partial p}{\partial x_i} + \frac{\partial}{\partial c_j}\left[\mu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) - \frac{2}{3}\mu\frac{\partial u_i}{\partial x_i} + \delta_{ij}\right] \tag{2}$$

Energy equation:

$$\frac{\partial}{\partial x_i}(u_i(\rho E + p)) = \frac{-\partial}{\partial x_i}\left[\left(\lambda + \frac{C_P}{\mathrm{Pr}_t}\right) - \frac{\partial T}{\partial x_j} + \mu(T_{ij})_{eff}\right] \tag{3}$$

Where, ρ, u, μ, p, λ, Cp and T represent fluid density, velocity flow, dynamic flow viscosity, pressure, thermal conductivity, specific heat, and temperature.

Flow Parameters and Boundary Conditions

The no-slip condition was considered on the walls of the pipe. The working fluid for the simulations was water. The physical properties include a density of 998.2 kg/m^3, specific heat (Cp) of 4.18 J/kg.K, dynamics viscosity (μ) of 0.001003 kg/(ms) and thermal conductivity (λ) of 0.6 W/kg.K. Temperature at inlet was 298.15 K. The Reynolds number at the inlet was specified as 5000. Boundary conditions are inlet velocity with inlet temperature, and outlet pressure at zero static pressure (Po).

RESULTS AND DISCUSSIONS

Three models having different corrugated ring diameters (CRD) (1.5, 2.5, and 3.5 mm) were investigated. Figure 3 represents the static pressure along the length of the pipe. As can be observed from the figure, the pressure decreases substantially as the length increases. The flow field is divided into three main regions. The inlet area has a higher pressure, the middle area has a lower pressure, and finally, the outlet area has the lowest pressure.

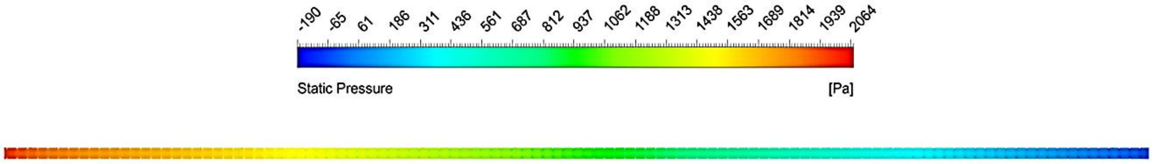


Figure 3. Static pressure-drop across the length of the pipe.

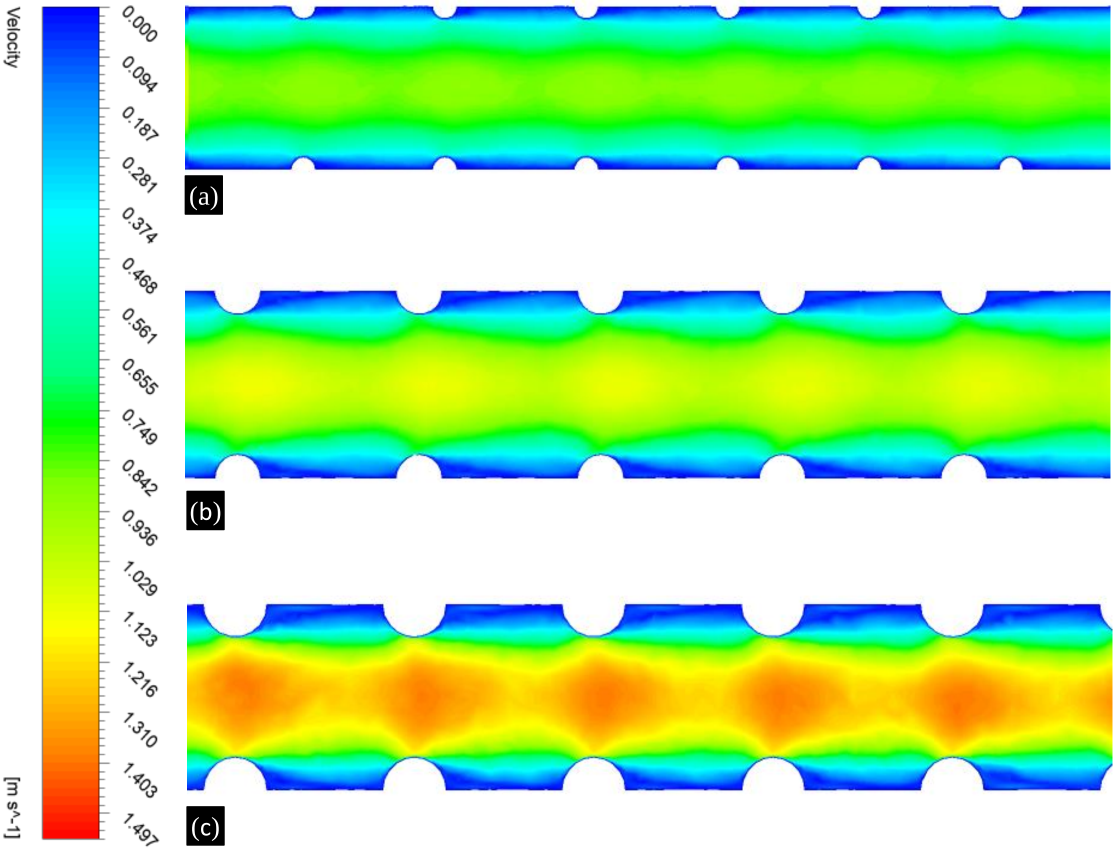


Figure 4. Velocity distribution along the pipe for different CRD values. (a) CRD=1.5 mm; (b) CRD=2.5 mm; (c) CRD=3.5 mm.

Figure 4 shows the velocity distribution contour for different values of CRD. For the smaller value of CRD of 1.5 mm, the local velocity is maximum at the center of the pipe and decreases significantly as one move towards the pipe walls. However, for CRD values of 2.5 and 3.5 mm, the velocity starts to increase even at corrugated ring walls.

In Figure 5, the contours of the turbulent kinetic energy (TKE) are shown for different corrugated ring diameters. It can be seen that the increase in CRD significantly increases the TKE variations in the pipe. This is due to the disturbance caused by the corrugated ring surfaces in the flow direction. The increased variation in TKE suggests that there is better mixing and more turbulence in the flow. This leads to an increase in the Nusselt number, implying an increase in convective heat transfer.

The temperature distribution with different corrugated ring diameters (CRD) is shown in Figure 6. An increase in temperature further away from the wall can be seen with increasing CRD. This is due to more heat being transferred from the wall to the fluid, due to the corrugated rings. The highest temperature in each configuration is on the walls of the pipe. Based on these two observations, it is clear that heat transfer shows a direct correlation with corrugated ring diameter.

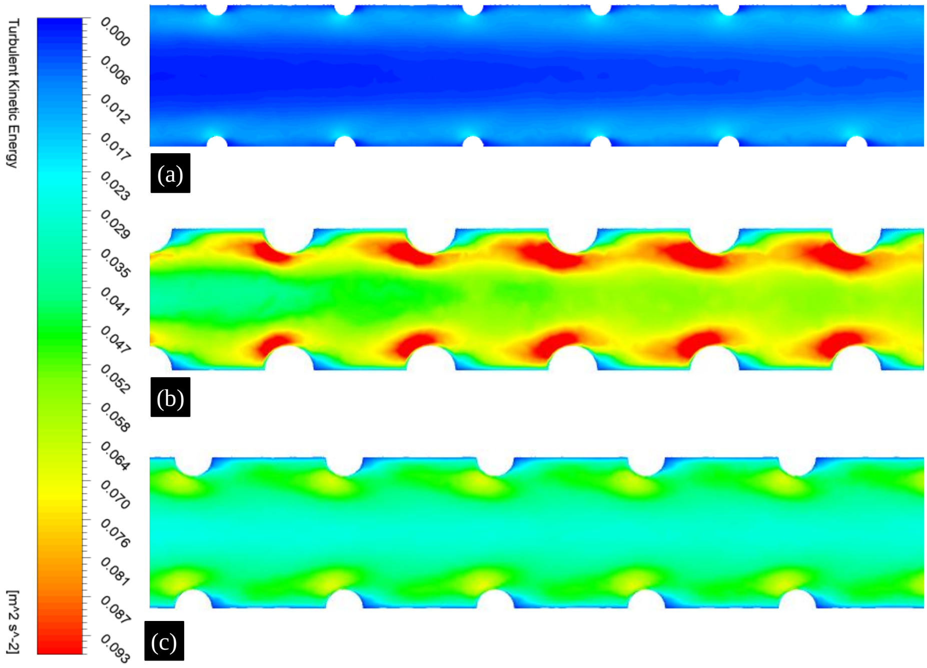


Figure 5. Kinetic energy for different CRD values. (a) CRD=1.5 mm; (b) CRD = 2.5 mm; (c) CRD = 3.5 mm.

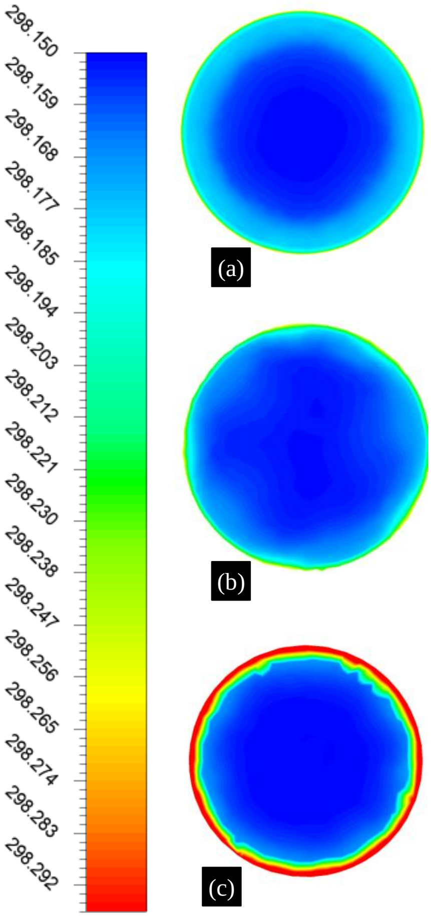


Figure 6. Temperature distribution comparison among different CRD values. (a) CRD = 1.5 mm; (b) CRD = 2.5 mm; (c) CRD = 3.5 mm.

CONCLUSIONS

From the above discussions we can conclude the following:

1. As we increase the diameter of the corrugated ring, the heat transfer coefficient increases and hence there is increase in heat transfer.
2. The increased diameter of the ring increases the disturbance in the flow direction and hence the transverse velocity increases, which leads to greater heat transfer coefficient.

REFERENCES

1. Wang W, Zhang Y, Li B, Li Y. Numerical investigation of tube-side fully developed turbulent flow and heat transfer in outward corrugated tubes. Int J Heat Mass Tran. 2017; 116(2018): 115–126.
2. Alhamid J, Al-Obaidi RA. Flow pattern investigation and thermohydraulic performance enhancement in three-dimensional circular pipe under varying corrugation configurations. In: J Phys: Conf Ser, IOP Publishing. 2021 Mar 1; 1845: 012061.
3. Wang W, Zhang Y, Lee KS, Li B. Optimal design of a double pipe heat exchanger based on the outward helically corrugated tube. Int J Heat Mass Tran. 2019; 135: 706–716.
4. Al-Obaidi AR, Chaer I. Study of the flow characteristics, pressure drop and augmentation of heat performance in a horizontal pipe with and without twisted tape inserts. Case Stud Therm Eng. 2021; 25: 100964.
5. Lee HS. Thermal Design: Heat Sinks, Thermoelectrics, Heat Pipes, Compact Heat Exchangers, and Solar Cells. New York: John Wiley & Sons; 2010.
6. Al-Obaidi AR. Experimental comparative investigations to evaluate cavitation conditions within a centrifugal pump based on vibration and acoustic analyses techniques. Arch Acoust Q. 2020; 45(3): 541–556.
7. Kurtulmus N, Sahin B. Experimental investigation of pulsating flow structures and heat transfer characteristics in sinusoidal channels. Int J Mech Sci. 2020; 167: 105268.
8. Wang W, Shuai Y, Li B, Li B, Lee KS. Enhanced heat transfer performance for multi-tube heat exchangers with various tube arrangements. Int J Heat Mass Tran. 2021; 168: 120905.
9. Al-Obaidi AR. Analysis of the effect of various impeller blade angles on characteristic of the axial pump with pressure fluctuations based on time-and frequency-domain investigations. Iran J Sci Technol Trans Mech Eng. 2021; 45(2): 441–459.
10. Tokgoz N, Sahin B. Experimental studies of flow characteristics in corrugated ducts. Int Commun Heat Mass Tran. 2019; 104: 41–50.
11. Kurtulmus N, Sahin B. A review of hydrodynamics and heat transfer through corrugated channels. Int Commun Heat Mass Tran. 2019; 108(17): 104307.
12. Alam T, Kim MH. A comprehensive review on single phase heat transfer enhancement techniques in heat exchanger applications. Renew Sustain Energy Rev. 2018; 81: 813–839.
13. Yildiz C, Biçer Y, Pehlivan D. Heat transfers and pressure drops in rotating helical pipes. Appl Energy. 1995; 50(1): 85–94.
14. Yang D, Guo Y, Zhang J. Evaluation of the thermal performance of an earth-to-air heat exchanger (EAHE) in a harmonic thermal environment. Energy Convers Manag. 2016; 109: 184–194.
15. Kwon HG, Hwang SD, Cho HH. Flow and heat/mass transfer in a wavy duct with various corrugation angles in two dimensional flow regimes. Heat Mass Tran. 2008; 45(2): 157–165.

Journal of Thermal Analysis and Calorimetry
https://doi.org/10.1007/s10973-023-12119-y

Thermodynamic analysis and experimental investigation of the water spray cooling of photovoltaic solar panels

Yunis Khan[1] · Roshan Raman[1,2] · Mohammad Mehdi Rashidi[3] · Hakan Caliskan[4] · Manish Kumar Chauhan[5] · Akhilesh Kumar Chauhan[5]

Received: 31 July 2022 / Accepted: 12 March 2023

Abstract

This paper investigates an alternative cooling method for photovoltaic (PV) solar panels by using water spray. For the assessment of the cooling process, the experimental setup of water spray cooling of the PV panel was established at Sultanpur (India). This setup was tested in a geographical location with different climate conditions. It was found that the temperature of the panel decreased from 53 to 23 °C and the total power was increased by 15.3% by the water spray cooling. The effectiveness of the system is also increased by its cleaning effects. The efficiency of this solar PV is reduced with the increase of panel temperature. The experiments showed that the PV cell efficiency was dropped by 0.5% with an increase of 1 °C in panel temperature. However, the electrical efficiency of the panel was increased by 0.28%/0.2 °C of temperature drop by the single nozzle spray cooling.

Keywords Thermodynamic · Spray cooling · Photovoltaic · Solar · Water spray method · Efficiency

List of symbols

A	Area/m^{-2}
R_e	Reynolds number
PV	Photovoltaic
WBT	Wet bulb temperature/°C
DBT	Dry bulb temperature/°C
Q	Heat transfer rate/W
$\dot{m}$	Mass flow rate/kg s^{-1}
T	Temperature/K
C_p	Specific heat at constant pressure/kJ kg^{-1} K^{-1}
PCM	Phase change material
G	Solar irradiation/W m^{-1}
I	Current/A
V	Voltage/V
h	Heat transfer coefficient/W m^{-2} K^{-1}
N_u	Nusselt number
P_r	Prandtl number
K	Thermal conductivity/W m^{-2} K^{-1}
v	Velocity/m s^{-2}

Greek symbols

α	Absorptivity
ε	Emissivity
σ	Boltzmann constant/W m^{-2} K^{-4}
η	Efficiency/%
μ	Dynamic viscosity/N s m^{-2}

Subscripts

p	Panel
loss	Loss
C	Convection
E	Evaporation
R	Radiation
0	Ambient condition
w	Wall
f	Fluid

✉ Roshan Raman
roshanraman@ncuindia.edu

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

[2] Department of Mechanical Engineering, The NorthCap University, Gurugram, Haryana 122017, India

[3] University of Electronic Science and Technology of China, Chengdu 610056, Sichuan, China

[4] Department of Mechanical Engineering, Faculty of Engineering, Usak University, 64000 Usak, Turkey

[5] Department of Mechanical Engineering, KNIT Sultanpur, Sultanpur, Utter Pradesh 228118, India

Published online: 27 March 2023

Introduction

Nowadays, environmental protection and efforts to reduce pollution, caused by industrial activities, on the one hand, and research on finding new and improved energy supply options, on the other, have become one of the concerns of governments all around the world [1–3]. In recent years, photovoltaic (PV) power plants due to their proven potential in most parts of the earth are rapidly developing. In several studies, the negative effect of increasing the surface temperature of photovoltaic modules on their electrical efficiency has been shown. The temperature coefficient for power generation depends on the photovoltaic technology of the solar cells [4, 5]. According to the literature review, the highest test temperature for the photovoltaic module was 125 °C in Libya. This temperature caused a 69% drop in the electricity production of PV modules [6]. PV cells are extensively used as one of the most important renewable energy applications because they can use solar energy by converting solar irradiance to direct current (DC) power [7]. According to the materials used to make PV cells, solar irradiance can be converted into direct electricity with varying conversion efficiency ratings ranging from 7 to 40% [8]. A range of about 80% of solar radiation incidents on the surface of PV cells can be absorbed; however, due to the conversion efficiency of PV cell manufacturing technology, only a tiny percentage of the absorbed incident solar energy is transformed into electrical energy [9]. The remaining absorbed energy overheats the PV cells which can reach an operating temperature of around 40 °C over the ambient atmospheric temperature [10–12]. PV cells overheat because they convert a certain band of the entering irradiance spectrum wavelength that is responsible for light direct conversion into electrical energy, while the remaining spectral wavelength causes the PV cells to overheat [13]. Elevated PV panel temperatures are considered one of the most essential issues, especially in hot climatic locations, as they cause a sequential decline in PV electrical conversion efficiency of roughly 0.5%/°C for every degree increase in PV panel temperature, lowering PV panel lifetime [14]. Excess overheating needs to be drained in order to reach an acceptable level of PV cell electricity conversion efficiency because high temperature reduces the performance of the PV cells. Therefore, it is a vital responsibility for enhancing PV cell performance and assuring an efficient conversion process, especially in sun-belt nations. It is necessary to integrate the cooling system for PV cells during the operation. Furthermore, the cooling system will aid in lowering the overall cost of solar cells, extending the lifetime of PV cells, promoting solar cell manufacturing and ensuring optimum output power from installed PV cells [15].Water and air are the most common cooling methods. Air cooling requires less energy than that of water cooling. However, water has a greater cooling capacity and can be circulated for recooling. Also, warm water can be utilized for other purposes such as bathing and washing [16]. Active and passive cooling systems are the most common methods of lowering solar panel temperatures. Water and air are used as a coolant in active cooling, which is driven by a fan or an electric motor. Passive cooling can be further categorized as conductive passive cooling, water cooling and air passive cooling. Heat exchangers, sinks [17] and heat pipes are used to create natural convection in other passive cooling technologies. There are various cooling technologies such as phase change materials (PCM), nanotechnology (nanofluid) sinks, thermoelectric generators, microchannel and other modern cooling methods [18]. In this direction, Shalaby et al. [19] investigated solar panel water cooling is more efficient than air cooling. Water spray cooling enhances the total power output of photovoltaic panels by 33.3%. Spray cooling of water reduced the temperature by 57.1% from 24.7 to 26.4 °C. Also in [20], the authors investigated the effect of evaporative cooling implemented on PV panels and the maximal detected a total increase in power output was around 19%. Direct PV panel cooling with an established water flow over the front side of the panel was investigated in [21] and it was possible to increase power output by 9.5%. Furthermore, in [22] the authors investigated a water spray cooling technique implemented just on the front side of the PV panel, and significant improvement in electrical efficiency was established. Apart from water cooling, there are various other techniques to cool solar PV panels such as microchannel heat exchanger cooling [23], solar panel nanofluid cooling [24], solar panel evaporative cooling [25] and PCM cooling. In the PCM cooling method, latent heat and melting point temperature are the important parameters that decide the cooling rate. PCM absorbs or releases significant quantities when it is realized to change their physical state, and PCM has a high absorber capacity during phase change [26]. In this field recently, Foteinis et al. [27] have experimented with the PCM cooling method of PV panel cooling. They examined that by cooling with PCM the PV panel electricity efficiency was increased by 9.4%. Also, the environmental footprint is reduced by 22% compared to the standalone PV cell. In another cooling method, Lupu et al. [28] examined that heat pipe cooling improved the thermal efficiency of the PV panel by 13.9%. Apart from this, heat pipe uses sealed pipe which should have a high value of thermal conductivity like copper–silver, etc. The heat pipe converted solar panel heat to air or water; this lowered the system heat and improved the system efficiency. It was observed from the literature survey that various researches were present on solar panel water spray cooling, but only a few manual works achieved above 25% efficiency. The novelty of the present work was to evaluate the performance of the PV panel with water spray cooling with a single nozzle and its thermal aspect for Sultanpur (India). The main objective of this paper was to develop an experimental setup and to investigate a water spray cooling technique for Sultanpur (India). The proposed water spray cooling technique can potentially increase PV panel performance due to an evaporation and self-cleaning effect, which is also a great benefit in terms of improved feasibility in the long run.

Experimental setup

The setup for an experiment was made to study the performance of a photovoltaic panel with spray cooling. The solar panel water spray cooling system remains on the roof of the hostel of KNIT Sultanpur, India, for several days during June 2022 on a typical clear summer day when average temperatures of the surrounding air ranged from 30 °C and up to 35 °C. Measurements were taken from 1 to 3 PM (the period of highest solar irradiation levels during June). During the series of measurements, irradiation ranged from 810 W m^{-2} to 850 W m^{-2}, in the already mentioned period of highest solar irradiation (the specific recorded peak value was 850 W m^{-2}) [29]. The geographical location of this place is the latitude, longitude and altitude, respectively, which are 26.27°N, 82.07°E and 95 m above sea level. The main components of this experiment are a solar panel nozzle strainer, a motor pump and a tank. The solar panel is located at some angle from the earth's surface. The panel was fixed under a specific angle of 18° (to obtain the highest electricity output for the fixed slope and which is characteristic of the specific geographical location where the panel was tested). It is situated on the four-pillar stand. Water is pulled from the tank by an electric motor which is dipped inside the reservoir (therefore, it is not visible in Fig. 1) and sprayed on the panel surface by the nozzle. The nozzles were on the front side of the panel and they were fixed at an angle of 40° (to prevent a wire frame shadowing effect and also to ensure a wider water spray dispersion over the PV panel as much as possible) and it did not analyze the nozzle angle influence on the panel performance. After spraying the cooled panel temperature, therefore, panel temperature falls from the rated temperature, and water fully slides on the panel surface so that with the cooling action, the cleaning action is also done. This water, again and again, recalculates via a softer pipe through the pump which is put inside the bucket. The mass flow of water is constant with the value of 2.3 kg s^{-1}. The water distribution system was mounted on a flexible pipe of diameter 4 mm. The distance between the nozzle and panel surface has been taken as 150 mm front side [30]. It was fixed for the maximum dispersion of the water on the panel. It was observed that as compared to the other cooling systems water spray cooling was more efficient, except

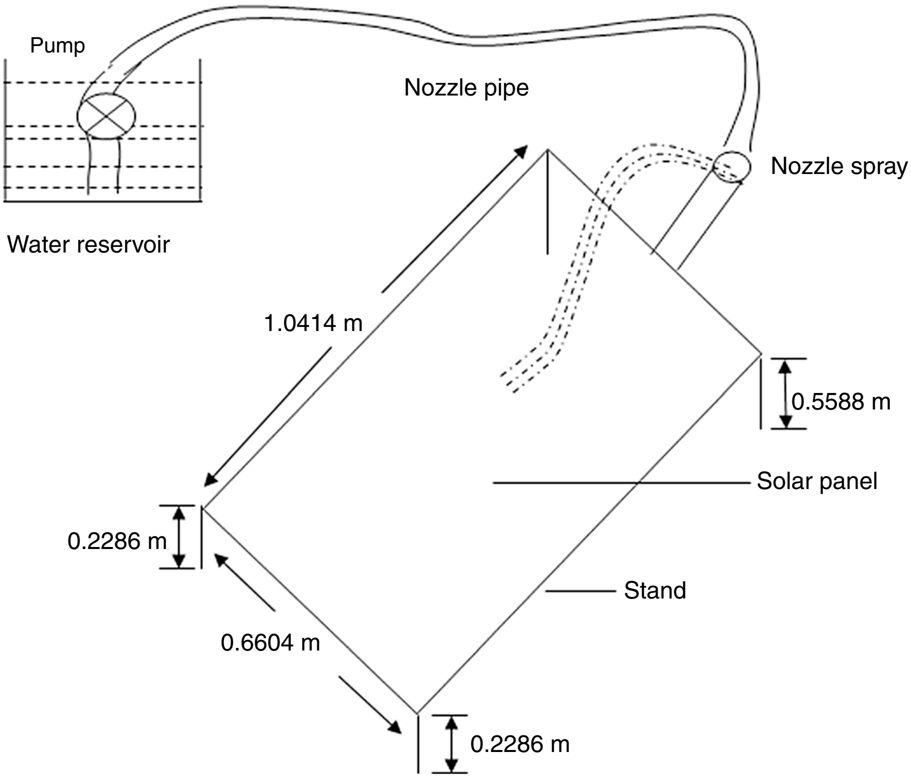
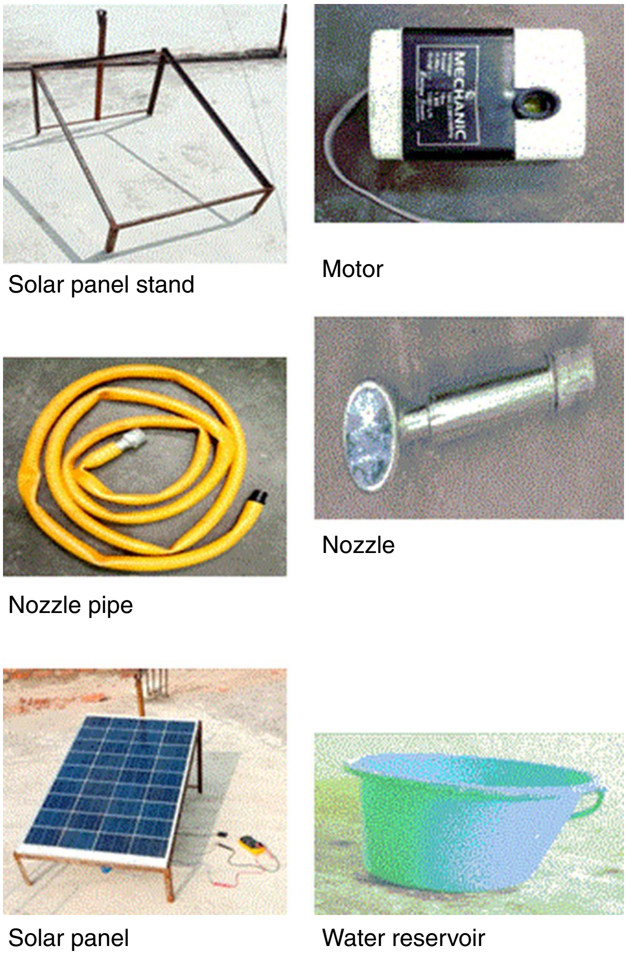


Fig. 1 Designed water spray cooling system for solar panels

for that water submerged method in which higher efficiency was obtained. Also, water spray works as a self-cleaning agent with a good cleaning action technique. The surface of the panel is made of a low coefficient of reflex ion material. Less reflection means more light is observed by the surface and also it has a lower mass than that of other materials. When there is solar irradiation incidence on the PV cells same time, infrared radiation was observed efficiently by the surface. The remaining light is converted into electricity by the photovoltaic cell. The setup of the line diagram and photograph is shown in Figs. 1 and 2, respectively.

The main components used in the solar panel water spray cooling system and their specifications are described as a *solar panel stand:* A stand is necessary for holding the panel. It has four legs. It is constructed of iron material and parts are welded having dimensions 1.0414 m × 0.6604 m. The lengths of the lower and higher legs are 0.2286 m and 0.5588 m, respectively. It has a nozzle mouth at the higher leg side. The nozzle diameter mouth is 1.51 mm as displayed in Fig. 3. *Nozzle pipe:* A PVC material pipe is used for transferring water from the bucket to the nozzle mouth. It has a length of 1.5 m and 1.38 mm in diameter as shown in Figs. 1 and 3. *Motor:* The motor is used for the conversion of mechanical energy into hydraulic energy. It generates flow with power and, according to the load, manages the pressure. An AC motor integrated pump 18 W, 165-230 V/50 Hz, model MEN265A (Mechanic Pvt. Limited) is used for pulling water from the reservoir through the pipe and send to the nozzle pipe. Its photograph is shown in Fig. 3. *Nozzle:* For spraying water, a 1.38 m diameter nozzle is used. It is made of steel material with many small holes to spray the water. *Solar panel:* Photovoltaic modules use the generation of electricity by the use of solar energy (photons). The modules utilized thin-film cells or crystalline silicon cells.


Solar panel stand

Motor

Nozzle pipe

Nozzle

Solar panel

Water reservoir

Fig. 3 Components for experimental setup

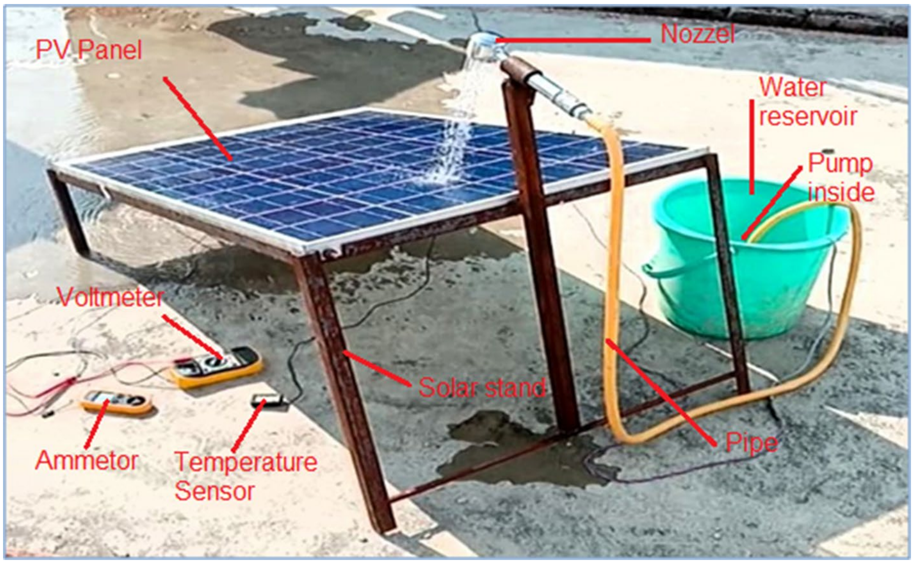


Fig. 2 Photograph of solar PV panel with water spray cooling system (experimental setup)

Table 1 Technical characteristics of solar PV panel

Characteristics*	
Model	LUM 12,165
Effective area/m^{-2}	0.6877
Maximum power output/W	165
Open-circuit voltage/maximal voltage/V	22.84 V/18.65 V
Short-circuit current/maximal current	9.55A/8.85A

*At solar irradiation (DNI) of 950 W m^{-2} and temperature of 25 ℃

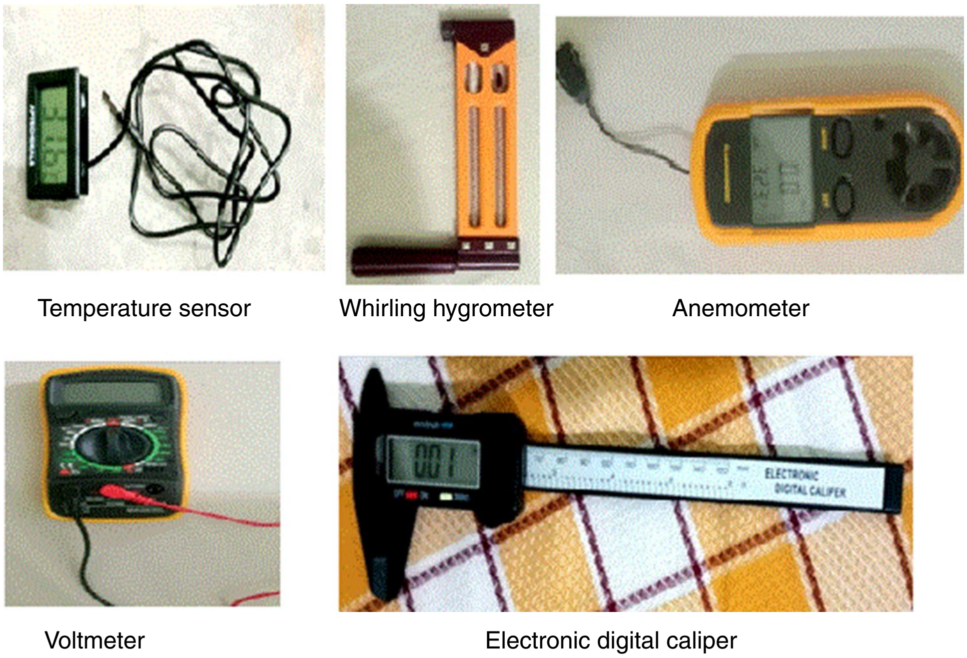
It was developed by Luminous Technologies Pvt. Limited, India. The characteristics of the panel are given in Table 1. In all cells, one electric cell is so connected that mechanical damage and vibration do not take place. A 12 V power solar panel is used for cooling purposes. Its length is 1.0414 m and its width is 0.6604 m. This solar panel is fixed on the solar panel stand. The conversion of solar energy into electrical energy by the solar panel is given in Figs. 1 and 3. *Water reservoir:* A water reservoir capacity of 20 L is used for the flow of water. It is used for the continuous supply of water. The motor pump is fully dipped inside it and pulled water from it.

The following instruments are used for the measurements of the system characteristics. *Temperature sensor:* The temperature at the required places (backside and front side as shown in Fig. 1) was measured by the digital sensor and is shown in Fig. 4. This sensor (model TPM-10) can be used to measure the temperature in the range of -50 °C to 110 °C with an accuracy of + 1 °C. This temperature sensor is manufactured by ApTechDeals Company. *Whirling hygrometer:* Whirling hygrometer is used for the measurement of measuring wet and dry bulb temperatures. It can measure temperatures up to 20°F as shown in Fig. 4. *Anemometer:* The wind velocity affects the temperature of the solar panel; its speed is measured by an anemometer as shown in Fig. 4. The range of velocity of the anemometer is 0–45 m s^{-1} with a range of operating temperature – 10 to 50 °C and humidity 40–85% RH, respectively. *Voltmeter:* Voltmeter is used for measuring voltage and current value. It can measure the maximum voltage as 600 V, and the maximum current is 10A. It has three poles: One is voltage, another is current, and the third, between the two, is a neutral pole. *Electronic digital caliper:* Electronic digital caliper is used for measuring the dimension of different sections. It can be measured in units like mm and inch. It can measure maximum dimensions in the range of 150 mm. Its photograph is shown in Fig. 4.

In summer conditions, the temperature of the solar panel becomes very high; consequently, the value of power output is reduced. Therefore, it is necessary to cool the solar panel for better performance.

Measurement procedure and thermodynamics

The water is taken out from the water reservoir by the pump. This water goes through the pipe and is sprayed on the panel by the nozzle. The panel is cleaned and also cooled with this water. The reading was taken every 5–10-min intervals, and the values of voltage, current, dry bulb temperature (DBT) and wet bulb temperature (WBT) were noted. Also, the air velocity was noted because velocity affects the panel temperature due to forced convection.



Temperature sensor

Whirling hygrometer

Anemometer

Voltmeter

Electronic digital caliper

Fig. 4 Instrument used in the experimental setup

The water spray cooling technique aims to achieve a lower solar panel temperature in this way by increasing the output power of PV. Let G_S be the incoming solar irradiation and A_P be the total solar panel area. Now we are interested in converting the use of full power output of this incoming solar radiation.

According to this Fig. 5, Q_{loss} is the total heat loss, in convection, it is Q_C, Q_E is the total evaporation heat and Q_R is the radiation heat loss.

Then the total solar irradiation is expressed as [30, 31]:

$$Q_{solar} = \alpha \cdot G_S \cdot A_p \tag{1}$$

where α, G_S and A_p absorptive coefficient, solar irradiation and panel area, respectively.

Overall photovoltaic heat loss can be calculated by [30, 31]:

$$Q_C + Q_R + Q_E = Q_{loss} \tag{2}$$

where Q_E, Q_R and Q_C are evaporative, radiation and convection heat loss, respectively.

Total convection heat loss in front and backside is [30, 31]:

$$Q_{C,F} + Q_{C,B} = Q_C \tag{3}$$

Convection heat loss from the front side is also calculated by given below [30, 31]:

$$Q_{C,F} = h_{front} \cdot A_p \cdot \left(T_{panel\ front} - T_{air\ front}\right) \tag{4}$$

Same as from the backside [30, 31]:

$$Q_{C,B} = h_{backside} \cdot A_p \cdot \left(T_{panel\ backside} - T_{air\ backside}\right) \tag{5}$$

Total radiator heat loss is expressed as [30, 31]:

$$Q_{R,F} + Q_{R,B} = Q_R \tag{6}$$

Total heat radiation Q_R can also be calculated by [30, 31]:

$$Q_R = \sigma \cdot \varepsilon \cdot A_P \cdot F_{xy}\left(T_x^4 - T_y^4\right) \tag{7}$$

where x subscription denoted that front side any description denoted that backside, and $F_{xy,}$, T_x, σ, A_P and ε are the view factor of both side, front side temperature, backside temperature, Stefan–Boltzmann constant, panel area and emissivity, respectively.

The total evaporation heat loss depends on both relative humidity and surrounding air temperature and also the velocity of air, respectively. Total evaporation heat loss in PV panels can be explained as [30, 31]:

$$Q_E + Q_{EF} = Q_{E,\ B} \tag{8}$$

The evaporation heat loss also can be written as [30, 31]:

$$Q_E = e \cdot A_p \cdot \left(P_s - P_d\right) \cdot \mathrm{r} \tag{9}$$

The purpose of the water spray is to increase overall heat reflection by evaporation, which is strongly influenced by the evaporation coefficient (e). P_s and P_d is the partial pressure, and the total evaporation heat loss depends on the relative humidity velocity of the air and the temperature of the water spray, where the evaporation coefficient greatly depends on the air velocity; in case of turbulent flow, it directly depends on the concoction heat transfer coefficient. Additionally, it relies on the temperature of narrow flow separation, relative humidity of the surrounding air and other factors.

Also, the following equations have been considered for the performance calculation. Hence, it is clear that heat transfers via force convection.

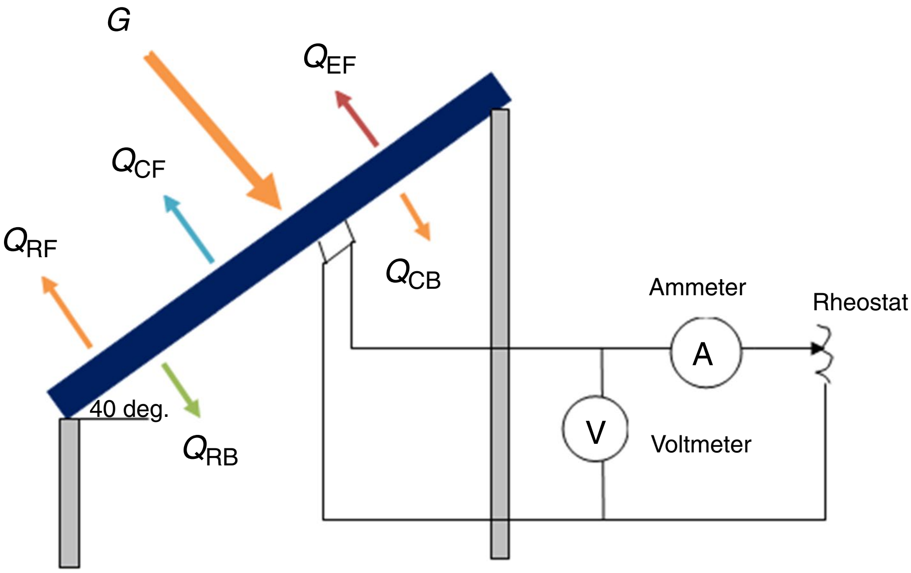


Fig. 5 Image of net heat loss from solar panel

According to Newton's cooling law [30, 31]:

$$Q = h \cdot A_s \cdot (T_w - T_f) \tag{10}$$

Here T_f and T_w are the temperature of fluid and wall, respectively. Q is related to overall heat transfer.

According to the second energy balance [30, 31]:

$$Q = \dot{m} \cdot C_p \cdot (T_e - T_0) \tag{11}$$

The flow is checked for laminar or turbulent flows by the Reynolds number [30, 31]:

$$R_e = \frac{\rho \cdot V \cdot d}{\mu} \tag{12}$$

where ρ, v and μ are density, velocity and dynamic viscosity, respectively,

The velocity of fluid can be determined by [30, 31]:

$$\dot{m} = \rho \cdot A \cdot V \tag{13}$$

Nusselt number can be calculated as [30]:

$$N_u = 0.332(R_e)^{1/2}(P_r)^{1/3} \quad \text{(laminar)} \tag{14}$$

$$N_u = 0.0288(R_e)^{4/5}(P_r)^{1/3} \quad \text{(turbulent)} \tag{15}$$

The formula for the heat transfer coefficient (h) is [30, 31]:

$$N_u = \frac{hd}{k} \tag{16}$$

where d is the hydraulic length.

The efficiency of the system can be calculated as [30, 31]:

$$= \frac{\text{Output power}}{A_p \cdot G_s}. \tag{17}$$

Results and discussions

In this section, performance of the system was discussed without cooling and with cooling one after the other. The results of the experiment were recorded without cooling of solar panel in the month of July 10, 2022, as shown listed in Table 1. The experiment analysis has been taken during the time 1:00 PM to 3:00 PM a day. In this period, there is maximum solar irradiation incidence on the PV panel [29]. This solar irradiation is direct normal irradiation (DNI) (W m.$^{-2}$) incidence on the PV panel. The wind velocity, DBT, WBT and module power have been recorded at some stage in the test and were used within the calculation of the PV module efficiency. The panel temperature initially at 60.6 °C (1:00 PM) is increased and reached the peak point of the day at 62.3 °C (2:00 PM). During the experiment, the panel temperature fluctuated, according to wind velocity and environmental conditions. The minimum temperature of the panel is 59.8 °C at 2:30 PM. After this time, the temperature becomes down gradually. It can be observed that both the value of voltage as well as the value of current drop with increasing cell temperature. It has been noted that as temperatures rise, the module efficiency marginally declines. The solar panel's initial temperature at 1:00 PM was 60.6 °C, and its efficiency was 2.88%. When the temperature slightly decreases to 60.1 °C, then efficiency slightly improved to 2.919%. All variation are listed in Table 2

Further in this section, all measurements were taken with the cooling of the solar panel on July 12, 2022, at Sultanpur, India. The recorded data are listed in Table 3. The wind velocity, DBT, WBT and module power were recorded at some stage in the test and used within the calculation of the PV module efficiency. The panel temperature initially at 45 °C (1:00 PM) is lowered and reaches the lowest point of the day at 42.8 °C (3:15 PM). During the experiment, the panel temperature fluctuated, according to wind velocity and environmental conditions. The minimum temperature of the panel is 42.8 °C at 3:15 PM. After cooling the temperature become down gradually. Here, it was observed that both the value of voltage and the quantity of current result

Table 2 Recorded data of solar panel without cooling on July 10, 2022

Time	1:00 PM	1:15 PM	1:30 PM	1:45 PM	2:00 PM	2:45 PM	3:00 PM
Temperature/°C	60.6	60.1	60.8	61.2	62.3	61.8	59.8
Voltage/V	19.08	19.30	19.24	19.11	19	19.03	19.28
Current/A	1.04	1.06	0.98	0.93	0.63	0.58	0.08
DBT/°C	89	89.1	90	90	91	89.9	88
WBT/°C)	82.5	82.4	83	83	83.5	03	82.9
Velocity of air/ms^{-1}	1.09	1.2	1.1	1.3	1.08	1.2	1.08
Output power/W	19.84	20.458	18.85	17.77	12.92	11.03	20.822
Efficiency/%	2.88	2.919	2.69	2.536	1.84	2.33	3.027

Table 3 Recorded data of solar panel with cooling on July 12, 2022

Time	1:00 PM	1:15 PM	1:30 PM	1:45 PM	2:00 PM	2:15 PM	2:30 PM	3:00 PM	3:15 PM
Temperature/°C	45.0	44.9	44.6	44.5	44.3	46.3	46.3	43	42.8
Voltage/V	20.6	21.9	21.2	20.9	21.9	19.6	20.6	22.1	22.3
Current/A	4.54	4.75	4.96	4.84	4.90	4.75	4.58	4.80	4.83
DBT/°C	80	79.9	79.8	79.7	79.8	80	80	79.9	79.9
WBT/°C	74	73.9	73.8	73.8	73.7	74	74	73.9	73.9
Velocity of air/m s^{-1}	1.1	1.09	1.33	1.08	1.45	1.06	1.3	1.09	1.8
Output power/W	93.5	100.7	105.152	101.156	107.31	91.14	95.264	106.08	107.709
Efficiency/%	13.5	14.64	15.29	14.709	15.60	13.25	13.856	15.42	15.662

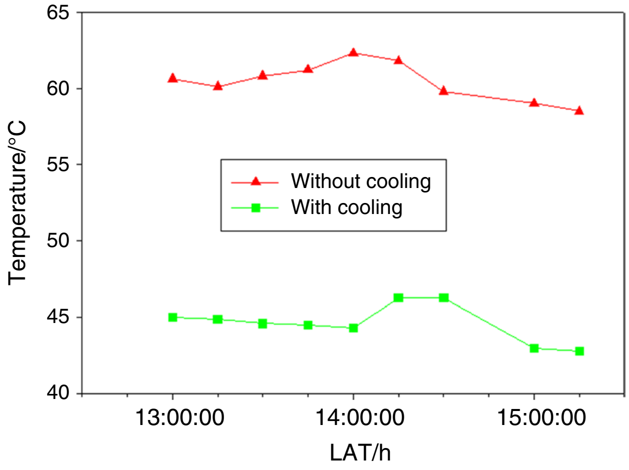


Fig. 6 Variation of temperature with time

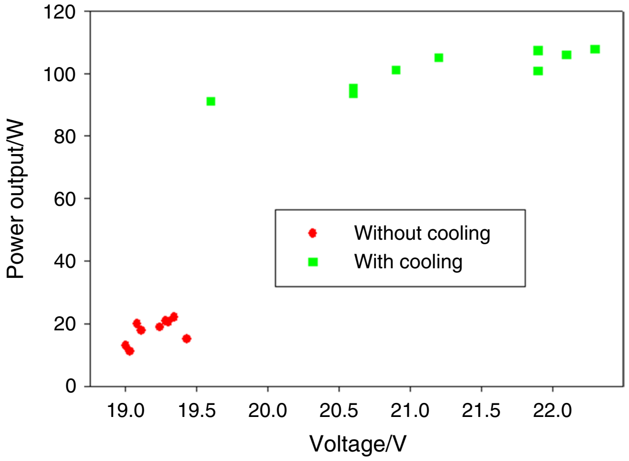


Fig.7 Power output variation with voltage

in decreasing cell temperature. Additionally, it is evident from the results of the experiment that the module efficiency steadily raises with temperature. At 1:00 PM, it is found that the initial temperature of a solar panel is 45 °C and the efficiency is 13.5%. When the temperature slightly decreases to 42.8 °C, then efficiency is slightly improved to 15.662%. At 1:15 PM, the efficiency is 14.64%; after the 15 min, the temperature becomes 44.6 °C which was lowered by 0.2 °C and the efficiency was improved to 15.2% from 14.64%. Therefore, the solar panel must be cooled for achieving better performance and efficiency.

Variation of temperature

Figure 6 shows the temperature variation of solar panel with time. It is observed that the maximum temperature is between 1:00 PM and 2:30 PM on July 10, 2022. At this time, a high amount of solar radiation was received in India. Therefore, energy transformation rate was also high. It can be observed that solar panel temperature fluctuated with the environmental condition and wind velocity also. Further it also shows the temperature variation with time with cooling. The maximum temperature was observed as 46.23 °C in the whole test range, i.e., between 1:00 PM and 3:15 PM on July 12, 2022. The minimum temperature was found as 42.8 °C at 3:15 PM while in the previous test, i.e., without cooling, it was 59.75 °C at the same time. Apart from this, panel temperature fluctuated with the environmental condition and wind velocity as observed in Fig. 6. Solar irradiation is the critical parameter for any solar-integrated study. The variation of surface temperature with direct normal irradiation (DNI) has been also discussed. There is no need to show the variation of the temperature with the DNI because it will show the same trend as the temperature with time, because the same trend showed DNI with the temperature [29].

Variation of power output with voltage

The maximum power was achieved at the time 3:00 PM when the minimum temperature was found. The maximum power was obtained as 20.822 W at 59.8 °C temperature. Figure 7 shows that power output decreased as the voltage increased, reaching a maximum of 19.28 V. It is concluded that when the panel temperature tends to maximum then their power output becomes minimum. Also, in this subsection variations of electrical power with voltage with cooling were discussed. The maximum was observed at 3:15 PM when the temperature was the lowest of all-temperature readings. Table 3 shows that at 42.8 °C maximum voltage and the maximum power were found as a maximum of 22.3 V and 107.709 W. It is concluded that when the panel temperature tends to maximize, the corresponding power would be minimized. The comparison showed that due to the cooling effect, the efficiency was improved. As discussed above, the electrical efficiency decreased with panel surface temperature. It means if it is cooled then it will improve. Figure 7 shows that the dispersion points without cooling are very nearly distributed. It means the variation of the power is very less with the voltage produced. It is because of the high temperature of the panel. Without cooling, power varied only 19.5–19 V during the same tested time and solar irradiation. However, with cooling the dispersion point on the graph is widely sprayed. It means due to the spray of the water over the panel temperature and cleaning is there. The variation of the power with voltage varied from 90.3 to 107.7W as the wide range of voltage varied from 19 to 22.5 V.

Variation of current with voltage

It was also observed that the current first increased with the voltage, and after some value of the voltage, it decreased. Keeping constant all other parameters, as the voltage increased the corresponding temperature also decreased which leads to the decrease in the current. The maximum current was found at 1.06 V of voltage. Figure 8 shows that the dropping rate of the current was fast corresponding to the voltage from 0.58 to 0.55 V. As discussed previously, for this reason the rate of temperature drop was also sharp. Further current variation with the voltage with cooling was also discussed. The pattern of variation was the same as that without cooling conditions. It was observed that the maximum current value occurred at the highest value of the voltage. As the panel temperature lowered, both temperature and voltage got the highest value. In comparison with Fig. 8, it was seen that with cooling the current was obtained more than without cooling. It can be seen that little bit of variation of the current with voltage without cooling the panel. Without cooling, the current was varied from 0.18 to 1.2A only because the temperature of the panel was high before cooling the duct and other impurities are there. However, a high current was achieved with the cooling, and the current varied from 4.8 to 4.9A in between fluctuating with the voltage. This variation occurred with a wide range of voltage, i.e., 19.0 to 22.1 V.

Variation of efficiency with temperature

It was observed that when the panel temperature reached a maximum value the corresponding electrical efficiency also obtained its minimum value as shown in Fig. 9. In this experiment, the maximum efficiency occurred at the lowest temperature which is 59.8 °C, and at the maximum temperature 60.8 °C, efficiency is the lowest value which is 2.69%. Electrical efficiency first increased slowly with temperature; after that, it increased sharply from the temperature of 46.3–43 °C. Maximum electrical efficiency was found as 22.32% at 43 °C overall temperature limits during the test with cooling, while lowest, at temperature 46.3 °C, efficiency has a value of 13.20% as displayed in Fig. 9. The comparison showed that the panel's electrical efficiency was more in cooling as compared to without cooling. This is due to a decrease in temperature and the cleaning action of water. The variation of the efficiency

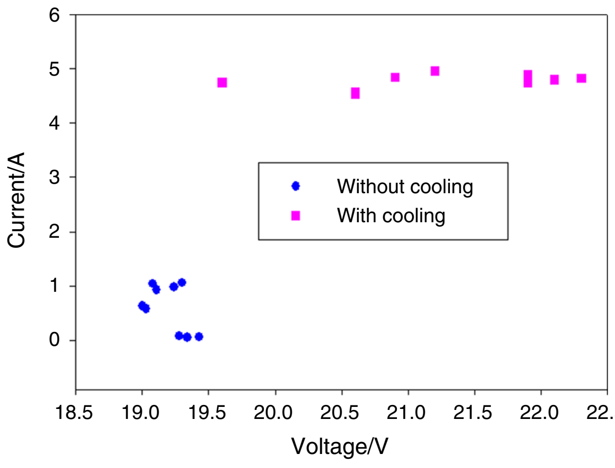


Fig. 8 Variation of current with voltage

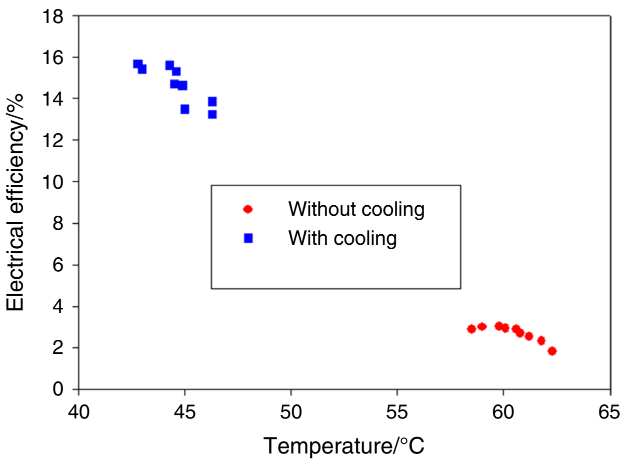


Fig. 9 Efficiency variations with temperature

with the temperature without cooling varied from 2.88 to 3.027% when the temperature increased from 60.6 to 59.8 °C during the tested time, i.e., 1:00 PM to 3:00 PM. However, the efficiency will increase with cooling with the temperature from 13.5 to 15.667% when the temperature increased from 45.0 to 42.8 °C. The temperature increased with time with cooling the efficiency also decreased.

Performance variation with temperature

The power of the module increases; hence, its efficiency increases. Higher temperatures of solar panels reduce both power and efficiency. It can be maintained by cooling the heated solar panel. Figure 10 shows the variation of power and efficiency with temperature. The efficiency is found below 20% and current power fluctuated between 80 and 110 W, while maximum power output and efficiency are obtained at 42.8 °C which is the minimum temperature in the experiment. The range of solar irradiation is 810–850 W m^{-2}. The power output will increase from 80 to 110W with the temperature from 42.8 to 46 °C. The temperature fluctuates. The highest temperature was found as 46 °C at which the efficiency and power were 13.25% and 91.5%, respectively. The highest electrical efficiency was found to be 110 kW at the 42.2 °C surface temperature.

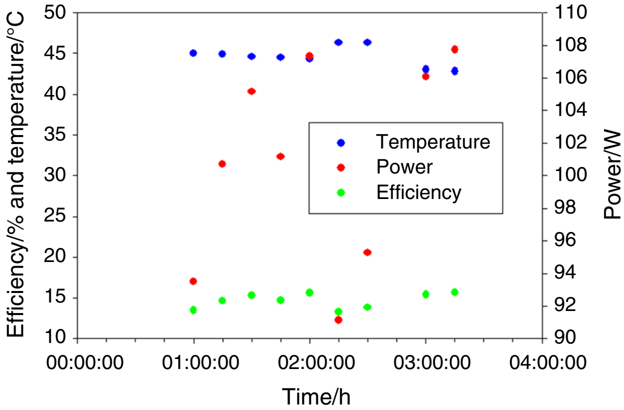


Fig. 10 Variation of power and efficiency with temperature

Comparisons of the result with previous work

The performance of the solar panel water cooling (July 2022) is compared with the previous spray water cooling system for photovoltaic panel work (June 2019) [31]. The compassion is made by considering the parameter of solar panel temperature, solar radiation and the solar panel system efficiency in both cooling and non-cooling systems. While the time frame for the current project was 1:00 PM to 3:15 PM at 15-min intervals, the prior project's time frame was 11:30 AM to 3:30 PM at 10-min intervals. The time and date for the current experiment are different from the previous study. In earlier research, the temperature range was between 24.2 and 27.8 °C with a spray cooling system. However, in the present research, the temperature range was between 60.8 and 59 °C with a spray cooling system. The narrowing of the temperature range is brought on by wind speed, water's particular heat, DBT and WBT. The results are listed in Table 4. In the reference, the panel's maximum power output was obtained values of 54W and 72W without and with spray cooling, respectively. In reference research work, where the highest power production is 22.822W without a cooling panel, employing the steady cooling system results in increases in maximum electrical power output of roughly 33.3%. This maximum power is increased to 107.7 W by continuous spray cooling in the present work as given in Table 4. The present research with the single point pray cooling increased the power output by 80.66%. In the case without a cooling system, electrical efficiency was found to mean an efficiency of 9.11%. Utilizing the stable cooling system, the prior work's electrical efficiency was obtained at 12.1%. The present work maximum efficiency without cooling was obtained as 3.027%. This efficiency improved by cooling to 15.662%. The maximum voltage value in the previous research work was recorded at the voltage of 13.5 V with the cooling system this is improved to 17 V, whereas in current research work 19.23 V in the case without a cooling system, this is improved to 22.3 V with spray cooling.

Table 4 Comparison of the results with previous studies under the same operating conditions*

Performance parameters	Reference [31]		This study	
	Without cooling	With cooling	Without cooling	With cooling
Electrical efficiency/%	9.1	12.1	3.02	15.66
Power output/W	54	72	20.82	107.7
Voltage/V	13.5	17	19.23	22.3

*Average DNI of 985 W m^{-2} and ambient temperature of 28–31 °C

Conclusions

The thermodynamic analysis is performed on the experimental setup of the water spray cooling of photovoltaic solar panels. The following main conclusions can be obtained from this study:

- The highest electrical efficiency and power of the PV panel were found as 15.66% and 107.8 W, respectively, at the minimum temperature of 42.8 °C at the time of 3:15 PM.
- The maximum temperature of the solar panel is found between 40 and 60 °C. Higher temperature prompt lowered the efficiency and power output of the solar panel.
- The electrical efficiency of the panel was increased by 0.28%/0.2 °C of temperature drop by the single nozzle spray cooling.
- The environment, wind speed, dry bulb and wet bulb temperatures, panel inclination, water temperatures, and distance between nozzle and solar panel influenced the efficiency and performance of the solar panels.
- By water spray cooling, many types of dust particles (an environmental impurity that is laid on the panel surface) are cleaned, and consequently, higher solar irradiation is achieved. Consequently, thermal performance was improved.

Author contributions YK and RR had written the complete paper which was conceptualized by MMR and HC. Moreover, the proofreading was carried out by MMR and HC. All the experiments and data collection were performed by MKC and AKC under the esteemed guidance.

References

1. Hooman AG, Hoseinzadeh S, Esmaeilion F, Memon S, Garcia DA, Assad MEH. A solar thermal driven ORC-VFR system employed in the subtropical Mediterranean climatic building. Energy. 2022;250: 123819.
2. Hoseinzadeh S, Nastasi B, Groppi D, Garcia DA. Exploring the penetration of renewable energy at increasing the boundaries of the urban energy system – The PRISMI plus toolkit application to Monachil, Spain. Sustain Energy Technol Assess. 2022;54: 102908.
3. Olia H, Torabi M, Bahiraei M, Ahmadi MH, Goodarzi M, Safaei MR. Application of nanofluids in thermal performance enhancement of parabolic trough solar collector: state-of-the-art. Appl Sci. 2019;9(3):463.
4. Ahmadi MH, Ghazvini M, Sadeghzadeh M, Nazari MA, Kumar R, Naeimi A, Ming T. Solar power technology for electricity generation: a critical review. Energy Sci Eng. 2018;6(5):340–61.
5. Said Z, Rahman S, Sharma P, Hachicha AA, Issa S. Performance characterization of a solar-powered shell and tube heat exchanger utilizing MWCNTs/water-based nanofluid: an experimental, numerical, and artificial intelligence approach. Appl Therm Eng. 2022;212: 118633.
6. Firoozzadeh M, Shiravi AM, Shafiee M. Thermodynamics assessment on cooling photovoltaic modules by phase change materials (PCMs) in critical operating temperature. J Therm Anal Calorim. 2023. https://doi.org/10.1007/s10973-020-09565-3.
7. Keith E. The rating of photovoltaic performance. IEEE Trans Electron Dev. 1999;46:1928–31.
8. Kurtz S. Opportunities and challenges for development of mature concentrating photovoltaic power industry, National Renewable Energy Laboratory. 2011, Technical Report (NREL/TP) 520–43208.
9. Helden WJGV, Zolingen RJCH, Zondag HAPV. Thermal systems: PV panels supplying renewable electricity and heat. Prog Photovolt Res Appl. 2004;12:415–26.
10. Ye Z, Nobre A, Reindl T, Luther J, Reise C. On PV module temperatures in tropical regions. Sol Energy. 2013;88:80–7.
11. Rahman MM, Hasanuzzaman M, Rahim NA. Effects of various parameters on PV module power and efficiency. Energy Convers Manage. 2015;103:348–58.
12. Abd-Elhady MS, Fouad M, Khalil T. Improving the performance of photovoltaic panels by oil coating. Energy Convers Manage. 2015;115:1–7.
13. Luque A, Hegedus S. Handbook of photovoltaic science and engineering. 2nd ed. West Sussex: John Wiley & Sons; 2011.
14. Skoplaki E, Palyvos JA. On the temperature dependence of photovoltaic module electrical performance: a review of efficiency/power correlations. Sol Energy. 2009;83:614–24.
15. Chatterjee S, Mani GT. BAPV arrays: side-by-side comparison with and without fan cooling. IEEE Photovolt Spec Conf (PVSC). 2011;7:537–42.
16. Elminshawy Nabil AS, Mohamed AMI, Morad K, Elhenawy Y, Alrobaian Y. Performance of PV panel coupled with geothermal air cooling system subjected to hot climatic. Appl Therm Eng. 2019;148:1–9.
17. Arifin Z, Tjahjana DDDP, Hadi S, Rachmanto RA, Setyohandoko G, Sutanto B. Numerical and experimental investigation of air cooling for photovoltaic panels using aluminum heat sinks. Int J Photoenergy. 2020. https://doi.org/10.1155/2020/1574274.
18. Shalaby SM, Elfakharany MK, Moharram BM, Abosheiasha HF. Experimental study on the performance of PV with water cooling. Energy Rep. 2022;8:957–61. https://doi.org/10.1016/j.egyr.2021.11.155.
19. Alami AH. Effects of evaporative cooling on efficiency of photovoltaic modules. Energy Convers Manage. 2014;77:668–79.
20. Dorobantu L, Popescu MO. Increasing the efficiency of photovoltaic panels through cooling water film, UPB. Sci Bull Ser C Electr Eng. 2013;75(4):223–32.
21. Abolzadeh M, Ameri M. Improving the effectiveness of a photovoltaic water pumping system by spraying water over the front of photovoltaic cells. Renew Energy. 2009;34(1):91–6.
22. Shittu S, et al. Experimental study and exergy analysis of photovoltaic-thermoelectric with flat plate micro-channel heat pipe. Energy Convers Manage. 2020;207: 112515. https://doi.org/10.1016/j.enconman.2020.112515.
23. Alrobaian AA, Alturki AS. Investigation of numerical and optimization method in the new concept of solar panel cooling under the variable condition using nanofluid. J Therm Anal Calorim. 2020;142:2173–87.
24. Advance cooling techniques of P.V. modules: A state of art Pushpendu Dwivedi, I Kirpichnikova, October 2020.
25. Rajvikram M, Leoponraj S, Ramkumar S, Akshaya H, Dheeraj A. Experimental investigation on the abasement of operating

temperature in solar photovoltaic panel using PCM and aluminium. Sol Energy. 2019;188:327–38.

26. Foteinis S, Savvakis N, Tsoutsos T. Energy and environmental performance of photovoltaic cooling using phase change material under the meditation climate. Energy. 2023;265: 126355.
27. Lupu AG, et al. A review of solar photovoltaic systems cooling technologies. IOP Conf Ser Mater Sci Eng. 2018;444: 082016. https://doi.org/10.1088/1757-899X/444/8/082016.
28. Sukhatme SP, Nayak JK. Solar energy: principles of thermal collection and storage, 3rd Edition, Tata McGraw-Hill Education, New Delhi, 2008, pp. 1–431.
29. Nizetic S, Coko D, Yadav AF, Cabo G. Water spray cooling technique applied on a photovoltaic panel: the performance response. Energy Convers Manage. 2016;108:287–96.
30. Hadipour A, Zargarabadi MR, Rashidi S. An efficienct pulsed-spray cooling for photovoltaic panels: experimental study and cost analysis. Renew Energy. 2021;164:867–75.
31. Said Z, Sharma P, Sundar LS, Li C, Tran DC, Pham NDK, Nguyen XP. Improving the thermal efficiency of a solar flat plate collector using MWCNT-Fe_3O_4/water hybrid nanofluids and ensemble machine learning. Case Stud Therm Eng. 2022;40: 102448.

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

Toeplitz determinants of Logarithmic coefficients for Starlike and Convex functions

Surya Giri[1] and S. Sivaprasad Kumar*

Abstract

In this study, we deal with the sharp bounds of certain Toeplitz determinants whose entries are the logarithmic coefficients of analytic univalent functions f such that the quantity $zf'(z)/f(z)$ takes values in a specific domain lying in the right half plane. The established results provide the bounds for the classes of starlike and convex functions, as well as various of their subclasses.

Keywords: Univalent functions; Starlike functions; Convex functions; Logarithmic coefficients; Toeplitz determinants.
AMS Subject Classification: 30C45, 30C50.

1 Introduction

Let $\mathcal{A}$ be the class of analytic functions f defined on the open unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ with the following Taylor series expansion:

$$f(z) = z + \sum_{n=2}^{\infty} a_n z^n. \tag{1.1}$$

The subclass of $\mathcal{A}$ consisting of all univalent functions is denoted by $\mathcal{S}$. Associated with each function $f \in \mathcal{S}$, consider

$$F_f(z) = \log \frac{f(z)}{z} = 2 \sum_{n=1}^{\infty} \gamma_n(f) z^n, \quad z \in \mathbb{D}, \quad \log 1 = 0. \tag{1.2}$$

The number $\gamma_n := \gamma_n(f)$, for each $n = 1, 2, 3, \cdots$, is called the logarithmic coefficients of f. Using the idea of logarithmic coefficients, Kayumov [13] proved the Brennan's conjecture for the conformal mappings. Also, logarithmic coefficients play an important role in Milin's conjecture ([8, p. 155], [18]). Contrary to the coefficients of $f \in \mathcal{S}$, a little exact information is known about the coefficients of $\log(f(z)/z)$ when $f \in \mathcal{S}$. The Koebe function leads to the natural conjecture $|\gamma_n| \leq 1/n$, $n \geq 1$ for the class $\mathcal{S}$. However, this is false, even in order of magnitude (see [8, Section 8.1]). For $f \in \mathcal{S}$, the only known bounds are

$$|\gamma_1| \leq 1 \text{ and } |\gamma_2| \leq \frac{1}{2} + \frac{1}{e^2}.$$

The problem of finding the estimates of $|\gamma_n|$ $(n \geq 3)$ for the class $\mathcal{S}$ is still open. In past few years, various authors examined the bounds of $|\gamma_n|$ for functions in the subclasses of $\mathcal{S}$ instead of the whole class (see [5, 1, 6, 23, 24]) and the references cited therein).

In geometric function theory, the classes of convex and starlike functions are the subclasses of $\mathcal{S}$ that have received the most attention. A function $f \in \mathcal{S}$ is said to be convex if $f(\mathbb{D})$ is convex. Let $\mathcal{C}$ denote the class of convex functions. It is well known that, $f \in \mathcal{C}$, if and only if $\operatorname{Re}((1 + zf''(z))/f'(z)) > 0$ for $z \in \mathbb{D}$. A function $f \in \mathcal{S}$ is said to be starlike if $f(\mathbb{D})$ is starlike with respect to the origin. Let $\mathcal{S}^*$ denote the class of starlike functions. Analytically, $f \in \mathcal{S}^*$, if and only if $\operatorname{Re}(zf'(z)/f(z)) > 0$ for $z \in \mathbb{D}$. Let Ω be the class of all Schwarz functions and $\mathcal{P}$ denote the class of analytic functions $p : \mathbb{D} \to \mathbb{C}$ such that $p(0) = 1$ and $\operatorname{Re} p(z) > 0$ for all $z \in \mathbb{D}$. An analytic function f is said to be subordinate to the analytic function g, if there exists a Schwarz function ω such that $f(z) = g(\omega(z))$ for all $z \in \mathbb{D}$. It is denoted by $f \prec g$. Ma and Minda [17] unified various subclasses of starlike and convex functions. They defined

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

and

$$\mathcal{C}(\varphi) = \left\{ f \in \mathcal{S} : 1 + \frac{zf''(z)}{f'(z)} \prec \varphi(z) \right\},$$

where $\varphi(z)$ is an analytic univalent functions with positive real part in $\mathbb{D}$, $\varphi(\mathbb{D})$ is symmetric with respect to the real axis starlike with respect to $\varphi(0) = 1$, and $\varphi'(0) > 0$. Let, for $z \in \mathbb{D}$, φ has the series expansion

$$\varphi(z) = 1 + B_1 z + B_2 z^2 + B_3 z^3 + \cdots, \quad B_1 > 0.$$

Since $\varphi(\mathbb{D})$ is symmetric about the real axis and $\varphi(0) = 1$, therefore all B_i's are real. Further, φ is a Carathéodory function, it follows that $|B_n| \leq 2$, $n \in \mathbb{N}$ [8, page-41].

If we take $\varphi(z) = (1 + Az)/(1 + Bz)$, $-1 \leq B < A \leq 1$, $\mathcal{S}^*(\varphi)$ and $\mathcal{C}(\varphi)$ reduce to the classes of Janowski starlike and convex functions, denoted by $\mathcal{S}^*[A, B]$ and $\mathcal{C}[A, B]$ respectively (see [11]). For $B = -1$ and $A = 1 - 2\alpha$, $(0 \leq \alpha < 1)$, the classes $\mathcal{S}^*(\alpha) = \mathcal{S}^*[1 - 2\alpha, -1]$ and $\mathcal{C}(\alpha) = \mathcal{C}[1 - 2\alpha, -1]$ are the well known classes of starlike and convex functions of order α $(0 \leq \alpha < 1)$ (see [8]).

Toeplitz matrices and Toeplitz determinants arise in the field of pure as well as applied mathematics [25]. They occur in analysis, integral equations, image processing, signal processing, quantum mechanics and among other areas. For more applications, we refer to the survey article [27]. Toeplitz matrices contain constant entries along their diagonals. For $f(z) = z + \sum_{n=2}^{\infty} a_n z^n \in \mathcal{A}$, the Toeplitz determinant is given by

$$T_{m,n}(f) = \begin{vmatrix} a_n & a_{n+1} & \cdots & a_{n+m-1} \\ a_{n+1} & a_n & \cdots & a_{n+m-2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n+m-1} & a_{n+m-2} & \cdots & a_n \end{vmatrix}, \tag{1.3}$$

where $m, n \in \mathbb{N}$. In case of the class $\mathcal{S}^*$ and $\mathcal{C}$, the bound of $|T_{2,n}(f)|$, $|T_{3,1}(f)|$ and $|T_{3,2}(f)|$ were examined by Ali et al. [3] in 2017. Motivated by this work, for small values of m and n, various authors studied the bounds of $|T_{m,n}(f)|$ for various subclasses of $\mathcal{S}$ in past few years [2, 7, 10, 16, 20].

Hankel and Toeplitz matrices are closely related to each other. Hankel matrices contain constant entries along the reverse diagonals. Ye and Lim [27] showed that any $n \times n$ matrix over $\mathbb{C}$ generically can be written as the product of some Toeplitz matrices or Hankel matrices. Recently, Kowalczyk and Lecko [14] introduced the Hankel determinant whose entries were the logarithmic coefficients of functions in $\mathcal{A}$. They studied the sharp estimates of second order Hankel determinant of logarithmic coefficients for functions belonging to $\mathcal{S}^*$ and $\mathcal{C}$, which is further generalized for the classes $\mathcal{S}^*(\alpha)$ and $\mathcal{C}(\alpha)$ by the same authors in [15]. Also, Mundalia and Kumar [19] studied the same problem for the certain subclasses of close-to-convex functions.

Motivated by these works and considering the significance of Toeplitz determinant and logarithmic coefficients, we define

$$T_{m,n}(\gamma_f) = \begin{vmatrix} \gamma_n & \gamma_{n+1} & \cdots & \gamma_{n+m-1} \\ \gamma_{n+1} & \gamma_n & \cdots & \gamma_{n+m-2} \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{n+m-1} & \gamma_{n+m-2} & \cdots & \gamma_n \end{vmatrix}. \tag{1.4}$$

Consequently, we obtain

$$T_{2,1}(\gamma_f) = \gamma_1^2 - \gamma_2^2 \quad \text{and} \quad T_{2,2}(\gamma_f) = \gamma_2^2 - \gamma_3^2.$$

A comparison of same powers of z in (1.2) yields that

$$\gamma_1 = \frac{a_2}{2}, \quad \gamma_2 = \frac{1}{4}(2a_3 - a_2^2) \text{ and } \gamma_3 = \frac{1}{2}\left(a_4 - a_2 a_3 + \frac{1}{3}a_2^3\right). \tag{1.5}$$

In this paper, we derive the sharp estimates of $|T_{2,1}(\gamma_f)|$, $|T_{2,2}(\gamma_f)|$ and $|T_{3,2}(f)|$ for the classes $\mathcal{S}^*(\varphi)$ and $\mathcal{C}(\varphi)$. The established bounds lead to a number of new and already known results for different subclasses of starlike and convex functions when φ is appropriately chosen.

The following lemmas are required to prove the main results.

Lemma 1.1. *[21] If $\omega(z)=\sum_{n=1}^{\infty} c_n z^n \in \Omega$ and $(\mu,\nu)\in \cup_{i=1}^{3} D_i$, then*

$$|c_3+\mu c_1 c_2+\nu c_1^3|\leq |\nu|,$$

where

$$D_1=\left\{(\mu,\nu): |\mu|\leq 2,\ \nu\geq 1\right\},\ \ D_2=\left\{(\mu,\nu): 2\leq|\mu|\leq 4,\ \nu\geq\frac{1}{12}(\mu^2+8)\right\},$$

and

$$D_3=\left\{(\mu,\nu): |\mu|\geq 4,\ \nu\geq\frac{2}{3}(|\mu|-1)\right\}.$$

Lemma 1.2. *[9, Theorem 1] Let $p(z)=1+\sum_{n=1}^{\infty} p_n z^n\in\mathcal{P}$ and $\mu\in\mathbb{C}$. Then*

$$|p_n-\mu p_k p_{n-k}|\leq 2\max\{1,|2\mu-1|\},\quad 1\leq k\leq n-1.$$

The inequality is sharp for the function $p(z)=(1+z)/(1-z)$ or its rotation when $|2\mu-1|\geq 1$. In case of $|2\mu-1|<1$, the inequality is sharp for $p(z)=(1+z^n)/(1-z^n)$ or its rotations.

2 Main results

We begin with the bounds of $|T_{2,1}(\gamma_f)|$ and $|T_{2,2}(\gamma_f)|$ for the classes $\mathcal{S}^*(\varphi)$ and $\mathcal{C}(\varphi)$.

Theorem 2.1. *Let $\varphi(z)=1+B_1z+B_2z^2+B_3z^3+\cdots$ and $f\in\mathcal{S}^*(\varphi)$. If $|B_2|\geq B_1$, then*

$$|\gamma_1^2-\gamma_2^2|\leq\frac{B_1^2}{4}+\frac{B_2^2}{16}.$$

The estimate is sharp.

Proof. Let $f\in\mathcal{S}^*(\varphi)$ be of the form (1.1). Then there exists a Schwarz function, say $\omega(z)=\sum_{n=1}^{\infty} c_n z^n$ such that

$$\frac{zf'(z)}{f(z)}=\varphi(\omega(z)),\quad z\in\mathbb{D}. \tag{2.1}$$

From the Taylor series expansions of f and φ, we obtain

$$\frac{zf'(z)}{f(z)}=1+a_2z+(-a_2^2+2a_3)z^2+(a_2^3-3a_2a_3+3a_4)z^3+\cdots \tag{2.2}$$

and

$$\varphi(\omega(z))=1+B_1c_1z+(B_2c_1^2+B_1c_2)z^2+(B_3c_1^3+2B_2c_1c_2+B_1c_3)z^3+\cdots. \tag{2.3}$$

By comparing the same powers in (2.1) using (2.2) and (2.3), coefficients a_2, a_3 and a_4 can be expressed as

$$a_2=B_1c_1,\ \ a_3=\frac{1}{2}(B_1^2c_1^2+B_2c_1^2+B_1c_2) \tag{2.4}$$

and

$$a_4=\frac{1}{48}((8B_1^3+24B_1B_2+16B_3)c_1^3+(24B_1^2+32B_2)c_1c_2+16B_1c_3). \tag{2.5}$$

Further, applying $|c_n|\leq 1$, we get

$$|a_2|\leq B_1. \tag{2.6}$$

Ali et al. [4, Theorem 1] established the bound of Fekete-Szegö functional for $p-$valent functions, which for $p = 1$ gives

$$|a_3 - \lambda a_2^2| \leq \begin{cases} \dfrac{1}{2}(B_1^2 + B_2 - 2\lambda B_1^2), & \text{if } 2\lambda B_1^2 \leq B_1^2 + B_2 - B_1; \\ \\ \dfrac{1}{2}B_1, & \text{if } B_1^2 + B_2 - B_1 \leq 2\lambda B_1^2 \leq B_1^2 + B_2 + B_1; \\ \\ \dfrac{1}{2}(-B_1^2 - B_2 + 2\lambda B_1^2), & \text{if } 2\lambda B_1^2 \geq B_1^2 + B_2 + B_1. \end{cases}$$

Since $|B_2| \geq B_1$, hence the above inequality directly yields

$$|a_3 - \frac{1}{2}a_2^2| \leq \frac{|B_2|}{2}. \tag{2.7}$$

From (1.5), we obtain

$$|\gamma_1^2 - \gamma_2^2| = \left|\frac{1}{4}\left(a_2^2 - \left(a_3 - \frac{a_2^2}{2}\right)^2\right)\right| \leq \frac{1}{4}\left(|a_2|^2 + \left|a_3 - \frac{a_2^2}{2}\right|^2\right). \tag{2.8}$$

The required bound follows from (2.8) by using the bounds of $|a_2|$ and $|a_3 - (a_2^2)/2|$ from (2.6) and (2.7) respectively.

To show the sharpness of the bound, consider the analytic function $k_\varphi : \mathbb{D} \to \mathbb{C}$ given by

$$k_\varphi(z) = z \exp \int_0^z \frac{\varphi(it) - 1}{t} dt = z + iB_1 z^2 - \frac{1}{2}(B_1^2 + B_2)z^3 + \cdots . \tag{2.9}$$

Clearly, $k_\varphi \in \mathcal{S}^*(\varphi)$ and for this function, a simple computation gives

$$|\gamma_1^2 - \gamma_2^2| = \frac{4B_1^2 + B_2^2}{16},$$

which shows that the bound is sharp.

Theorem 2.2. *Let $\varphi(z) = 1 + B_1 z + B_2 z^2 + B_3 z^3 + \cdots$ and $f \in \mathcal{C}(\varphi)$. If $|B_2 + \frac{1}{4}B_1^2| \geq B_1$, then*

$$|\gamma_1^2 - \gamma_2^2| \leq \frac{B_1^2}{16} + \frac{1}{144}\left(B_2 + \frac{B_1^2}{4}\right)^2. \tag{2.10}$$

The estimate is sharp.

Proof. Suppose $f \in \mathcal{C}(\varphi)$ be of the form (1.1). Then there exists a Schwarz function $\omega(z) = \sum_{n=1}^{\infty} c_n z^n$ such that

$$1 + \frac{zf''(z)}{f'(z)} = \varphi(\omega(z)), \quad z \in \mathbb{D}.$$

After comparing the coefficients of identical powers of z with the Taylor series expansion of f, φ and ω in the above equation, the coefficients a_2 and a_3 can be expressed as

$$a_2 = \frac{B_1 c_1}{2}, \quad a_3 = \frac{1}{6}(B_1^2 c_1^2 + B_2 c_1^2 + B_1 c_2) \tag{2.11}$$

and

$$a_4 = \frac{1}{12}((4B_1^3 + 3B_1 B_2 + B_3)c_1^3 + (3B_1^2 + 2B_2)c_1 c_2 + B_1 c_3). \tag{2.12}$$

Applying the bound $|c_n| \leq 1$, we obtain

$$|a_2| \leq \frac{B_1}{2}. \tag{2.13}$$

For $f \in \mathcal{C}(\varphi)$, Ma and Minda [17, Theorem 3] established the following bound

$$|a_3 - \lambda a_2^2| \leq \begin{cases} \dfrac{1}{6}(B_2 - \frac{3}{2}\lambda B_1^2 + B_1^2), & \text{if } 3\lambda B_1^2 \leq 2(B_1^2 + B_2 - B_1); \\ \dfrac{1}{6}B_1, & \text{if } 2(B_1^2 + B_2 - B_1) \leq 3\lambda B_1^2 \leq 2(B_1^2 + B_2 + B_1); \\ \dfrac{1}{6}(-B_2 + \frac{3}{2}\lambda B_1^2 - B_1^2), & \text{if } 2(B_1^2 + B_2 + B_1) \leq 3\lambda B_1^2. \end{cases}$$

Since $|B_2 + \frac{1}{4}B_1^2| \geq B_1$ holds, the above inequality directly gives

$$|a_3 - \frac{1}{2}a_2^2| \leq \frac{1}{6}|B_2 + \frac{1}{4}B_1^2|. \tag{2.14}$$

Using the bounds of $|a_2|$ and $|a_3 - (a_2^2)/2|$ for $f \in \mathcal{C}(\varphi)$ given in (2.13) and (2.14), respectively, we obtain

$$|\gamma_1^2 - \gamma_2^2| \leq \frac{1}{4}\left(|a_2|^2 + \left|a_3 - \frac{a_2^2}{2}\right|^2\right) \leq \frac{B_1^2}{16} + \frac{1}{144}\left(B_2 + \frac{B_1^2}{4}\right)^2.$$

The equality case in (2.10) holds for the function h_φ given by

$$1 + \frac{zh_\varphi''(z)}{h_\varphi'(z)} = \varphi(iz). \tag{2.15}$$

Clearly, $h_\varphi \in \mathcal{C}(\varphi)$ and for this function, we have

$$\gamma_1 = \frac{iB_1}{4} \quad \text{and} \quad \gamma_2 = -\frac{1}{12}(B_2 + \frac{B_1^2}{4}),$$

which shows that the bound in (2.10) is sharp.

Theorem 2.3. *Let $\varphi(z) = 1 + B_1z + B_2z^2 + B_3z^3 + \cdots$ and $f \in \mathcal{S}^*(\varphi)$. If $|B_2| \geq B_1$ and $(\mu_1, \nu_1) \in \cup_{i=1}^3 D_i$ hold, then*

$$|\gamma_2^2 - \gamma_3^2| \leq \frac{1}{144}(9B_2^2 + 4B_3^2),$$

where $\mu_1 = 2B_2/B_1$ and $\nu_1 = B_3/B_1$. The bound is sharp.

Proof. Suppose $f \in \mathcal{S}^*(\varphi)$ be of the form (1.1). Then from (1.5), we have

$$\begin{aligned} |\gamma_2^2 - \gamma_3^2| &= \frac{1}{4}\left|\left(a_3 - \frac{a_2^2}{2}\right)^2 - \left(\frac{a_2^3}{3} - a_2a_3 + a_4\right)^2\right| \\ &\leq \frac{1}{4}\left(\left|a_3 - \frac{a_2^2}{2}\right|^2 + \left|\frac{a_2^3}{3} - a_2a_3 + a_4\right|^2\right). \end{aligned} \tag{2.16}$$

From (2.4) and (2.5) for $f \in \mathcal{S}^*(\varphi)$, using the values of a_2, a_3 and a_4 ,we obtain

$$\left|\frac{a_2^3}{3} - a_2a_3 + a_4\right| = \frac{B_1}{3}|c_3 + \mu_1c_1c_2 + \nu_1c_1^3|,$$

where $\mu_1 = 2B_2/B_1$ and $\nu_1 = B_3/B_1$. Since $|B_2| \geq B_1$ holds, therefore (μ_1, ν_1) is a member of either D_1, D_2 or D_3. Thus, from Lemma 1.1, we get

$$\left|\frac{a_2^3}{3} - a_2a_3 + a_4\right| \leq \frac{|B_3|}{3}. \tag{2.17}$$

Using the bounds from (2.7) and (2.17) in the inequality (2.16), the required bound is obtained.

The sharpness of the bound can be seen by the function k_φ given by (2.9). As for this function, we have $\gamma_2 = -B_2/4$, $\gamma_3 = -iB_3/6$ and

$$\gamma_2^2 - \gamma_3^2 = \frac{1}{144}(9B_2^2 + 4B_3^2),$$

which proves the sharpness.

Theorem 2.4. *Let $\varphi(z) = 1 + B_1z + B_2z^2 + B_3z^3 + \cdots$ and $f \in \mathcal{C}(\varphi)$. If $|B_2 + \frac{1}{4}B_1^2| \geq B_1$ and $(\mu_2, \nu_2) \in \cup_{i=1}^3 D_i$ holds, then*

$$|\gamma_2^2 - \gamma_3^2| \leq \frac{B_1^4 + 8B_1^2B_2 + 16B_2^2 + B_1^2B_2^2 + 4B_1B_2B_3 + 4B_3^2}{2304},$$

where $\mu_2 = (B_1^2 + 4B_2)/(2B_1)$ and $\nu_2 = (B_1B_2 + 2B_3)/(2B_1)$. The bound is sharp.

Proof. In view of the equations (2.11) and (2.12) for $f(z) = z + \sum_{n=2}^{\infty} a_nz^n \in \mathcal{C}(\varphi)$, we have

$$\left|\frac{a_2^3}{3} - a_2a_3 + a_4\right| = \frac{B_1}{12}\left|c_3 + \mu_2c_1c_2 + \nu_2c_1^3\right|.$$

As by the hypothesis $|B_2 + \frac{1}{4}B_1^2| \geq B_1$ holds, therefore (μ_2, ν_2) belongs to either D_1, D_2 or D_3. Hence, from Lemma 1.1, we obtain

$$\left|\frac{a_2^3}{3} - a_2a_3 + a_4\right| \leq \frac{|B_1B_2 + 2B_3|}{24}. \tag{2.18}$$

Applying the bound from (2.14) and (2.18) in the inequality (2.16), we get

$$|\gamma_2^2 - \gamma_3^2| \leq \frac{B_1^4 + 8B_1^2B_2 + 16B_2^2 + B_1^2B_2^2 + 4B_1B_2B_3 + 4B_3^2}{2304}.$$

It is a simple exercise to check that the equality case holds for the function $h_\varphi \in \mathcal{C}(\varphi)$ given by (2.15).

2.1 Some Special Cases

Since the classes $\mathcal{S}^*(\varphi)$ and $\mathcal{C}(\varphi)$ generalize various subclasses of starlike and convex functions, therefore, for the appropriate choice of φ, whenever the Taylor series coefficients of φ satisfy the conditions in Theorem 2.1-2.4, we obtain the sharp bounds of $|T_{2,1}(\gamma_f)|$ and $|T_{2,2}(\gamma_f)|$ for the corresponding class.

In case of $\varphi(z) = (1 + Az)/(1 + Bz)$ $(-1 \leq B < A \leq 1)$, we have $\mathcal{S}^*[A, B] = \mathcal{S}^*((1 + Az)/(1 + Bz))$ and $\mathcal{C}[A, B] = \mathcal{C}((1 + Az)/(1 + Bz))$. The series expansion of $(1 + Az)/(1 + Bz)$ shows that $B_1 = (A - B)$, $B_2 = B^2 - AB$ and $B_3 = AB^2 - B^3$. Thus, Theorem 2.1-2.4 lead us to the following:

Corollary 2.5. *Let $f \in \mathcal{S}^*[A, B]$ be of the form (1.1), where $-1 \leq B < A \leq 1$.*

(i) If $|B^2 - AB| \geq A - B$, then

$$|\gamma_1^2 - \gamma_2^2| \leq \frac{(A - B)^2(4 + B^2)}{16}.$$

(ii) If $|B^2 - AB| \geq A - B$, and $(\mu_1, \nu_1) \in \cup_{i=1}^3 D_i$, then

$$|\gamma_2^2 - \gamma_3^2| \leq \frac{(A - B)^2B^2(4B^2 + 9)}{144},$$

where $\mu_1 = -2B$ and $\nu_1 = B^2$.

Corollary 2.6. *Let $f \in \mathcal{C}[A, B]$ be of the form (1.1), where $-1 \leq B < A \leq 1$.*

(i) If $|A^2 - 6AB + 5B^2| \geq 4(A - B)$, then

$$|\gamma_1^2 - \gamma_2^2| \leq \frac{(A - B)^2(A^2 + 25B^2 - 10AB + 144)}{2304}.$$

(ii) If $|A^2 - 6AB + 5B^2| \geq 4(A - B)$ and $(\mu_2, \nu_2) \in \cup_{i=1}^3 D_i$, then

$$|\gamma_2^2 - \gamma_3^2| \leq \frac{(A - B)^2(A^2(B^2 + 1) + B^2(9B^2 + 25) - 2AB(3B^2 + 5))}{2304},$$

where $\mu_2 = (A - 5B)/2$ and $\nu_2 = (B(3B - A))/2$.

By taking $A = 1 - 2\alpha$, $0 \leq \alpha < 1$ and $B = -1$, the following results follow from Corollary 2.5 and Corollary 2.6.

Corollary 2.7. *If $f \in \mathcal{S}^*(\alpha)$, $0 \leq \alpha < 1$, then*

$$|\gamma_1^2 - \gamma_2^2| \leq \frac{5}{16}(2-2\alpha)^2 \ \ \textit{and} \ \ |\gamma_2^2 - \gamma_3^2| \leq \frac{13}{144}(2-2\alpha)^2.$$

Corollary 2.8. *If $f \in \mathcal{C}(\alpha)$, $0 \leq \alpha < 1$, then*

$$|\gamma_1^2 - \gamma_2^2| \leq \frac{(\alpha-1)^2(\alpha^2 - 6\alpha + 45)}{144} \ \ \textit{and} \ \ |\gamma_2^2 - \gamma_3^2| \leq \frac{(\alpha-1)^2(2\alpha^2 - 10\alpha + 13)}{144}.$$

In particular, for $\alpha = 0$, Corollary 2.7 and Corollary 2.8 give the bounds for the classes $\mathcal{S}^*$ and $\mathcal{C}$ respectively.

Corollary 2.9. *If $f \in \mathcal{S}^*$, then*

$$|\gamma_1^2 - \gamma_2^2| \leq \frac{5}{4} \ \ \textit{and} \ \ |\gamma_2^2 - \gamma_3^2| \leq \frac{13}{36}.$$

Corollary 2.10. *If $f \in \mathcal{C}$, then*

$$|\gamma_1^2 - \gamma_2^2| \leq \frac{5}{16} \ \ \textit{and} \quad |\gamma_2^2 - \gamma_3^2| \leq \frac{13}{144}.$$

3 Bounds of $|\det T_{3,2}(f)|$

Ali et al. [4, Theorem 1] derived the sharp estimates of Fekete-Szegö functional for $p-$valent functions belonging to $\mathcal{S}^*(\varphi)$, which for $p = 1$ immediately gives the following estimates of $|a_4|$.

Lemma 3.1. *[4, Theorem 1] Let $\varphi(z) = 1 + B_1 z + B_2 z^2 + B_3 z^3 + \cdots$, and*

$$q_1 = \frac{3B_1^2 + 4B_2}{2B_1}, \quad q_2 = \frac{B_1^3 + 3B_1B_2 + 2B_3}{2B_1}.$$

If $f \in \mathcal{S}^*(\varphi)$ is of the form (1.1) such that $(q_1, q_2) \in \cup_{i=1}^3 D_i$, then

$$|a_4| \leq \frac{B_1^3 + 3B_1B_2 + 2B_3}{6}.$$

The bound is sharp.

Theorem 3.1. *Let $\varphi(z) = 1 + B_1 z + B_2 z^2 + B_3 z^3 + \cdots$ such that*

$$6B_1^2 \leq B_1(3B_1^2 + 2B_2) \leq B_1^2 + 2B_1^4 + 3B_1^2B_2 + 3B_2^2 - 2B_1B_3,$$

and

$$q_1 = \frac{3B_1^2 + 4B_2}{2B_1}, \quad q_2 = \frac{B_1^3 + 3B_1B_2 + 2B_3}{2B_1}.$$

If $f \in \mathcal{S}^*(\varphi)$ and $(q_1, q_2) \in \cup_{i=1}^3 D_i$, then

$$|T_{3,2}(f)| \leq \left(B_1 + \frac{B_1^3 + 3B_1B_2 + 2B_3}{6}\right)\left(B_1^2 + \frac{B_1^4}{3} + \frac{B_1^2B_2}{2} + \frac{B_2^2}{2} - \frac{B_1B_3}{3}\right).$$

The bound is sharp.

Proof. Let $f \in \mathcal{S}^*(\varphi)$ be of the form (1.1). Then from (2.1), we have

$$zf'(z) = f(z)\varphi(\omega(z)), \quad z \in \mathbb{D}.$$

Corresponding to the Schwarz function ω, there exists $p(z) = 1 + \sum_{n=1}^{\infty} p_n z^n \in \mathcal{P}$ such that $w(z) = (p(z)-1)/(p(z)+1)$. The comparison of identical powers of z using the power series expansions of f, φ and p yield

$$a_2 = \frac{B_1 p_1}{2}, \ a_3 = \frac{1}{8}(B_1^2 - B_1 + B_2)p_1^2 + 2B_1 p_2)$$

and

$$a_4 = \frac{1}{48}\bigg((B_1^3 - 3B_1^2 + 2B_1 - 4B_2 + 3B_1B_2 + 2B_3)p_1^3 + (6B_1^2 - 8B_1 + 8B_2)p_1p_2 + 8B_1p_3\bigg).$$

Using these values of a_2, a_3 and a_4 in terms of p_1, p_2 and p_3, it follows that

$$\begin{aligned}|a_2^2 - 2a_3^2 + a_2a_4| = \bigg| &\frac{B_1^2p_1^2}{4} - \frac{(B_1^2 - 3B_1^3 + 2B_1^4 - 2B_1B_2 + 3B_1^2B_2 + 3B_2^2 - 2B_1B_3)p_1^4}{96} \\ &- \frac{B_1(3B_1^2 - 2B_1 + 2B_2)p_1^2p_2}{48} - \frac{B_1^2}{8}p_2^2 + \frac{B_1^2}{12}p_1p_3\bigg|.\end{aligned}$$

Keeping in mind that $B_1^2 + 2B_1^4 + 3B_1^2B_2 + 3B_2^2 - 2B_1B_3 - B_1(3B_1^2 + 2B_2) \geq 0$ and by applying the bound $|p_n| \leq 2$, $n \in \mathbb{N}$ (see [8, Page- 41]), we get

$$\begin{aligned}|a_2^2 - 2a_3^2 + a_2a_4| \leq &\frac{3B_1^2}{2} + \frac{(B_1^2 - 3B_1^3 + 2B_1^4 - 2B_1B_2 + 3B_1^2B_2 + 3B_2^2 - 2B_1B_3)}{6} \\ &+ \frac{B_1^2}{6}\bigg|p_3 - \bigg(\frac{3B_1^2 - 2B_1 + 2B_2}{4B_1}\bigg)p_1p_2\bigg|.\end{aligned}$$

Since $3B_1^2 + 6B_2 \geq 6B_1$, therefore from Lemma 1.2, we obtain

$$|a_2^2 - 2a_3^2 + a_2a_4| \leq B_1^2 + \frac{B_1^4}{3} + \frac{B_1^2B_2}{2} + \frac{B_2^2}{2} - \frac{B_1B_3}{3}. \tag{3.1}$$

Further, we have $|a_2 - a_4| \leq |a_2| + |a_4|$. Using the bounds of $|a_2|$ and $|a_4|$ from (2.6) and Lemma 3.1 respectively, we get

$$|a_2 - a_4| \leq B_1 + \frac{B_1^3 + 3B_1B_2 + 2B_3}{6}.$$

From (1.3), a simple computation reveals that

$$|T_{3,2}(f)| = |(a_2 - a_4)(a_2^2 - 2a_3^2 + a_2a_4)|. \tag{3.2}$$

The required estimated is determined by putting the bounds given in (3.1) and (3.2) in the above equation.

The function k_φ defined by (2.9) plays the role of extremal functions. As for this function, we have

$$a_2 = iB_1, \quad a_3 = -\frac{1}{2}(B_1^2 + B_2), \quad a_4 = -\frac{i}{6}(B_1^3 + 3B_1B_2 + 2B_3)$$

and

$$|T_{3,2}(k_\phi)| = \bigg(B_1 + \frac{B_1^3 + 3B_1B_2 + 2B_3}{6}\bigg)\bigg(B_1^2 + \frac{B_1^4}{3} + \frac{B_1^2B_2}{2} + \frac{B_2^2}{2} - \frac{B_1B_3}{3}\bigg)$$

proving the sharpness.

Theorem 3.2. *Let* $\varphi(z) = 1 + B_1z + B_2z^2 + B_3z^3 + \cdots$ *such that*

$$16B_1^2 - 4B_1B_2 \leq 7B_1^3 \leq 5B_1^4 + 2B_1^2 - 4B_1B_2 + 7B_1^2B_2 + 8B_2^2 - 6B_1B_3, \tag{3.3}$$

and

$$q_1 = \frac{3B_1^2 + 4B_2}{2B_1}, \quad q_2 = \frac{B_1^3 + 3B_1B_2 + 2B_3}{2B_1}.$$

If $f \in \mathcal{C}(\varphi)$ *and* $(q_1, q_2) \in \cup_{i=1}^3 D_i$, *then*

$$|T_{3,2}(f)| \leq \frac{1}{144}\left(\frac{B_1}{2} + \frac{B_1^3 + 3B_1B_2 + 2B_3}{24}\right)(5B_1^4 + 36B_1^2 + 7B_1^2B_2 + 8B_2^2 - 6B_1B_3).$$

The bound is sharp.

Proof. Suppose $f \in \mathcal{C}(\varphi)$ be of the form (1.1), then we have

$$f'(z) + zf''(z) = f'(z)\varphi(\omega(z)).$$

Corresponding to the Schwarz function $\omega(z) = \sum_{n=1}^\infty c_nz^n$, there exists $p(z) = 1 + \sum_{n=1}^\infty p_nz^n \in \mathcal{P}$ such that $w(z) = (p(z) - 1)/(p(z) + 1)$. The comparison of same powers of z in the above equation after the series expansions yield that

$$a_2 = \frac{B_1p_1}{4}, \quad a_3 = \frac{1}{24}((B_1^2 - B_1 + B_2)p_1^2 + 2B_1p_2)$$

and

$$a_4 = \frac{1}{192}\bigg((B_1^3 - 3B_1^2 + 2B_1 - 4B_2 + 3B_1B_2 + 2B_3)p_1^3 + (6B_1^2 + 8B_2 - 8B_1)p_1p_2 + 8B_1p_3\bigg). \tag{3.4}$$

Using these expressions for a_2, a_3 and a_4 in terms of the coefficients p_1, p_2 and p_3, a simple computation gives

$$\begin{aligned}|a_2^2 - 2a_3^2 + a_2a_4| = \bigg|\frac{1}{2304}\bigg(&(2B_1^2 - 7B_1^3 + 5B_1^4 - 4B_1B_2 + 7B_1^2B_2 + 8B_2^2 - 6B_1B_3)p_1^4 \\ &+ 32B_1^2p_2^2 - 144B_1^2p_1^2 - 24B_1^2p_1\left(p_3 - \frac{(14B_1^3 - 8B_1^2 + 8B_1B_2)}{24B_1^2}p_1p_2\right)\bigg)\bigg|.\end{aligned}$$

In view of the hypothesis $2B_1^2 + 5B_1^4 - 4B_1B_2 + 7B_1^2B_2 + 8B_2^2 - 6B_1B_3 \geq 7B_1^3$ and by applying the bound $|p_n| \leq 2$ $(n \in \mathbb{N})$, we get

$$\begin{aligned}|a_2^2 - 2a_3^2 + a_2a_4| \leq \frac{1}{2304}\bigg(&16(2B_1^2 - 7B_1^3 + 5B_1^4 - 4B_1B_2 + 7B_1^2B_2 + 8B_2^2 - 6B_1B_3) \\ &+ 128B_1^2 + 576B_1^2 + 48B_1^2\left(\left|p_3 - \frac{(14B_1^3 - 8B_1^2 + 8B_1B_2)}{24B_1^2}p_1p_2\right|\right)\bigg).\end{aligned}$$

Since $7B_1^2 + 4B_2 \geq 16B_1$ holds, therefore from Lemma 3.1, it follows that

$$|a_2^2 - 2a_3^2 + a_2a_4| \leq \frac{1}{144}(36B_1^2 + 5B_1^4 + 7B_1^2B_2 + 8B_2^2 - 6B_1B_3). \tag{3.5}$$

Now, we only need to maximize $|a_2 - a_4|$ for $f \in \mathcal{C}(\varphi)$. By the one to one correspondence between the class $\mathcal{P}$ and the class of Schwarz functions, the coefficients a_4 in (3.4) can be expressed as

$$a_4 = \frac{1}{12}B_1(c_3 + q_1c_1c_2 + q_2c_1^3),$$

where $q_1 = (3B_1^2 + 4B_2)/(2B_1)$ and $q_2 = (B_1^3 + 3B_1B_2 + 2B_3)/(2B_1)$. As by the hypothesis $(q_1, q_2) \in \cup_{i=1}^3 D_i$, from Lemma 1.1, we obtain

$$|a_4| \leq \frac{B_1^3 + 3B_1B_2 + 2B_3}{24}. \tag{3.6}$$

Employing the bounds of $|a_2|$ and $|a_4|$ from (2.13) and (3.6) respectively, we get

$$|a_2 - a_4| \leq |a_2| + |a_4| \leq \frac{B_1}{2} + \frac{B_1^3 + 3B_1B_2 + 2B_3}{24}. \tag{3.7}$$

Thus, applying the bounds of $|a_2^2 - 2a_3^2 + a_2a_4|$ and $|a_2 - a_4|$ from (3.5) and (3.7) respectively in (3.2), we get the desired result.

The result is sharp for the function h_φ defined in (2.15). As for this function, we have $a_2 = iB_1/2$, $a_3 = -(B_1^2 + B_2)/6$, $a_4 = -i(B_1^3 + 3B_1B_2 + 2B_3)/24$ and

$$|T_{3,2}(f)| = \frac{1}{144}\left(\frac{B_1}{2} + \frac{B_1^3 + 3B_1B_2 + 2B_3}{24}\right)(5B_1^4 + 36B_1^2 + 7B_1^2B_2 + 8B_2^2 - 6B_1B_3)$$

proving the sharpness of the bound.

3.1 Special Cases

For the classes $\mathcal{S}^*[A,B]$ and $C[A,B]$, we have $\varphi(z) = (1+Az)/(1+Bz)$ and the series expansion gives $B_1 = A - B$, $B_2 = B^2 - AB$ and $B_3 = AB^2 - B^3$. Hence, we deduce the following results immediately from Theorem 3.1 and Theorem 3.2.

Corollary 3.3. *For $-1 \leq B < A \leq 1$, let*

$$6(A-B)^2 \leq (3A - 5B)(A-B)^2 \leq (A-B)^2(2A^2 - 7AB + 6B^2 + 1),$$

and

$$q_1 = \frac{3A - 7B}{2}, \quad q_2 = \frac{A^2 - 5AB + 6B^2}{2}.$$

If $f \in \mathcal{S}^*[A,B]$ and $(q_1, q_2) \in \cup_{i=1}^3 D_i$, then

$$|T_{3,2}(f)| \leq \frac{1}{36}(A-B)^2(2A^2 - 7AB + 6B^2 + 6)(A^3 + 6A - 6B - 6A^2B + 11AB^2 - 6B^3).$$

The estimates is sharp.

Corollary 3.4. *For $-1 \leq B < A \leq 1$, let*

$$4(A-B)^2(4+B) \leq 7(A-B)^3 \leq (A-B)^2(2 + 5A^2 + 4B - 17AB + 14B^2)$$

and

$$q_1 = \frac{3A - 7B}{2}, \quad q_2 = \frac{A^2 - 5AB + 6B^2}{2}.$$

If $f \in \mathcal{C}[A,B]$ and $(q_1, q_2) \in \cup_{i=1}^3 D_i$, then

$$|T_{3,2}(f)| \leq \frac{1}{3456}(A-B)^2(5A^2 - 17AB + 14B^2 + 36)(A^3 + 12A - 12B - 6A^2B + 11AB^2 - 6B^3).$$

The estimates is sharp.

When $A = 1 - 2\alpha$ and $B = -1$, the conditions in Corollary 3.3 and 3.4 are true and $(q_1, q_2) \in D_3$ for $\alpha \in [0, 1/7]$. Thus, we obtain the following bounds for the classes $\mathcal{S}^*(\alpha)$ and $\mathcal{C}(\alpha)$.

Corollary 3.5. *If $f \in \mathcal{S}^*(\alpha)$, then*

$$|T_{3,2}(f)| \leq \frac{4}{9}(1-\alpha)^3(16\alpha^4 - 100\alpha^3 + 268\alpha^2 - 345\alpha + 189)$$

for $\alpha \in [0, 1/7]$. The bound is sharp.

Corollary 3.6. *If $f \in \mathcal{C}(\alpha)$, then*

$$|T_{3,2}(f)| \leq \frac{1}{108}(1-\alpha)^3(20\alpha^4 - 124\alpha^3 + 381\alpha^2 - 576\alpha + 432)$$

for $\alpha \in [0, 1/7]$. The bound is sharp.

Remark 3.1. In particular, when $\alpha = 0$, the following bounds for the classes $\mathcal{S}^*$ and $\mathcal{C}$ follow as special case proved in [3].

(i) If $f \in \mathcal{S}^*$, then $|T_{3,2}(f)| \leq 84$ [3, Theorem 2.3].

(ii) If $f \in \mathcal{C}$, then $|T_{3,2}(f)| \leq 4$ [3, Theorem 2.11].

In case of $\varphi(z) = ((1+z)/(1-z))^\beta$, $0 < \beta \leq 1$, the classes $\mathcal{S}^*(\varphi)$ and $\mathcal{C}(\varphi)$ reduce to the class of strongly starlike functions of order β and the class of strongly convex functions of order β, denoted by $\mathcal{SS}^*(\beta)$ and $\mathcal{CC}(\beta)$ respectively (see [8]).

Corollary 3.7. *If $f \in \mathcal{SS}^*(\beta)$, then*

$$|T_{3,2}(f)| \leq \frac{4}{81}\beta^3(160 + 742\beta^2 + 799\beta^4)$$

for $\beta \in [3/4, 1]$. The bound is sharp.

Corollary 3.8. *If $f \in \mathcal{CC}(\beta)$, then*

$$|T_{3,2}(f)| \leq \frac{1}{324}\beta^3(323 + 650\beta^2 + 323\beta^4)$$

for $\beta \in [8/9, 1]$. The bound is sharp.

For $-1/2 < \lambda \leq 1$ and $f \in \mathcal{A}$ such that f is a locally univalent functions, Robertson [22] considered the class

$$\mathcal{F}(\lambda) = \left\{ f \in \mathcal{A} : \operatorname{Re}\left(1 + \frac{zf''(z)}{f'(z)}\right) > \frac{1}{2} - \lambda \right\}.$$

Clearly, when $1/2 \leq \lambda \leq 1$, functions in $\mathcal{F}(\lambda)$ are close-to-convex [12]. For $-1/2 < \lambda \leq 1/2$, the functions in $\mathcal{F}(\lambda)$ are convex. Vasudevarao et al. [26] derived the sharp bound of $|T_{3,2}(f)|$ for $f \in \mathcal{F}(\lambda)$ when $1/2 \leq \lambda \leq 1$, that is the class of Ozaki close-to-convex functions. Consider

$$\varphi_\lambda(z) = \frac{1 + 2\lambda z}{1 - z}, \quad z \in \mathbb{D}.$$

The function φ_λ maps the unit disk onto the right half plane for $-1/2 < \lambda \leq 1/2$ such that $\operatorname{Re}\varphi_\lambda > (1/2 - \lambda)$. Clearly, $\mathcal{C}(\varphi_\lambda) \subset \mathcal{F}(\lambda)$ for $\lambda \in (-1/2, 1]$ and $\mathcal{C}(\varphi_\lambda) = \mathcal{F}(\lambda)$ when $\lambda \in (-1/2, 1/2]$. The Taylor's series expansion of φ_λ gives $B_1 = B_2 = B_3 = (1 + 2\lambda)$, which satisfy the condition (3.3) for $\lambda \in [5/14, 1/2]$. Thus, from Theorem 3.2, we obtain the following sharp bound of $|T_{3,2}(f)|$ for the class $\mathcal{F}(\lambda)$ when $5/14 \leq \lambda \leq 1/2$.

Corollary 3.9. *If $f \in \mathcal{F}(\lambda)$, $5/14 \leq \lambda \leq 1/2$, then*

$$|T_{3,2}(f)| \leq \frac{1}{864}(1 + 2\alpha)^3(9 + 5\alpha + 2\alpha^2)(25 + 17\alpha + 10\alpha^2).$$

Remark 3.2. Vasudevarao et al. [26, Theorem 4.3] proved the same bound as given in Corollary 3.9 for $1/2 \leq \lambda \leq 1$. Thus, Corollary 3.9 shows that the result is also true for $5/14 \leq \lambda \leq 1/2$.

Declarations

Funding

The work of the Surya Giri is supported by University Grant Commission, New-Delhi, India under UGC-Ref. No. 1112/(CSIR-UGC NET JUNE 2019).

Conflict of interest

The authors declare that they have no conflict of interest.

Author Contribution

Each author contributed equally to the research and preparation of manuscript.

Data Availability

Not Applicable.

References

[1] E. A. Adegani, N. E. Cho and M. Jafari, Logarithmic coefficients for univalent functions defined by subordination, Mathematics 7, no. 5 (2019), 408.

[2] O. P. Ahuja, K. Khatter and V. Ravichandran, Toeplitz determinants associated with Ma-Minda classes of starlike and convex functions, Iran. J. Sci. Technol. Trans. A Sci. **45** (2021), no. 6, 2021–2027.

[3] M. F. Ali, D. K. Thomas and A. Vasudevarao, Toeplitz determinants whose elements are the coefficients of analytic and univalent functions, Bull. Aust. Math. Soc. **97** (2018), no. 2, 253–264.

[4] R. M. Ali, V. Ravichandran and N. Seenivasagan, Coefficient bounds for p-valent functions, Applied Mathematics and Computation 187.1 (2007): 35-46.

[5] M. F. Ali and A. Vasudevarao, On logarithmic coefficients of some close-to-convex functions, Proc. Amer. Math. Soc. **146** (2018), no. 3, 1131–1142.

[6] N. E. Cho, B. Kowalczyk, O. S. Kwon, A. Lecko and Y. J. Sim, On the third logarithmic coefficient in some subclasses of close-to-convex functions, Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat. RACSAM **114** (2020), no. 2, Paper No. 52, 14 pp.

[7] K. Cudna, O. S. Kwon, A. Lecko, Y. J. Sim, and B. Śmiarowska, The second and third-order Hermitian Toeplitz determinants for starlike and convex functions of order α, Bol. Soc. Mat. Mex. (3) **26** (2020), no. 2, 361–375.

[8] P. L. Duren, Univalent functions, Grundlehren der mathematischen Wissenschaften, 259, Springer, New York, 1983.

[9] I. Efraimidis, A generalization of Livingston's coefficient inequalities for functions with positive real part, J. Math. Anal. Appl. **435** (2016), no. 1, 369–379.

[10] S. Giri and S. S. Kumar, Hermitian-Toeplitz determinants for certain univalent functions, Anal. Math. Phys. (*accepted*).

[11] W. Janowski, Some extremal problems for certain families of analytic functions. I, Ann. Polon. Math. **28** (1973), 297–326.

[12] W. Kaplan, Close-to-convex schlicht functions, Michigan Math. J. **1** (1952), 169–185 (1953).

[13] I. R. Kayumov, On Brennan's conjecture for a special class of functions, Math. Notes **78** (2005), no. 3-4, 498–502; translated from Mat. Zametki **78** (2005), no. 4, 537–541.

[14] B. Kowalczyk and A. Lecko, Second Hankel determinant of logarithmic coefficients of convex and starlike functions, Bull. Aust. Math. Soc. **105** (2022), no. 3, 458–467.

[15] B. Kowalczyk and A. Lecko, Second Hankel determinant of logarithmic coefficients of convex and starlike functions of order alpha, Bull. Malays. Math. Sci. Soc. **45** (2022), no. 2, 727–740.

[16] A. Lecko, Y. J. Sim and B. Śmiarowska, The fourth-order Hermitian Toeplitz determinant for convex functions, Anal. Math. Phys. **10** (2020), no. 3, Paper No. 39, 11 pp.

[17] W. Ma and D. Minda, A unified treatment of some special classes of univalent functions, Proceedings of the Conference on Complex Analysis, 1992. International Press Inc., 1992.

[18] I. M. Milin, Univalent Functions and Orthonormal Systems, Izdat. "Nauka", Moscow (1971) (in Russian); English transl., American Mathematical Society, Providence (1977).

[19] M. Mundalia and S. S. Kumar, Coefficient Problems for Certain Close-to-Convex Functions, Bull. Iranian Math. Soc. **49** (2023), no. 1, Paper No. 5.

[20] M. Obradović and N. Tuneski, Hermitian Toeplitz determinants for the class of univalent functions. Armenian Journal of Mathematics, 13(4), 1–10 (2021).

[21] D. V. Prokhorov and J. Szynal, Inverse coefficients for (α, β)-convex functions, Ann. Univ. Mariae Curie-Skłodowska Sect. A **35** (1981), 125–143 (1984).

[22] M. S. Robertson, On the theory of univalent functions, Ann. of Math. (2) **37** (1936), no. 2, 374–408.

[23] D. K. Thomas, On the coefficients of Bazilevič functions with logarithmic growth, Indian J. Math. **57** (2015), no. 3, 403–418.

[24] D. K. Thomas, N. Tuneski, A. Vasudevarao, Univalent Functions: A Primer. Walter de Gruyter GmbH, Berlin (2018).

[25] O. Toeplitz, Zur Transformation der Scharen bilinearer Formen von unendlichvielen Veränderlichen. Mathematischphysikalis- che, Klasse, Nachr. der Kgl. Gessellschaft derWissenschaften zu Göttingen, pp 110–115 (1907).

[26] A. Vasudevarao, A. Lecko and D. K. Thomas, Hankel, Toeplitz, and Hermitian-Toeplitz determinants for certain close-to-convex functions, Mediterr. J. Math. **19** (2022), no. 1, Paper No. 22, 17 pp.

[27] K. Ye and L.-H. Lim, Every matrix is a product of Toeplitz matrices, Found. Comput. Math. **16** (2016), no. 3, 577–598.

[1]Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India
E-mail address: suryagiri456@gmail.com

[*]Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India
E-mail address: spkumar@dtu.ac.in

Eur. Phys. J. D (2023)77:42
https://doi.org/10.1140/epjd/s10053-023-00615-7

THE EUROPEAN
PHYSICAL JOURNAL D

Regular Article – Optical Phenomena and Photonics

Ultra-narrow band perfect absorber for sensing applications in the visible region

Ritika Ranga[1,a], Yogita Kalra[1,b], Kamal Kishor[1,c], and Nishant Shankhwar[2,d]

[1] Advanced Simulation Lab, Centre of Relevance and Excellence in Fiber Optics and Optical Communication, Delhi Technological University, Delhi 110042, India
[2] Hansraj College, University of Delhi, Delhi 110007, India

Received 20 October 2022 / Accepted 16 February 2023

Abstract. Plasmonics is widely used for converting electromagnetic radiation into energy and confining electromagnetic radiation below the diffraction limit. However, the ultra narrowband and high electromagnetic field cannot be obtained simultaneously because of resistive loss and radiation damping in the metals. In this article, a metallic ultra-narrow band perfect absorber has been proposed consisting of an array of four squares on a silver layer. The structure shows more than 99% absorption and full width at half maxima less than 2 nm at resonance wavelength. The absorption mechanism has been revealed by calculating electric and magnetic field profiles. The dependence of the structure on the geometrical parameters has been studied and the structure has thus been optimized at 692 nm i.e. in the visible range of frequency. The proposed structure is then investigated for sensing application. The structure shows high sensitivity of 680 nm/RIU in the visible range of wavelength and a high figure of merit of 348.72.

1 Introduction

Plasmonic materials have attracted much attention in previous years for varying applications in biosensors [1], optical filters [2], and photodetectors [3]. Collective oscillation of free electrons on the surface of metal due to an external electric field is known as localized surface plasmons [4]. The charges/plasmons accumulated at the opposite ends of the nanopatiocle. If gap is created between the two nanoparticles, than it is evident that the order of magnitude of surface plasmons is more in the gap region rather than in the vicinity of elongated nanoparticles [5]. By choosing shape and size, amplitude and bandwidth of resonance wavelength can be controlled. This resonance wavelength also varies with the refractive index of the surrounding environment. This property can be utilised in the sensing application [6]. Plasmonic metamaterials due to localised surface plasmon convert far-field radiation into localised energy or vice-versa. Hence, they can be used in the design of optical antennas [7]. High field enhancement can be obtained in the gap region of plasmonic nanoantennas [8].

On account of being made of metals, plasmonic nanoantennas exhibit substantial ohmic loss, which facilitates their utilisation for absorption of light [9]. By choosing a suitable design and optimising geometrical parameters one can obtain absolute absorption in a range of frequencies, which can be very useful in device applications [10].

Absorbers can be categorized into broadband and narrowband on the basis of bandwidth. Broadband absorbers can be used in solar panels utilized in solar energy harvesting [11, 12]. Narrowband absorbers can be used in sensing [13–15], thermal radiation tailoring [16], absorption filters [17] etc. The first absorber was proposed and demonstrated experimentally by Lardy et al. in 2008 for sensing application in the infrared region [18]. In this design like most of the previous designs, triple layers of metal-dielectric-metal (MDM) have been chosen where a dielectric spacer enables strong plasmonic coupling between the top resonator and bottom metal film [19–21]. The above-mentioned absorbers can be considered as resonators coupled to a transmission line with a dielectric spacer region, whose thickness influences the radiative damping rates and resonance frequency. However, due to strong radiative damping and inherent metal loss, the bandwidth of the absorbers is relatively broad (> 40 nm). Sensitivity and Figure of Merit (FOM) are the two important parameter to specify the performance of sensors. Sensitivity is measure of change in the output quantity with respect

[a] e-mail: ritikaranga821@gmil.com
[b] e-mail: dryogitakalra@gmail.com (corresponding author)
[c] e-mail: kishorkamal1@rediffmail.com
[d] e-mail: nishant.shankhwar@gmail.com

Published online: 15 March 2023

to change in the input quantity. The term ”bulk sensitivity” refers to the shift in the resonance wavelength with respect to the surrounding refractive index that is used to calculate the sensitivity of the presented plasmonic sensors [22–24]. Higher FOM indicates a high ability to sense changes in the environment. It is defined by the ratio of the sensitivity to the bandwidth at half maxima of the peak. Therefore, attempts have been made to reduce the absorption bandwidth in order to increase the corresponding figure of merit (FOM). In 2016, Z. Yong et al. proposed a narrow band absorber for sensing application of FWHM 7.5 nm in the near IR region with FOM 110/RIU [25]. Many absorbers from UV to near IR have been demonstrated till now [20, 26–28]. However, narrowband absorbers in visible range have not been studied much. Most of them show resonance in a broad range of wavelengths and have a very low FOM [10, 15, 29]. Much work has been done by researchers practically and experimentally to obtain high FOM in the visible range. In 2014, Emiko and Tetsu experimentally studied localised surface plasmon resonance sensors based on spectral dips [30]. G. Liu et al. in 2016 proposed theoretically a network type plasmonic nanostructure with multiple reflection bands reaching full width at half maximum of 3 nm and FOM of 68.57/RIU [31]. In 2017, Wu et al. proposed grating absorbers with nanoribbons in between metal dielectric metal (MDM) having sensing applications in the visible range and FOM equal to 233.5/RIU [32]. In 2020, a narrow band perfect absorber of Al_2O_3 based on dielectric-dielectric-metal configuration was proposed by M. Pan et al. having a sensitivity of 108 nm/RIU reaching FOM up to 240.7/RIU [33].

In this article, all metallic plasmonic narrowband absorber has been designed. The design confines and enhances electric as well as magnetic field at a single hot spot and shows a narrow absorption peak of full width at half maximum (FWHM) less than 2 nm in the visible wavelength. The design shows the sensitivity of 680 nm/RIU and FOM of 348.72/RIU, which is higher than the earlier reported results in the visible wavelength [25, 30–33]. Further, the performance of the absorber has been investigated by varying geometrical parameters and the design is optimised for maximum absorption and high FOM. The mechanism of absorption has been revealed by observing electric and magnetic field profiles. The proposed design holds great potential for sensing and near-field optics.

2 Design and modelling

The geometry comprises of four cuboids of silver, facing each other at tips, placed over a silver layer of thickness (t) of 100 nm. The whole structure is placed on a glass substrate and surrounded by air on top. The design and geometrical parameters are demonstrated in Fig. 1. Length and width of cuboids have been taken as ‘a’ and height has been taken as ‘h’. The gap ‘g’ between the tips of the opposite cuboid has been taken as 30 nm. The set of four cuboids has been arranged in the form of an array of periodicity ‘p’.

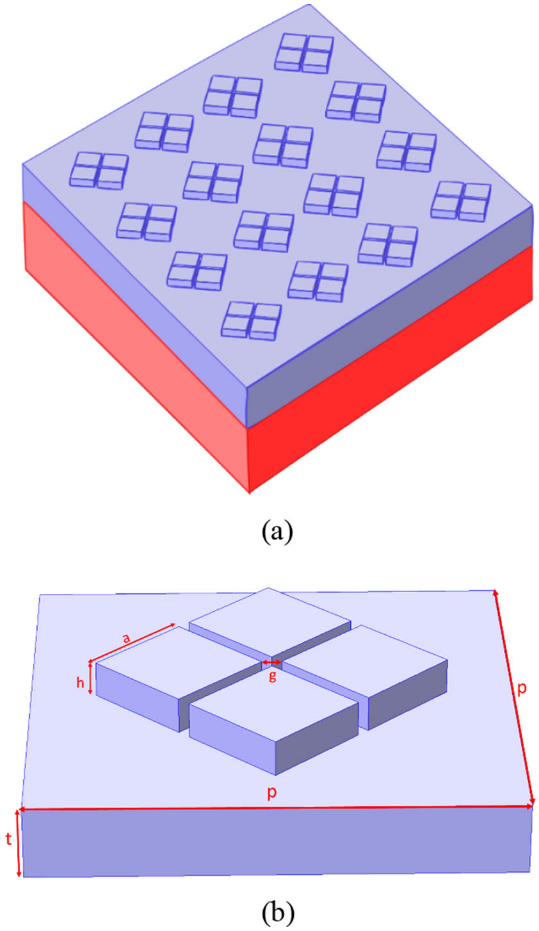


Fig. 1 **a** Schematic of the proposed metallic absorber, **b** unit cell with geometrical parameters

Silver has been chosen as the material because of its high reflectivity. The thickness of the silver layer has been chosen such that no light is transmitted through this layer and the maximum absorption takes place at resonance wavelength. The value of relative permittivity for silver has been taken from literature [34].

The transverse magnetic (TM) light, polarised along x-direction is allowed to fall on the surface of the structure from the top. The distance between light source and the structure is taken more than half of resonant wavelength. The simulation is performed on single unit enclosed in a cuboid having dimensions of periodicity, p along x and y axis. The periodic boundary conditions have been applied along x and y plane on the boundary of cuboid enclosing unit cell to represent as infinite array in x–y plane. The refractive index of glass substrate is taken to be 1.5. The maximum mesh size is taken as 10 nm. The reflection coefficient has been calculated using the finite element method via COMSOL Multiphysics. The absorption coefficient has been obtained by using relationship A = 1 − R.

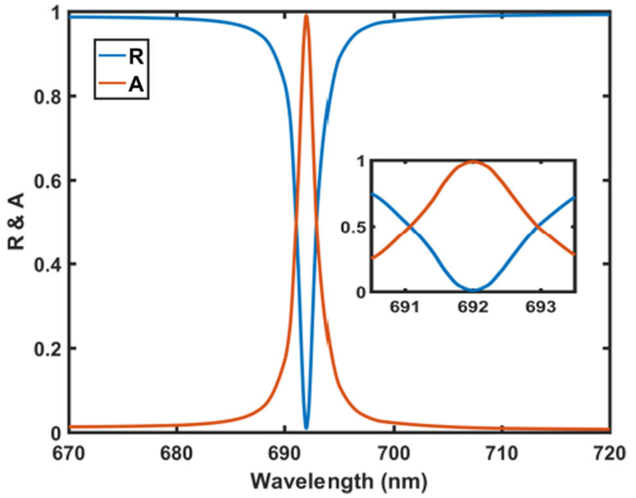


Fig. 2 Refraction and absorption coefficient for a = 140 nm, h = 30 nm and p = 650 nm showing resonance at 692 nm and FWHM = 1.95 nm

The resonance wavelength of plasmonic structures depends upon shape, size, geometry, constituting material as well as the surrounding medium. This property of changing resonance wavelength with refractive index can be used in sensing applications. The sensors are characterised by their sensitivity and FOM. Sensitivity, S for refractive index sensor is defined by the ratio of the spectral shift with the change of refractive index of the surrounding medium and FOM is the ratio of sensitivity to the FWHM [22–24].

$$S = \frac{\Delta\lambda}{\Delta n} \quad \text{FOM} = \frac{S}{\text{FWHM}}$$

3 Results and discussion

Absorption and reflection spectra for the structure has been shown in Fig. 2. As the thickness of the silver film is greater than the skin depth, the absorption is reduced to A = 1-R. The absorption spectrum shows an absorption peak at 692 nm of which the full width at half maxima is 1.95 nm, which is much narrower than the previously reported design in the visible range [10, 15, 29–31]. A maximum of 99.45% absorption is obtained at the resonance wavelength. Further, the quality factor Q given by the ratio of the resonance wavelength to the FWHM of resonance peak ($\lambda_{res}/\Delta\lambda$) which comes out to be 354.87 is much higher than previously reported designs [25, 30–33].

3.1 Effect of length of the nanoantenna

3.1.1 Effect of height, ‘h’

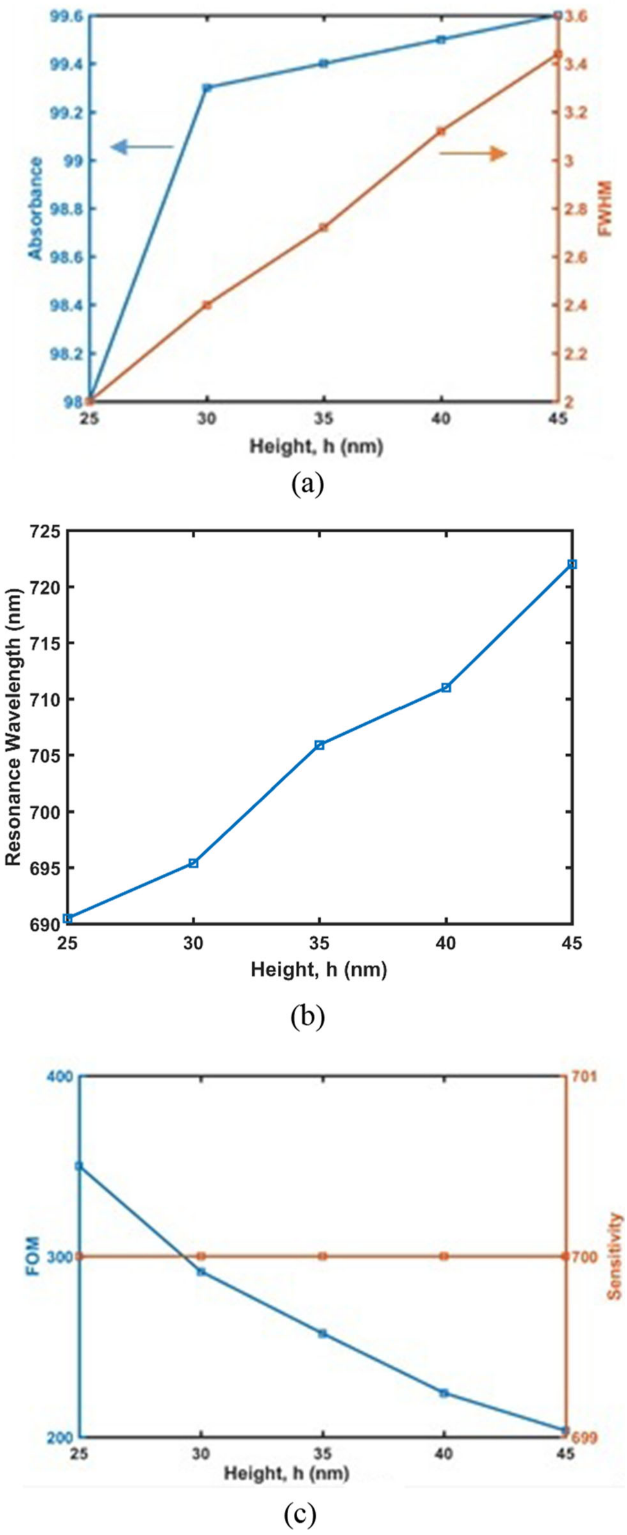
The study of dependence of the size of cuboidal pillars on the absorption peak has been studied. First, the height of the silver cuboids has been varied, and the periodicity and sides of cuboids have been fixed at 650 nm and 160 nm respectively. The plot showing the variation of absorption coefficient and FWHM with the variation of height from 25 to 45 nm in the step of 5 nm, has been depicted in Fig. 3a and the variation of resonance wavelength with height has been shown in Fig. 3b. As there is a vertical waveguide in between four silver pillars, the increment in the height eventually increases the length of the waveguide. Therefore, the resonance shifts toward longer wavelengths, as seen in Fig. 3b. Although FWHM decreases with a decrease in the height of silver pillars, but below 30 nm height, the absorption decreases. Hence, the height has been fixed at 30 nm. Further, the effect of height of the pillars on the sensitivity has been studied by keeping all other parameters same as above and corresponding FOM has been calculated. The graph showing variation of the sensitivity and FOM of the structure with the variation in height of the nanopillars of the nanoantennas has been shown in Fig. 3c. It has been observed from the figure that there is no effect of height of the nanopillars on the sensitivity of structure but there is fall in the FOM of the structure with increase in the height because of increase in the value of FWHM with the height of nanopillars.

An increase in the value of parameter ‘h’ increases the absorption. The increase in the values of ‘h’ increases the volume of cuboidal pillars on the silver layer. This increases the regions of high E-field alongside the wall of the cuboidal pillars. The increase in the high E-field regions increases the shift in the plasmon resonance wavelength Fig. 3b. However, a higher value of ‘h’ also reduces the probability of coupling surface plasmons on top of the cuboid pillars as the evanescent tail of the plasmon will no longer be able to reach the top surface [23]. This may reduce the values of sensitivity. The FWHM is increasing with height resulting in a constant sensitivity. To achieve higher absorption, sensitivity and FOM, the height ‘h’ was adopted as 30 nm.

3.1.2 Effect of side length, ‘a’

Further, the sides of silver cubes have varied from 120 to 160 nm in the step size of 10 nm, keeping the periodicity and height of cuboids at 650 nm and 30 nm respectively. The effect of varying sides of cubes on absorption coefficient and FWHM has been shown in Fig. 4a and the shift in the resonance wavelength has been shown in Fig. 4b. It has been observed that with the increase in ‘a’, there is a slight redshift in the resonance wavelength, and full width at half maxima decreases. Red shift in the resonance wavelength with the increase in side length ‘a’ can be understood in analogy to a half-wave dipole antenna whose resonant wavelength is directly proportional to antenna length, given by,

$$\lambda_{\text{resonance}} = 2 \times \text{length}$$

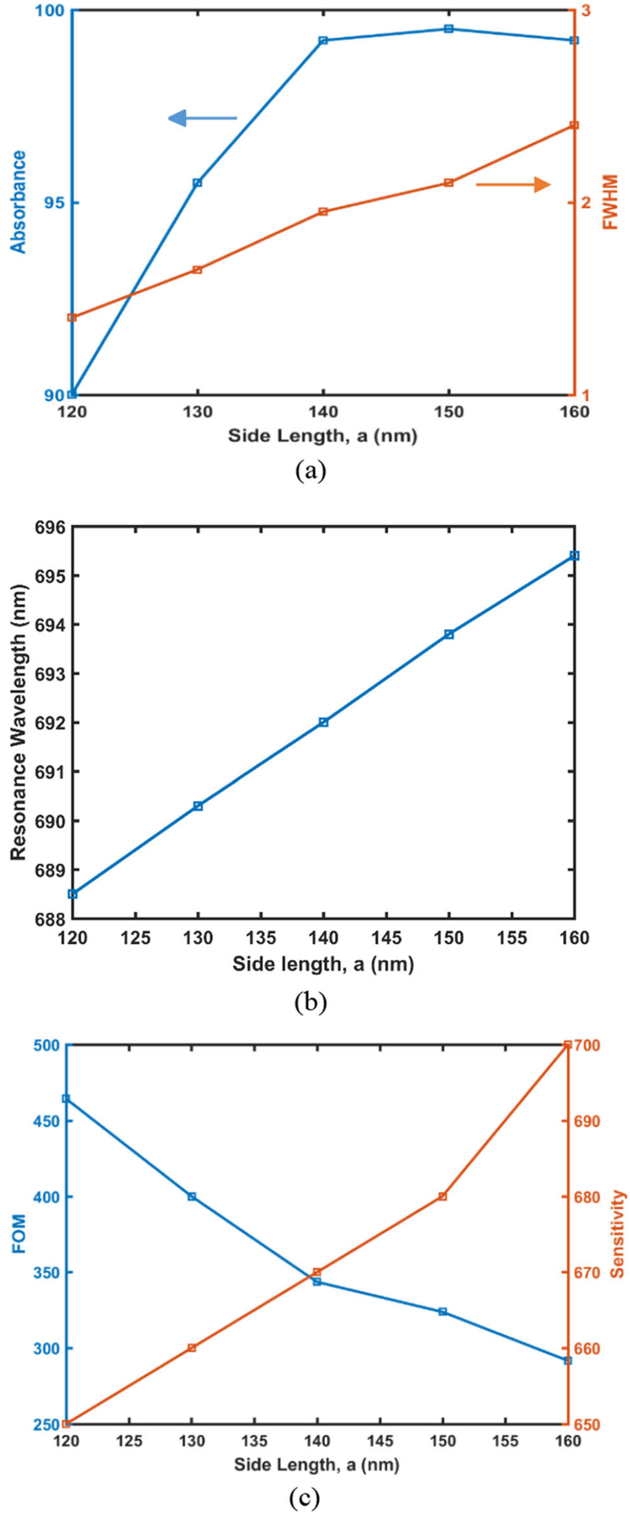


(a)

(b)

(c)

Fig. 3 Variation of **a** absorption coefficient and FWHM, **b** resonance wavelength and **c** FOM and sensitivity with the variation of height, ‘h’



(a)

(b)

(c)

Fig. 4 Variation of **a** absorption coefficient and FWHM, **b** resonance wavelength and **c** FOM and sensitivity with the variation of sides of pillars, ‘a’

As the length of the dipole increase so does the resonant wavelength [8]. In fact, it can be established as a general truth for all the antenna designs that increase in the size of the resonant components results in increase or red shift in the resonant wavelength. Hence, in the reported design the resonant wavelength increases with increase in side length ‘a’.

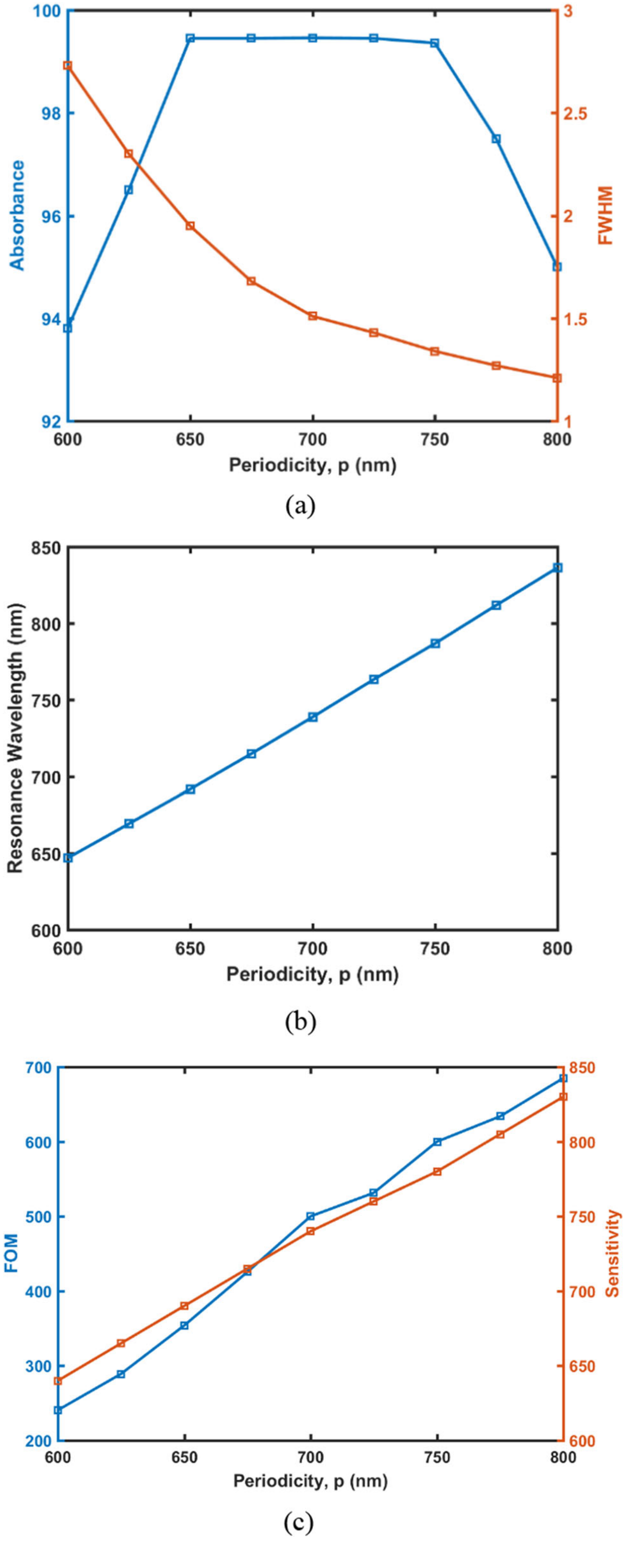
A small value of FWHM can lead to high FOM for sensing applications which can be achieved by decreasing the side lengths of pillars. But if side length is reduced below 140 nm, absorption decreases. The effect of sides of nanopillars, ‘a’ on the sensitivity and FOM of the structure can be illustrated from Fig. 4c. Sensitivity of the structure increase with increase in ‘a’ while FOM decreases with increase in ‘a’. This is because increase in value of FWHM is more than the increase in the value of sensitivity with ‘a’.

3.1.3 Effect of periodicity, ‘p’

Keeping ‘a’ and ‘h’ constant at 140 nm and 30 nm respectively, the periodicity of the structure has been varied from 600 nm to 800 nm in the step size of 50 nm, and absorbance and FWHM have been recorded. The variation of maximum absorption and FWHM with the variation of periodicity ‘p’ has been shown in Fig. 5a. It has been observed from the figure that FWHM decreases with an increase in periodicity and the absorption coefficient first increases then starts decreasing with a further increasing in the periodicity. Also, there is a linear shift in the resonance wavelength with an increase in periodicity as shown in Fig. 5b. The design is optimised at a wavelength of 692 nm to operate in the visible spectrum. The value of the sensitivity and the FOM has been calculated and plotted in Fig. 5c. It can be clearly shown from the Fig. 5c that both the sensitivity and FOM increases with increase in periodicity ‘p’. The shifting of the resonance wavelength with the variation of a period can be explained by using LC model. When the period decreased, the gap between adjacent unit cells will also decrease and then the E-filed enhancement capacitance between adjacent unit cells will increase. So, the resonance wavelength will shift towards a higher value [6, 23, 35, 36].

3.1.4 Effect of gap, ‘g’

Keeping ‘a’ and ‘h’ constant at 140 nm and 30 nm respectively and the periodicity ‘p’ of the structure at 650 nm, the gap ‘g’ between the opposite two pillars has been varied from 20 to 60 nm in the step of 10 nm. The absorbance and FWHM has been calculated and plotted in Fig. 6a. It has been observed from the figure that there is maximum absorption at the gap of 30 nm while there is no effect of gap ‘g’ on the FWHM. There is slight shift in the resonance wavelength with the gap ‘g’ as observed from Fig. 6b. The variation of the sensitivity and FOM with gap ‘g’ is shown in Fig. 6c. It can be stated that there is no effect on sensitivity of the structure with the gap between the opposite pillars. Therefore, FOM remain unaltered with the gap ‘g’ as both the sensitivity and FWHM remains constant with the gap ‘g’.



(a)

(b)

(c)

Fig. 5 Variation of **a** Absorption coefficient and FWHM, **b** resonance wavelength and **c** FOM and sensitivity with the variation of periodicity, ‘p’

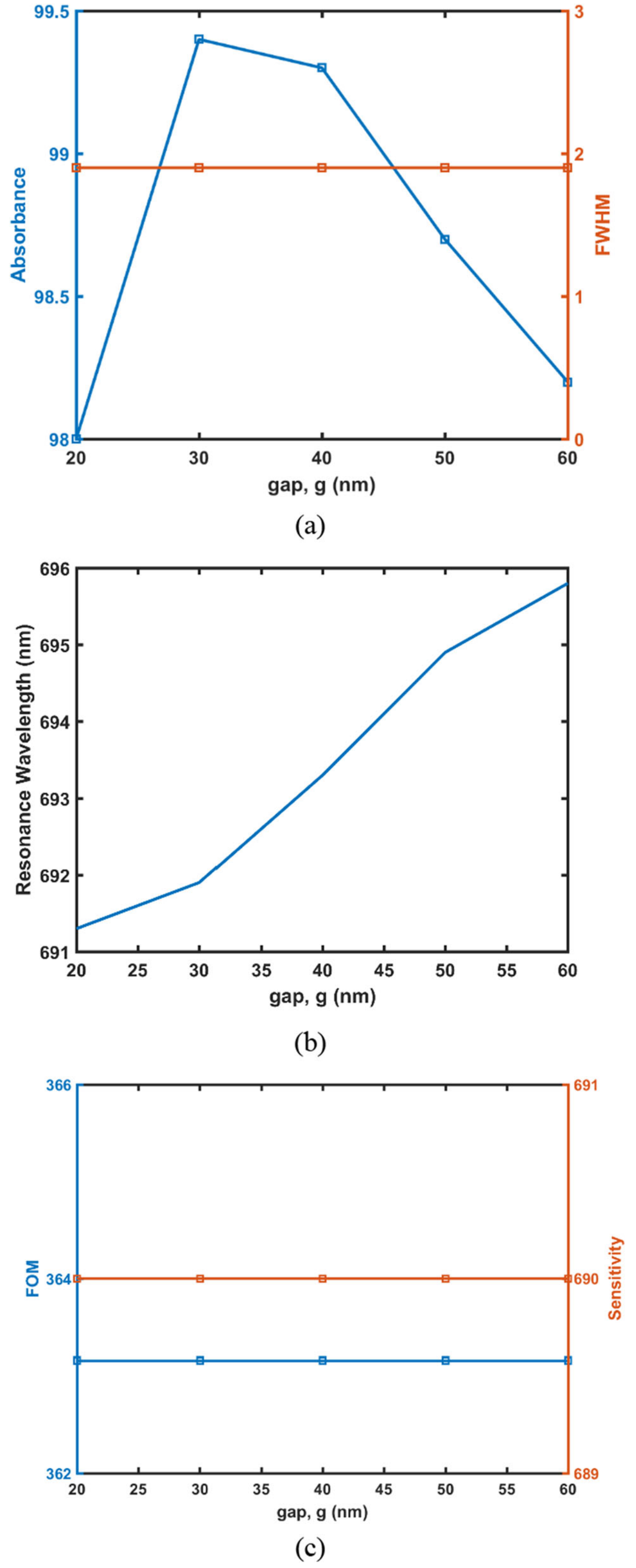


Fig. 6 Variation of **a** absorption coefficient and FWHM, **b** resonance wavelength and **c** FOM and sensitivity with the variation of gap, 'g' between opposite pillars

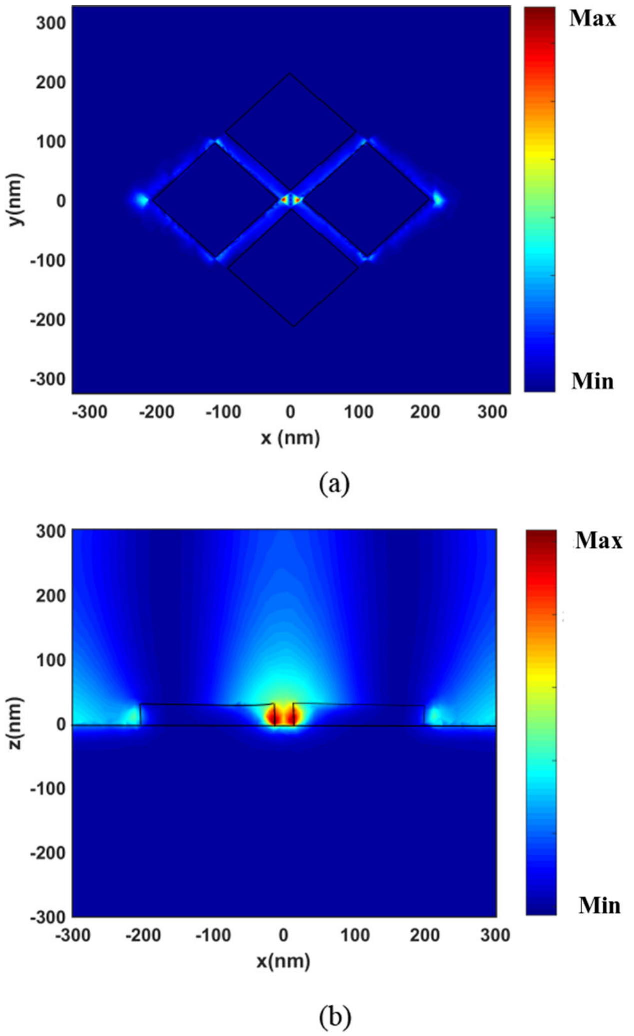


Fig. 7 **a** Distribution of electric field intensity along x–y plane, **b** distribution of magnetic field intensity along xz plane

3.2 Mechanism

To better understand the mechanism of absorption, electric and magnetic field spatial distribution plots have been drawn. Here geometrical parameters are set as a = 140 nm, h = 30 nm, g = 30 nm and p = 650 nm. The plane polarized light of wavelength 692 nm is allowed to fall on the absorber from the top. The corresponding electric field and magnetic field distribution has been drawn in the xy and xz plane respectively and shown in Fig. 7a, b. It has been observed from the figure that structure is not only a perfect absorber but also enhances the electric field intensity which is desirable in bio-sensors. The Fig. 7 shows that electric and magnetic fields are concentrated at the central gap. The electric field is observed due to the formation of localised surface plasmons. Localised surface plasmons tend to accumulate at the tip or the edges, that's why they are concentrated at the upper part more than the base of the cuboids. Electric field enhancement is more

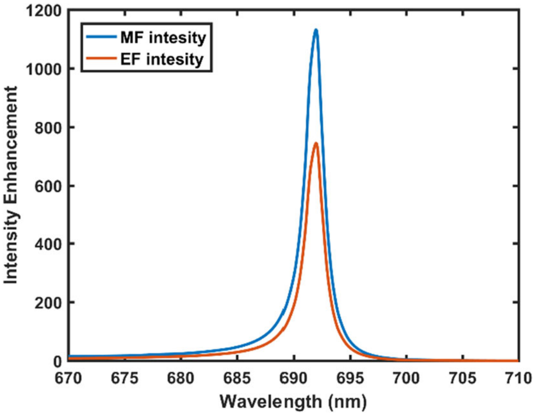


Fig. 8 Electric and magnetic field intensity enhancement at the central gap at a height of 30 nm above the silver film

at the gap rather than the other edges. Two opposite silver pillars and the connecting silver base together act as a split ring resonator. The charges accumulated at the edges flow through the conducting base, exhibiting a current carrying loop type configuration, which results in the induction of the magnetic field in the central gap. Therefore, the magnetic field is mainly concentrated at the center of the vertical gap. Figure 7a shows the electric field intensity distribution profile from the top view and Fig. 7b shows the magnetic field intensity distribution from the side view. Both electric, as well as magnetic fields, are concentrated at the same hot spot. Figure 8 shows the electric field and magnetic field enhancement for the same configuration as mentioned above at the center of the four pillars at height of 30 nm from the silver layer, where both the curves show a narrow resonance peak.

3.3 Plasmonic sensing capability

As discussed, the resonance wavelength of plasmonic structures depends upon shape, size, geometry, constituting material as well as the surrounding medium. This property of changing resonance wavelength with refractive index can be used in sensing applications. The sensors are characterised by their sensitivity and FOM. Sensitivity, S for refractive index sensor is defined by the ratio of the spectral shift with the change of refractive index of the surrounding medium, and FOM is the ratio of sensitivity to the FWHM [22–24].

$$S = \frac{\Delta\lambda}{\Delta n} \quad \text{FOM} = \frac{S}{\text{FWHM}} \quad S^* = \frac{\Delta I}{\Delta n} \quad \text{FOM}^* = \left[\frac{S^*}{I}\right]_{max}$$

Moreover, the change of intensity is also considered for sensing as the refractive index of the surrounding environment changes. S* is defined by the maximum change of intensity of reflected light with a unit change in refractive index. I is the absolute intensity of reflected light at resonance wavelength. FOM* is defined by the maximum S*/ I ratio.

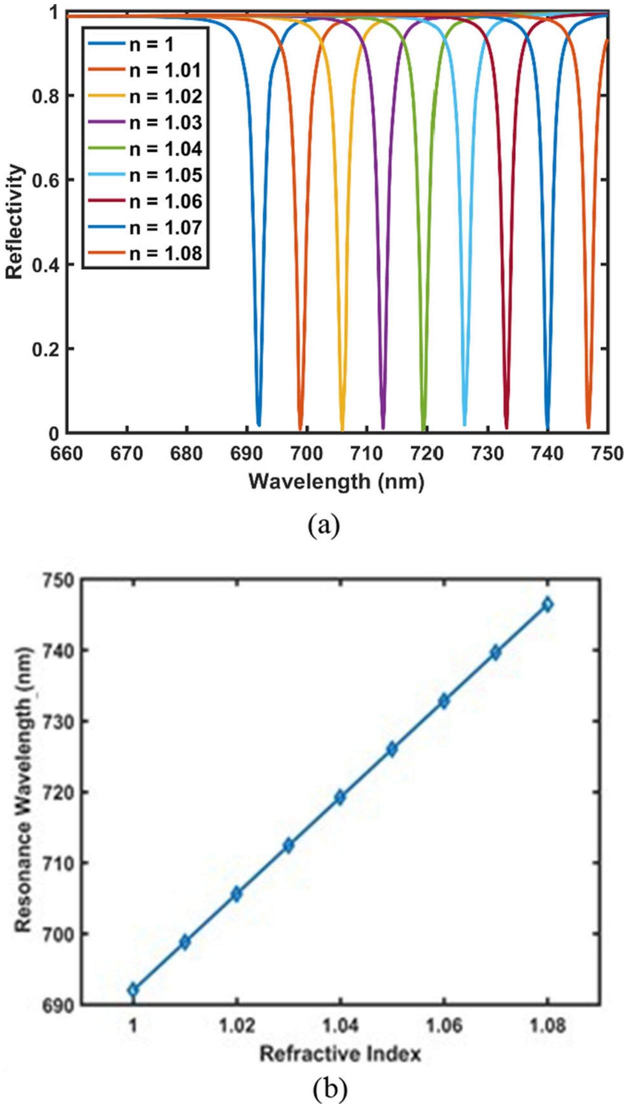


Fig. 9 **a** Reflection spectrum of the proposed all metallic absorber with the variation of wavelength for different values of refractive index varying from 1 to 1.08. **b** Shift in resonance wavelength with variation of refractive index

Air like reflective index sensors with values ranging from 1 to 1.08 are most commonly used in gas sensors [6]. The optimized design has been investigated for air-like refractive index sensor for the different values of surrounding refractive index. The refractive index of the surroundings has varied from 1 to 1.08 and corresponding absorption spectra were recorded. The reflection spectrum for different values of refractive index varying from 1 to 1.08 has been shown in Fig. 9a. The variation in resonance wavelength with the variation of the refractive index of the surrounding medium has been shown in Fig. 9b. There is a redshift in the resonance wavelength with an increase in the refracive index of the surrounding. The sensitivity and FOM come out to be 680 nm/RIU and 348.72/RIU. The value of S* and FOM* as calculated at 692 nm comes out to be 99.68 and 17,580.

4 Conclusion

Ultra narrow-band perfect absorber structure has been designed and optimised. Further, this ultra-narrowband absorber will used in refractive index sensing applications. The structure offers 99.45% absorption and has a very narrow peak at resonance wavelength. The mechanism of absorption has been revealed via observing electric and magnetic field distribution. The structure offers a high-quality factor of 354.87. The structure shows a good sensing capability with a sensitivity 680 nm/RIU and high figure of merit of 348.72/RIU.

Acknowledgements The authors gratefully acknowledge the initiatives and support provided by the establishment of the TIFAC-Centre of Relevance and Excellence in Fiber Optics and Optical Communication at Delhi Technological University (Formerly Delhi College of Engineering) Delhi, through the "Mission REACH" program of Technology Vision-2020 of the Government of India.

Author's contribution

All authors has contributed equally to the paper.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: There are no associated data available.]

References

1. A.M. Shrivastav, U. Cvelbar, I. Abdulhalim, Commun. Biol. **4**(1), 70 (2021)
2. M.N. Baitha, K. Kim, IEEE Photonics Technol. Lett. **34**(24), 1329–1332 (2022)
3. A. Dorodnyy, Y. Salamin, P. Ma, J.V. Plestina, N. Lassaline, D. Mikulik, P. Romero-Gomez, A.F. i-Morral, J. Leuthold, IEEE J. Sel. Top. Quantum Electron. **24**(6), 1–13 (2018)
4. H. Reather, *Springer Tracts in Modern Physics*, vol. 111 (1988), pp. 1–3.
5. A.A. Maradudin, J.R. Sambles, W.L. Barnes, *Modern Plasmonics* (Elsevier, Amsterdam, 2014)
6. S. Pevec, D. Donlagic, Miniature fiber-optic Fabry–Perot refractive index sensor for gas sensing with a resolution of 5x10^{-9} RIU. Opt. Express **26**(18), 23868–23882 (2018)
7. S. Verma, B. Rahman, Sensors **23**(3), 1290 (2023)
8. A.E. Krasnok, I.S. Maksymov, A.I. Denisyuk, P.A. Belov, A.E. Miroshnichenko, C.R. Simovski, Y.S. Kivshar, Phys. Usp. **56**(6), 539 (2013)
9. R. Ranga, Y. Kalra, K. Kishor, Opt. Commun. **481**, 126511 (2021)
10. P. Agarwal, K. Kishor, R.K. Sinha, Opt. Commun. **522**, 128667 (2022)
11. Y. Wang, T. Sun, T. Paudel, Y. Zhang, Z. Ren, K. Kempa, Nano Lett. **12**(1), 440–445 (2012)
12. C. Fei-Guo, T. Sun, F. Cao, Q. Liu, Z. Ren, Light Sci. Appl. **3**(4), e161 (2014)
13. Y. Li, L. Su, C. Shou, C. Yu, J. Deng, Y. Fang, Sci. Rep. **3**(1), 2865 (2013)
14. A.A. Jamali, B. Witzigmann, Plasmonics **9**, 1265–1270 (2014)
15. W. Zhou, K. Li, C. Song, P. Hao, M. Chi, M. Yu, Y. Wu, Opt. Express **23**(11), A413–A418 (2015)
16. X. Liu, T. Tyler, T. Starr, A.F. Starr, N.M. Jokerst, W.J. Padilla, Phys. Rev. Lett. **107**(4), 045901 (2011)
17. K.T. Lee, S. Seo, L.J. Guo, Adv. Opt. Mater. **3**(3), 347–352 (2015)
18. N.I. Landy, S. Sajuyigbe, J.J. Mock, D.R. Smith, W.J. Padilla, Phys. Rev. Lett. **100**(20), 207402 (2008)
19. Q. Li, Z. Li, X. Xiang, T. Wang, H. Yang, X. Wang, Y. Gong, J. Gao, Coatings **9**(6), 393 (2019)
20. J. Becker, A. Trügler, A. Jakab, U. Hohenester, C. Sönnichsen, Plasmonics **5**, 161–167 (2010)
21. C. Huang, J. Ye, S. Wang, T. Stakenborg, L. Lagae, Appl. Phys. Lett. **100**(17), 173114 (2012)
22. A.K. Agrawal, A. Ninawe, A. Dhawan, IEEE Sens. J. **22**(7), 6491–6508 (2021)
23. A.K. Agrawal, A. Suchitta, A. Dhawan, RSC Adv. **12**(2), 929–938 (2022)
24. A.K. Agrawal, A. Suchitta, A. Dhawan, IEEE Access **9**, 10136–10152 (2021)
25. Z. Yong, S. Zhang, C. Gong, S. He, Sci. Rep. **6**(1), 24063 (2016)
26. N. Liu, M. Mesch, T. Weiss, M. Hentschel, H. Giessen, Nano Lett. **10**(7), 2342–2348 (2010)
27. D. Wu, Y. Liu, R. Li, L. Chen, R. Ma, C. Liu, H. Ye, Nanoscale Res. Lett. **11**(1), 1–9 (2016)
28. S. Luo, J. Zhao, D. Zuo, X. Wang, Opt. Express **24**(9), 9288–9294 (2016)
29. S.-Y. Cho, J.L. Briscoe, I.A. Hansen, J.K. Smith, Y. Chang, I. Brener, IEEE Sens. J. **14**(5), 1399–1404 (2013)
30. E. Kazuma, T. Tatsuma, Nanoscale **6**(4), 2397–2405 (2014)
31. G. Liu, M. Yu, Z. Liu, P. Pan, X. Liu, S. Huang, Y. Wang, Plasmonics **11**, 677–682 (2016)
32. D. Wu, R. Li, Y. Liu, Z. Yu, L. Yu, L. Chen, C. Liu, R. Ma, H. Ye, Nanoscale Res. Lett. **12**(1), 1–11 (2017)
33. M. Pan, Z. Su, Z. Yu, P. Wu, H. Jile, Z. Yi, Z. Chen, Results Phys. **19**, 103415 (2020)
34. P.B. Johnson, R.-W. Christy, Phys. Rev. B **6**(12), 4370 (1972)
35. Z. Ma, F. Ding, Ultra-broadband metamaterial absorber in Terahertz regime, in *2012 Asia Communications and Photonics Conference (ACP)*, 7–10 Nov. 2012 (2012), pp. 1–3.
36. B.-X. Wang, X. Zhai, G.-Z. Wang, W.-Q. Huang, L.-L. Wang, J. Appl. Phys. **117**(1), 014504 (2015)

Using tropical cyclone characteristics and considering local factors, the radius of maximum wind over the North Indian Basin is evaluated

Monu Yadav[1] and Laxminarayan Das[1*]

[1*]Department of Applied Mathematics, Delhi Technological University, New Delhi, India.

*Corresponding author(s). E-mail(s): lndas@dtu.ac.in;
Contributing authors: yadavm012@gmail.com;

Abstract

Several coastal regions across the world suffer devastating as a result of Tropical Cyclones (TCs). In managing disasters and the economy, learning of TC wind direction, speed and radius can be quite useful. Evaluation of TC wind paramters requires a strong representation of local characteristics that include the effects of spatially varied terrain and land cover. In this study, we create an equation to determine the radius of maximum wind using TC parameters with statistical analysis. Using local regression models, the missing parameter in the best track database is reconstructed. Considering the influence of geographical factors like topography and geo-surface roughness as well. Finally, two TC cases are discussed in the paper. We find that the suggested method's error percentage with the best track data provided by the Indian Meteorological Department ranges from approximately **−50%** to **50%** and other two reference's [25, 26] error precentage with the best track data provided by the IMD, ranges from approximately **−26%** to **200%**

Keywords: Tropical Cyclone, Radius of Maximum Wind, Location of TC Centre, Estimated Pressure Drop at the Centre

1 Introduction

Tropical cyclones (TCs) in coastal regions around the world result in fatalities and significant economic damage [1]. There are approximately 90 TCs held worldwide in each year [2]. One of the main risks associated with TCs is wild wind, which not only causes property damage but also moderates the strength of storm surges and other secondary risks such as ocean waves risen. The wind speed map of the return period is commonly used to delineate and quantify the statistical distribution of storm intensity and frequency for tropical cyclonic wind danger [3]. Typically, approaches like the extreme value theory (EVT), which is based on historical meteorological records, which are used to examine the frequency of wind speeds for a single site [4]. Ground measurements for local areas may be limited, necessitating the use of historical or stochastic simulated TC modelling based on the Monte-Carlo approach [5, 6]. Methods utilising basin-wide stochastic simulations of complete TC tracks have been developed for wider regions [7, 8].

It is possible to simulate a TC wind field using numerical or parametric techniques. Although numerical models, like the Weather Research and Forecasting (WRF) model, have been extensively used in wind field reconstruction and forecasting [9, 10], parametric wind field models have become more popular in TC wind hazard assessment due to their satisfactory modelling accuracy with constrained TC parameters as inputs. Typically, historical TC track datasets can be used to derive the critical parameters, and considerable efforts have been made to assemble the world's severe TC record [2]. However, generated parameter values can occasionally have low dependability, and the accuracy of TC parameters may have an impact on the accuracy of parametric models.

Assessment of the TC wind hazard has progressed unevenly across nations and areas. Numerous TC-prone nations have developed and extensively used wind hazard models at the national level. For example, the HAZUS hurricane model has been modified since 1997 in the United States [11], and the Florida Office of Insurance Regulation has supported the creation of the Florida Public Hurricane Loss Model [12]. The Comprehensive Approach to Probabilistic Risk Assessment (CAPRA) for numerous Central American nations has been produced [13], and the Tropical Cyclone Risk Model (TCRM) for Australia was created in 2008 [14] and published in 2011 [15]. A parametric wind field model (Holland model) was used in 2011, with UNEP cooperation, to recreate the historical TC wind fields globally for all basins from 1970 to 2010 [21, 22]. Giuliani and Peduzzi [16] developed global wind speed maps based on this research, and methodologies were slightly enhanced in 2013 and 2015 [17–19].

The improvement of the assessment of the worldwide TC wind danger still has a significant gap. Generally speaking, wind dangers are geographically variable, and local elements like latitude, slope and direction aspect, land use, and land cover have a significant impact on TC wind profiles at the local scale [11, 20]. However, the majority of recent studies at the national or international level oversimplify or ignore the impact of local conditions on TC wind dangers

[7, 14, 18, 19, 21, 22]. The modelling of TC wind fields is severely hampered by the absence of measurements of critical TC characteristics such as locations of TC centre, estimated central pressure, and estimated pressure drop at the centre [2, 23, 24]. Because of this, there aren't many TC wind hazard products for the Northwest Pacific (NP), North Indian (NOI), and South Indian (SOI) basins that are publicly accessible. However, a few exclusive models have been created and released in the form of a black box [12].

Using tropical cyclone parameters including the location of the TC centre, estimated central pressure, and estimated pressure drop at the centre, this research aims to determine the value of radius of maximum wind using historical observations or reconstruction TC parameters. Considering, as well, the effects of geographical elements like topography and geo-surface roughness. In this paper,section 2, data and methodology are described. The results of the proposed method are shown in section 3 and are compared with those from the other two expression mention in paper [25, 26]. In section 4, the conclusion is provided. The section 5 indicates the study's future scope.

2 Data and Methodology

The study utilized best track dataset from the Regional Specialized Meteorological Center for Tropical Cyclones over North Indian Ocean, generally known as the Indian Meteorological Department (IMD) to evaluate the value of the suggested method. The datasets include TC metrics such as TC number, time (year, month, day, and hour), locations (longitude and latitude of TC centre), estimated central pressure (P_c), maximum wind speed (V_m), and estimated pressure drop at the centre (P_d), which are stored in the datasets at a time interval of 3 hours. Additionally, we use the radius of maximum wind data from IMD bulletins to compare with the suggested method.

The main objective of this study is to create a method for calculating the radius of maximum wind using TC characteristics, such as the latitude and longitude of the TC centre and the estimated pressure drop at the TC centre. Considering the influence of geographical factors like topography and geo-surface roughness as well.

We conclude at the relationship for determining the radius of maximum wind (r_{max})after analysing all the variables that affect it as well as the historical observation.

$$r_{max} = 10.16812 \times e^{(-0.213\sqrt{P_d}+0.0169\phi)} + 23.817 \tag{1}$$

where ϕ is latitude of the TC centre and P_d is estimated pressure drop at the TC centre.

This relationship, which determines the value of r_{max}, is valid for values of P_d that are less than or equal to 12 hPa.

We compare the suggested method to determine the r_{max} with the expression proposed by Willoughby et al.[25] (Eq. 2) , Tan and Fang [26] (Eq. 3) and the best data provided by the IMD.

$$r_{max} = 46.6 \times e^{(-0.015V_{max}+0.0169\phi)} \tag{2}$$

$$r_{max} = -26.73 \times ln(1013.25 - P_c) + 142.41 \tag{3}$$

3 Result

Two TC cases were used to evaluate the mentioned method to determining the value of r_{max}. Only two TC cases are discussed in the section 3 of the paper in order to keep it manageable. The supplementary material for the study discusses five additional TC cases. To determine whether the suggested method is accurate, we calculate the error percentage between the best data provide by the IMD and suggested method. Verify that the suggested method has a lower error percentage than the other two references by comparing the two.

$$\text{Error percentage} = \frac{\text{Experimental Value} - \text{Actual Value}}{\text{Actual Value}}$$

Extremely Severe Cyclonic Storm, Tauktae is the first TC case shown in Table 1; we find that the error percentage of the suggested method ranges from -9% to -50%, while that of the expressions of Willoughby et al. [25] and Tan and Fang [26] ranges from -3% to 66%. We get to the conclusion that the suggested method underestimates the value of r_{max}, whereas the other two expressions occasionally overestimate and occasionally underestimate the value of r_{max} in the case of Extremely Severe Cyclonic Storm, Tauktae.

Date/Time	IMD	Suggested	Willoughby et al.	Tan and Fang	E_1	E_2	E_3
14/06	40	31.8147	38.571	66.288	-20.46	-3.57	65.72
15/00	60	30.7164	31.69	60.7136	-48.80	-47.18	1.18
15/06	60	30.2973	29.65	58.30	-49.50	-50.58	-2.83
15/12	32	29.4437	25.785	53.10	-7.98	-19.42	65.93
15/18	32	29.1063	24.207	50.4	-9.04	-24.35	57.5

Table 1: Results of first TC case, where E_1 indicates the error percentage between suggested method and IMD, E_2 represents the error percentage between Willoughby et al.'s expression and IMD, and E_3 represents the error percentage between Tan and Fang's expression and IMD

Cyclonic Storm, Gulab is the second case shown in Table 2; we find that the error precentage of the suggested method ranges from -42% to 51%, while that of the expressions of Willoughby et al. [25] , and Tan and Fang [26] ranges from -26% to 202%.

After looking at the cases, we can conclude that the accuracy of the suggested method is higher than that of the other two sources.

Date/Time	IMD	Suggested	Willoughby et al.	Tan and Fang	E_1	E_2	E_3
25/00	55	32.43	40.55	71.39	-41.02	-26.27	29.8
25/06	55	32.05	40.55	71.39	-41.72	-26.27	29.8
25/12	24	31.70	37.55	65.28	32.09	56.45	174.5
25/18	24	31.70	37.55	66.28	32.09	56.45	176.16
26/00	30	31.40	34.84	63.35	4.67	16.13	111.16
26/06	24	30.89	32.38	60.71	28.71	34.91	152.95
26/12	24	30.89	32.38	60.71	28.71	34.91	152.95
26/18	21	31.71	37.62	63.35	51.02	79.14	201.66

Table 2: Results of Second TC case, where E_1 indicates the error percentage between suggested method and IMD, E_2 represents the error percentage between Willoughby et al.'s expression and IMD, and E_3 represents the error percentage between Tan and Fang's expression and IMD

4 Conclusion

To find the value of r_{max} in this study, a new expression has been developed. The equation included TC parameters including the center's position (Lat. and log.), as well as the estimated pressure drop at the center. Considering the influence of geographical factors like topography and geo-surface roughness as well. This expression may produce results that are better to those obtained using the other expression.

$$r_{max} = 10.16812 \times e^{(-0.213\sqrt{P_d}+0.0169\phi)} + 23.817$$

where ϕ is latitude of the TC centre and P_d is estimated pressure drop at the centre.

This relationship, which determines the value of r_{max}, is valid for values of P_d that are less than or equal to 12 hPa and no any condition on the latitude of the TC centre over the North Indian basin.

5 Future Scope

We investigate more TC situations and work to create an expression for estimated pressure drops at the centre that are greater than 12 hPa. To determine the value of r_{max}, try validating this expression on additional basins, such as the South Pacific and North Pacific basins with considering the influence of geographical factors like topography and surface roughness as well.

Acknowledgement

The best track data and metrics for tropical cyclone characteristics are freely available thanks to the Indian Meteorological Department, which is appreciated by the authors.

Supplementary Material

This paper's supplementary material includes five additional TC cases over the North Indian Basin, with the same notation as in the paper.

Declarations

Conflict of Interest

Conflict of interest is not declared by any of the authors.

Author Statement

Research and manuscript preparation were done equally by each author.

Data Availability

The datasets include TC metrics such as TC number, time (year, month, day, and hour), locations (longitude and latitude of TC centre), estimated central pressure (P_c), maximum wind speed (V_m), and estimated pressure drop at the centre (P_d), which are stored in the datasets at a time interval of 3 hours is freely available on the website of IMD.

References

[1] Guha-Sapir, D., Below, R., Hoyois, P.: EM-DAT. International Disaster Database (Brussels—Belgium: Université Catholique de Louvain). Accesed on January 30, 2023 (2019)

[2] Knapp, K.R., Kruk, M.C., Levinson, D.H., Diamond, H.J., Neumann, C.J.: The international best track archive for climate stewardship (ibtracs) unifying tropical cyclone data. Bulletin of the American Meteorological Society **91**(3), 363–376 (2010)

[3] Fang, W., Lin, W.: A review on typhoon wind field modeling for disaster risk assessment. Prog Geogr **32**(6), 852–867 (2013)

[4] Elsner, J.B., Jagger, T.H., Liu, K.-b.: Comparison of hurricane return levels using historical and geological records. Journal of Applied Meteorology and Climatology **47**(2), 368–374 (2008)

[5] Russell, L.R.: Probability Distributions for Texas Gulf Coast Hurricane Effects of Engineering Interest. Stanford University., ??? (1969)

[6] Huang, Z., Rosowsky, D.V., Sparks, P.R.: Long-term hurricane risk assessment and expected damage to residential structures. Reliability engineering & system safety **74**(3), 239–249 (2001)

[7] Vickery, P., Skerlj, P., Twisdale, L.: Simulation of hurricane risk in the us using empirical track model. Journal of structural engineering **126**(10), 1222–1237 (2000)

[8] Emanuel, K., Ravela, S., Vivant, E., Risi, C.: A statistical deterministic approach to hurricane risk assessment. Bulletin of the American Meteorological Society **87**(3), 299–314 (2006)

[9] Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Wang, W., Powers, J.G.: A description of the advanced research wrf version 2. Technical report, National Center For Atmospheric Research Boulder Co Mesoscale and Microscale . . . (2005)

[10] Nolan, D.S., Zhang, J.A., Stern, D.P.: Evaluation of planetary boundary layer parameterizations in tropical cyclones by comparison of in situ observations and high-resolution simulations of hurricane isabel (2003). part i: Initialization, maximum winds, and the outer-core boundary layer. Monthly weather review **137**(11), 3651–3674 (2009)

[11] Agency), F.F.E.M.: HAZUS-MH 2.1 hurricane model technical manual. FEMA Washington, DC (2012)

[12] University), F.F.I.: Florida Public Hurricane Loss Model 4.1. Miami, FL: Florida International University

[13] Cardona, O.D., Ordaz, M., Reinoso, E., Yamín, L., Barbat, A.: Capra–comprehensive approach to probabilistic risk assessment: international initiative for risk management effectiveness. In: Proceedings of the 15th World Conference on Earthquake Engineering, vol. 1 (2012)

[14] Arthur, W., Schofield, A., Cechet, R., Sanabria, L.: Return period cyclonic wind hazard in the australian region. In: 28th AMS Conference on Hurricanes and Tropical Meteorology (2008). Citeseer

[15] : Summons, C. N. & Arthur: Tropical Cyclone Risk Model User Guide. Canberra, Australia: Commonwealth of Australia (Geoscience Australia). Accessed on 26 January, 2023. https://storage.googleapis.com/google-codearchive-downloads/v2/code.google.com/t user_guide.pdf.

[16] Giuliani, G., Peduzzi, P.: The preview global risk data platform: a geoportal to serve and share global data on risk to natural hazards. Natural hazards and earth system sciences **11**(1), 53–66 (2011)

[17] CIMNE, E., INGENIAR, I.: Probabilistic modeling of natural risks at the global level: global risk model. Background paper prepared for the 2013 global assessment report on disaster risk reduction, UNISDR, Geneva,

Switzerland. UNISDR, Geneva, Switzerland. http://www. preventionweb. net/gar (2013)

[18] UNISDR, G.: Global assessment report on disaster risk reduction. United Nations International Strategy for Disaster Reduction Secretariat Geneva, Switzerland (2013)

[19] UNISDR, G.: Global assessment report on disaster risk reduction. United Nations International Strategy for Disaster Reduction Secretariat Geneva, Switzerland (2015)

[20] Lee, Y.-K., Lee, S., Kim, H.-S.: Evaluation of wind hazard over jeju island. Wind Engineering (2009)

[21] UNISDR, G.: Global assessment report on disaster risk reduction. United Nations International Strategy for Disaster Reduction Secretariat Geneva, Switzerland (2009)

[22] UNISDR, G.: Global assessment report on disaster risk reduction. United Nations International Strategy for Disaster Reduction Secretariat Geneva, Switzerland (2011)

[23] Landsea, C.W., Franklin, J.L.: Atlantic hurricane database uncertainty and presentation of a new database format. Monthly Weather Review **141**(10), 3576–3592 (2013)

[24] Ying, M., Zhang, W., Yu, H., Lu, X., Feng, J., Fan, Y., Zhu, Y., Chen, D.: An overview of the china meteorological administration tropical cyclone database. Journal of Atmospheric and Oceanic Technology **31**(2), 287–301 (2014)

[25] Willoughby, H.E., Darling, R.W.R., Rahn, M.E.: Parametric representation of the primary hurricane vortex. part ii: A new family of sectionally continuous profiles. Monthly Weather Review **134**(4), 1102–1120 (2006). https://doi.org/10.1175/MWR3106.1

[26] Tan, C., Fang, W.: Mapping the wind hazard of global tropical cyclones with parametric wind field models by considering the effects of local factors - International Journal of Disaster Risk Science. Beijing Normal University Press (2018). https://link.springer.com/article/10.1007/s13753-018-0161-1

Supplementary Material

Results of TC cases, where E_1 indicates the error percentage between suggested method and IMD, E_2 represents the error percentage between Willoughby et al.'s expression and IMD, and E_3 represents the error percentage between Tan and Fang's expression and IMD

Date/Time	IMD	Suggested	Willoughby et al.	Tan and Fang	E_1	E_2	E_3
07/00	64	31.12	34.36	71.39	-51.37	-46.31	11.54
07/06	64	31.13	34.41	69.58	-51.35	-46.23	8.71
07/12	64	30.83	34.53	69.58	-51.82	-46.04	8.71
07/18	34	30.57	32.203	67.88	-10.08	-5.28	-76.82
08/00	34	30.34	29.97	64.78	-10.76	-11.85	90.52
08/06	34	29.92	27.95	60.71	-12	-17.79	78.55
08/12	22	29.58	26.15	58.309	34.45	18.86	165.04
08/18	22	29.63	26.33	58.309	34.68	19.68	165.04
09/00	22	29.67	26.509	58.309	34.86	20.49	165.04
09/06	34	30.103	28.76	60.71	-11.46	-15.41	78.55
09/12	34	30.39	31.32	62.002	-10.61	-7.88	82.35
09/18	34	30.96	34.05	64.781	-8.94	0.14	90.53
10/00	64	31.67	36.95	69.58	-50.51	-42.26	8.71

Table 3: Severe Cyclonic Storm "MANDOUS" over Bay of Bengal during $6-10$ December, 2022

Date/Time	IMD	Suggested	Willoughby et al.	Tan and Fang	E_1	E_2	E_3
07/12	55	31.33	35.303	64.78	-43.03	-35.81	17.78
07/18	55	31.39	35.66	63.35	-42.91	-35.16	15.18
08/00	30	30.81	33.31	60.71	2.70	11.03	102.36
08/06	30	30.12	28.86	56.103	0.41	-3.8	87.001
11/00	30	30.58	30.98	56.103	1.96	3.26	87.01
11/06	30	31.11	33.51	58.309	3.707	11.7	94.36
11/12	55	31.76	36.309	60.71	-42.24	-33.98	10.38

Table 4: Severe Cyclonic storm ASANI over Bay of Bengal during $07-12$ May, 2022

Date/Time	IMD	Suggested	Willoughby et al.	Tan and Fang	E_1	E_2	E_3
23/00	64	31.99	38.48	75.43	-50.00	-39.87	17.85
23/06	64	31.71	38.87	73.33	-50.45	-39.26	14.57
23/12	34	31.49	36.55	71.39	-7.37	7.5	109.97
23/1	34	31.57	36.92	69.58	-7.14	8.58	104.64
24/00	34	31.09	34.66	66.28	-8.55	1.94	94.94
24/06	34	31.001	32.87	66.28	-8.82	-3.32	94.94
24/12	34	31.14	33.55	66.28	-8.38	-1.32	94.94
24/18	34	32.09	35.29	66.28	-5.59	3.79	94.94

Table 5: Cyclonic Storm “SITRANG” over Bay of Bengal during 22 – 25 October, 2022.

Date/Time	IMD	Suggested	Willoughby et al.	Tan and Fang	E_1	E_2	E_3
03/00	69	31.73	37.26	77.71	-54.003	-46	12.62
03/12	42	31.05	33.23	73.33	-26.06	-20.87	74.59
03/06	42	31.21	35.21	75.43	-25.68	-16.16	79.59
03/18	42	31.09	35.45	73.33	-25.95	-15.59	74.59
04/00	42	31.13	33.62	73.33	-25.86	-19.95	74.59
04/06	42	31.45	36.37	75.43	-25.11	-13.40	79.59
04/12	55	31.84	39.53	77.71	-42.09	-28.12	41.29

Table 6: Cyclonic Storm, “JAWAD” over the Bay of Bengal during 02 – 06 December, 2021

Date/Time	IMD	Suggested	Willoughby et al.	Tan and Fang	E_1	E_2	E_3
23/18	85	32.13	39.13	60.71	-62.19	-53.96	-28.57
24/00	58	31.44	36.309	58.309	-45.79	-37.39	0.53
24/06	58	31.16	33.74	26.103	-46.27	-41.82	-3.27
24/12	58	30.73	31.67	54.06	-47.00	-45.39	-6.79
24/18	38	30.36	29.63	52.17	-20.09	-22.02	37.28
25/00	38	29.69	27.68	48.74	-21.85	-27.15	28.26
25/06	38	29.46	25.98	47.18	-22.45	-31.63	24.15

Table 7: Very Severe Cyclonic Storm, “YAAS” over the Bay of Bengal during 23 – 28 May, 2021

Weight-based teasing and depressive symptoms among Indian college students: exploring the moderating effect of gratitude

Suhans Bansal and Naval Garg
USME, Delhi Technological University, Delhi, India, and
Jagvinder Singh
Department of DOR, University of Delhi, Delhi, India

Received 28 October 2022
Revised 10 January 2023
8 February 2023
Accepted 9 February 2023

Abstract

Purpose – This instant study explores the relationship between weight-based teasing and depressive symptoms in Indian college students. It further investigates the moderating effect of gratitude on depressive symptoms occurring due to weight-based teasing.

Design/methodology/approach – The study is theoretically based on Fredrickson's broaden-and-built theory (2001). PROCESS macro in IBM SPSS v21 was used to analyze the effect of gratitude in moderation of weight-based teasing and depressive symptoms. The study used correlation and regression analysis to assess the relationship between weight-based teasing and depressive symptoms.

Findings – The study has confirmed that weight-based teasing results in the development of depressive symptoms in Indian college students. The study has also revealed that gratitude casts a significant moderating effect on depression due to weight-based teasing, i.e. a reduction in regression weight of weight-based teasing.

Originality/value – This study is the first of its kind in India and will significantly add to the national literature on teasing and depression. Further, the study will help stakeholders like educators and policymakers to formulate psychological programs based on positive psychology 2.0 and gratitude to combat the rising issue of body shaming in India.

Keywords Teasing, Weight-based teasing, Depression, Gratitude, Moderation, Depressive symptoms

Paper type Research paper

Introduction

The meaning and intention of teasing can have varied interpretations (Keltner *et al.*, 2001). Teasing can be hostile or friendly (Myers *et al.*, 2013). Myers *et al.* (2013) suggest that teasing may lead to specific positive outcomes. Barnett *et al.* (2004) found that teasing affected interhuman relationships positively. However, teasing can be dangerous if done to hurt people and can significantly cause physical and emotional damage (Demirbağ *et al.*, 2017). Moreover, the studies have found the association of teasing with other mental health concerns like low self-esteem (Merlo *et al.*, 2018), social anxiety (Webb *et al.*, 2020; Demirbağ *et al.*, 2017), sense of rejection (Webb *et al.*, 2020) and depression (Szwimer *et al.*, 2020; Merlo *et al.*, 2018). Even children have associated teasing with aggression and do not enjoy the event (Szwimer *et al.*, 2020; Shapiro *et al.*, 1991). Also, teasing impacts adolescents' and children's psychological well-being, especially concerning their body type and weight (Greenleaf *et al.*, 2013; Schmidt and Martin, 2019; Blanco *et al.*, 2019). Teasing based on body weight is referred to as weight-based teasing (Greenleaf *et al.*, 2017). It can deter various adverse effects on its victims. The victims may form negative views against their physical appearance

The authors are thankful to all the anonymous reviewers for their inputs that have made this research paper more insightful. The authors also express gratitude towards the participants of this study.

Conflict of interest: The authors of the present study declare no conflict of interest.

Kybernetes
© Emerald Publishing Limited
0368-492X
DOI 10.1108/K-10-2022-1486

(Ievers-Landis *et al.*, 2019) or may see an increased rate of development in mental health issues (Greenleaf *et al.*, 2017).

Mental health issues have become more pertinent with the rising global health crisis (Wang *et al.*, 2021). Mental health issues generally start as minor symptoms, which, left untreated, can cause significant damage to an individual. One such issue is depression. Depression is not a new subject of study for scientists around the globe (Bansal *et al.*, 2022). Still, the rising cases of depression pose a significant concern to researchers and medical practitioners (Santomauro *et al.*, 2021). In 2021, the World Health Organization (WHO) reported that nearly 3.8% of the global population was affected by depression (WHO, 2021). The report further revealed that of the affected 3.8% of people, 5% were adults, and among these adults, 5.7% were over 60 (WHO, 2021). Goldschmidt *et al.* (2016) also explain that, though depressive symptoms may set on at an early age, like adolescents, they may last into adulthood. Depression also has serious repercussions on its victims. Persistent depression may cause its victims to develop other health issues like heart disease (Charlson *et al.*, 2011; Abu Sumaqa and Hayajneh, 2022). Further, the number of cases related to depression increased recently due to the COVID-19 pandemic.

With the onset of the COVID-19 pandemic, governments worldwide imposed stringent regulations like lockdowns. During the COVID-19 pandemic, the global youth depression cases doubled to 25.2% (Racine *et al.*, 2021). Ettman *et al.* (2020) suggested that depression cases rose threefold in the US during the COVID pandemic. A few reasons for this spike include a lack of social interaction and home isolation (Hosen *et al.*, 2021; Asanov *et al.*, 2021). Therefore, it has become necessary to study not only depression at the global level but at the country level also. India, with a population of over 1.2 billion people, is said to be home to 57 million people (18% of the global population with depression) affected by depression (Gandhi and Kishore, 2020). The cases further increased due to the onset of COVID 19 pandemic. The pandemic caused trauma because of lockdowns, isolations and death tolls (Banerjee, 2022). Based on the mentioned factors, the trauma led to a spike in depression cases among Indian citizens (Banerjee, 2022). Apart from COVID-19 and mental health stigma, depression in students can arise due to other factors. For instance, Giacolini *et al.* (2021) suggested that compared to the cortical, the asymmetrical development of the subcortical areas exposes an individual to neuro-hormonal and psychological dynamics of Dominance/Submission systems' interaction with Attachment/Care systems. Also, pre-existing tendencies and subclinical or personality traits can be the causes of depressive symptoms in an individual, suggested Giacolini *et al.* (2021). Some other factors include academic stress (Krukowski *et al.*, 2009), mobile addiction (Zhang *et al.*, 2022) and weight-based teasing (Webb *et al.*, 2020). Literature consists of studies on weight-based teasing, majorly based on school-going children. In India, it is somewhat easy to curb weight-based teasing in school students as they are under the constant monitoring and guidance of their parents and teachers (Bansal *et al.*, 2022). When we consider college students, they have relatively more freedom, and their increased use of information communication technology (ICT) has made them a target of the same even through social networking sites (SNS) (Verma *et al.*, 2022). Consequently, this has also resulted in the increased reporting of mental health issues among college students (Bansal *et al.*, 2022). Further, it is also suggested in Beck's (1967) cognitive theory on depression that people tend to evaluate events negatively due to their systematic bias in the thinking process (Mcleod, 2015) and their experiences foster their behaviours and vice versa. Therefore, it became imperative to study the impact of weight-based teasing on Indian college students. Hence, our study investigates the role of weight-based teasing in developing depressive symptoms in Indian college students.

In addition, with the advent of the second wave of positive psychology (positive psychology 2.0), the demand to explore the role of positive paradigms like mindfulness and gratitude in reducing the ills like depression and anxiety is rising (Garg *et al.*, 2021).

Regardless of the circumstances a person may face, positive psychology 2.0 will not only help combat the ills, as mentioned earlier but will also help people to better their lives and their potential (Wong, 2016). Therefore, the present study, which is based on the broaden-and-build theory (Fredrickson, 2001), forms the basis of this present research which studies the moderating role gratitude plays in weight-based teasing and depressive symptoms. The broaden-and-build theory (Fredrickson, 2004), based on positive psychology 2.0, postulates that positive emotions like gratitude, enjoyment and optimism can broaden an individual's awareness (Garg *et al.*, 2021). Positive emotions also build psychological resources like resilience that will help an individual to build resistance to depression (Susman, 2021) and improve the meaning of life (Fredrickson, 2004). The present study is novel because it focuses on explaining the role of gratitude in reducing the harmful consequences of weight-based teasing at the individual level. For instance, positive emotions like gratitude (Garg *et al.*, 2022) will act as a cushion against negative emotions as they will help an individual increase resilience (Garg *et al.*, 2022) against negative comments on their body type or weight. Consequently, the present study hypothesizes that a student who experiences more gratitude is less likely to develop depressive symptoms due to weight-based teasing, as it helps improve an individual's positive behaviour. To our best knowledge, the findings provided in the research are unique, and no previous research has explored the moderating role of gratitude on the relationship between weight-based teasing and depressive symptoms.

The study is majorly presented in four sections. The first section offers the theoretical substance of the relationship between weight-related teasing, depressive symptoms and gratitude, followed by a description of the proposed moderation effect of gratitude. The second part describes the sampling design, checks the reliability and validity of testing and explains the methodology used. Various hypotheses subjected to strict statistical testing are provided in the third section. Finally, the last section discusses the results and concludes the research. Also, the last section presents the theoretical and practical implications and limitations of the study and future scope for researchers.

Literature review and hypothesis development

Teasing

For ages, human beings have used teasing to communicate certain emotions effectively (Eckert *et al.*, 2020). These emotions resulted in either pro-social or dangerous behaviour (Myers *et al.*, 2013). Shapiro *et al.* (1991) defined teasing as "a personal communication directed by an agent toward a target that includes three components: aggression, humour and ambiguity" (p. 460). Winfrey (1993), in their definition, suggested that the core spirit of teasing is subject to the interpretation at the receivers' end. They further suggest that the meaning of teasing is beyond words. Hence, the definitions given by researchers suggest that teasing is a personal communication, not limited to words, are subject to the interpretation by the receiver and may be an intentional act. However, the more recent literature on teasing does not stop at defining teasing; it instead suggests that teasing can result in either pro-social or antisocial behaviours. As a pro-social behaviour, teasing can result in stronger friendships (Myers *et al.*, 2013; Barnett *et al.*, 2004). Barnett *et al.* (2004) found that teasing affected interhuman relationships positively. When one child jokingly commented on one of his friend's shoe sizes by comparing it with that of basketball legend Shaquill O'Neal, his friend appreciated the joke (Barnett *et al.*, 2004). According to them, this incident resulted in pro-social teasing. Further, the literature on sexual teasing refers to friendly teasing (also known as flirting) as a medium of communication used by couples to arouse sexual feelings between them (Meston and O'Sullivan, 2007). However, when teased with malicious intentions, the victim of dangerous teasing can suffer many psychological disorders. For instance, the world is witnessing swelling rates of obesity, especially in the Sub-African and Asian regions (Obesity, 2020).

In 2016, more than 1.9 billion people above 18 were overweight in these regions (Obesity and Overweight, 2021). This high number of obese populations drew researchers' attention, especially in children and adolescents. Studies have indicated that the frequency of teasing increased by approximately 60% in children with higher body mass indexes (Szwimer *et al.*, 2020). Studies have also suggested that weight-based teasing adversely affects adolescents, like peer rejection (Webb *et al.*, 2020). When teased based on weight, body shape, or size, an individual can develop a negative perception of self and suffer from eating disorders (Day *et al.*, 2021; Schmidt *et al.*, 2016). Goldfield *et al.* (2010) and Lattie *et al.* (2019) also talk about a rapid increase in other harmful effects like depressive symptoms an individual develops due to weight-related teasing by peers, colleagues and parents.

Depressive symptoms. Depressive symptoms differ from other psychological problems like mood swings and anxiety (WHO, 2021). Psychological problems, like anxiety, are generally short-lived compared to depression (WHO, 2021). Depressive symptoms are regarded as a crucial health indicator because they can lead to the development of various medical and psychological diseases (Smith *et al.*, 2022). Various serious medical and psychological disorders can occur due to the persistence or neglect of depressive symptoms that remain untreated (Goldfield *et al.*, 2010). According to Wang *et al.*, depressive symptoms are considered the major independent factor that can develop coronary heart disease and may trigger other cardiovascular diseases (2021). An increase in pain can also be associated with depressive symptoms, as patients suffering from depression are considered more sensitive to pain (Velly and Mohit, 2018). In severe cases, these symptoms may prompt the victim to commit suicide (Kerr *et al.*, 2017; Conejero *et al.*, 2018). Various other factors contribute towards developing depressive symptoms apart from teasing. For instance, Chaudhary *et al.* (2021) suggested that university students were more prone to mental health issues, like depression, due to rules and restrictions imposed during the COVID-19 pandemic. Their study further revealed that 28.7% of the respondents were affected by depression. Further, in India, people, especially children, find it difficult to talk about mental wellness as it remains taboo (Sethi, 2021). Hence, this also becomes a significant factor in the rise in depression cases among children (Sethi, 2021), especially in the COVID times. Nevertheless, depressive symptoms are still a cause of concern (Szwimer *et al.*, 2020). If not treated, they can adversely affect the victims.

Teasing and depressive symptoms. The incidences of weight-related teasing are both psychologically and physically wounding, especially when near and dear ones like parents are the perpetrators (Dahill *et al.*, 2021). Victims can suffer from depression (Wright, 2018) and counterproductive work behaviour (Bansal *et al.*, 2022). A study by Eisenberg *et al.* (2003) suggested that the subjects (91.7%) whose parents or peers teased developed depressive symptoms in contrast to those who were not. Goldschmidt *et al.* (2016) further suggested that because of teasing at an early age, depressive symptoms may not develop at an early age but can arise into adulthood. In another study, Eisenberg *et al.* (2006) further proposed that teasing at early ages is associated with developing depressive symptoms in young adults. In addition, many studies present contrasting views on the relationship between age and early depressive symptoms (Puhl and Latner, 2007; Szwimer *et al.*, 2020). Also, the literature reports conflicting results, like not every individual is affected with the same intensity by weight-based teasing (Szwimer *et al.*, 2020). Also, Schmidt and Martin (2019) also suggested uncertain results based on the gender of the victim and the source of weight-related teasing, like family or significant others. Further, previous studies have rarely explored the relationship between weight-based teasing and depression in college students, specifically in Indian college students. Therefore, weight-based teasing has become a serious issue and must be tackled effectively. Further, the theoretical premise of this study is based on the cognitive approach to depression. The cognitive approach is based on Beck's (1967) Cognitive Theory and suggests that depression is an outcome of systematic negative bias in the thinking process" (Mcleod, 2015). This approach further states that cognitive abnormalities are the reason behind behavioural and emotional changes. The approach also assumes that changes in the thinking process arise prior to the onset of depressive symptoms. This means that a person who has developed depressive symptoms thinks differently from normal people. Beck's (1967) theory further "considers the subjective symptoms such as a negative view of self, world and future defining features of depression" (Southam-Gerow *et al.*, 2011, p. 102). The theory's central idea is that an individual's thoughts influence their behavioural and emotional experiences. Also, the experiences of an individual can affect their thoughts (Southam-Gerow *et al.*, 2011). Further, Beck and various other social scientists have expanded the theory and suggested that when this theory is applied in therapies and research, it should be emphasized to modulate an individual's thoughts as they will alter the feelings and behaviours. Also, researchers have suggested using tools of positive psychology 2.0 to modulate an individual's thoughts (Liu *et al.*, 2020; Pössel and Smith, 2019). Accordingly, this study empirically examines a theoretical relation that predicts and explains that an individual can develop depressive symptoms due to weight-based teasing at the college level. In doing so, we assume that incidences of weight-based teasing act as negative bias in the thinking process that affects victim's emotional and behavioural experience. Hence, the following hypothesis is proposed based on the premises mentioned above.

H1. Weight-Based Teasing is positively associated with depression among college students.

Gratitude and moderation effect of gratitude

According to social scientists, gratitude is both a state and a trait (Jans-Beken *et al.*, 2019). As a state, gratitude is the ability to appraise benefit(s) received as positive outcomes and the ability to recognize that these positive outcomes stem from an external source (Jans-Beken *et al.*, 2019). Gratitude as a trait is the characteristic of an individual to be thankful and notice the positives in the creation (Jans-Beken *et al.*, 2019; Garg *et al.*, 2021). According to Wood *et al.* (2010), gratitude at the trait level is also known as a positive orientation towards life and appreciation of it. Also, McCullough *et al.* have defined gratitude as an affective trait as a "general tendency to recognize and respond with grateful emotions to the roles of other people's benevolence in the positive experiences and outcomes that one obtains" (2002, p. 112). Furthermore, Watkins (2014) recommends Ronsenberg's (1998) analysis of gratitude's complexities to understand the complex nature of grateful behaviour. Rosenberg (1998) argues that grateful behaviour has three levels: emotion, trait and mood. He further suggests that when gratitude is studied on a continuum, traits and emotions reside at the very extremities (Garg *et al.*, 2021). He also argues that gratitude has a long and subtle impact on consciousness. Even in Bhagawad Gita, Lord Shree Krishna has also said that showing grateful behaviour to the Lord can also be shown by being thankful towards nature (Upadhyaya, 2018; Prabhupada, 1997). In this way, an individual can show gratitude to the Lord and experience gratitude from others or the external stimuli in their life (Upadhyaya, 2018; Prabhupada, 1997). Even positive psychology 2.0 suggests the practice of positive behaviours like gratitude to resist and reduce the ills of stress and depression (Bansal *et al.*, 2022). Researchers like Garg and Sarkar (2020) have also suggested gratitude can be termed grateful behaviour, and both terms can be used interchangeably. Studies conducted by Bartlett and DeSteno (2006) and Wood *et al.* (2008) suggest that gratitude promotes pro-social behaviour. For instance, as per Bartlett and DeSteno's (2006) study, gratitude help cultivate incidental emotions. It may encourage an increase in the help provided even to strangers. They suggest that gratitude differs from mere awareness of the pro-social norms and helps promote pro-social behaviour. Hence, gratitude positively affects an individual's mental health and consciousness and may moderate behaviour or actions (Bartlett and DeSteno, 2006). Researchers like Garg and Sarkar (2020) have also suggested that gratitude promotes vitality in Indian college students. Thus, gratitude can be used as one of the interventions to promote positive behaviour and hinder the ill effects of events like teasing, cyberbullying (Bansal *et al.*, 2022) and depression.

Eisenberg *et al.* (2011) discuss the impact of weight-based teasing perpetrated by family or significant others and suggest methods to combat it. They have suggested measures like community-wide campaigns or interventions to reduce weight-based talks. Blanco *et al.* have further suggested introducing training programs at the school level to help educators redesign the school policies to minimize students' exposure to weight-related talks (2019). However, the need for more accurate and preventive measures at the individual level is still there. The proposed postulate of gratitude's moderation effect is based on the broaden-and-built theory by Fredrickson (2004) in positive psychology. The theory suggests that positive emotions like happiness, kindness and gratitude "broaden people's momentary thought-action repertoires and build their enduring psychological and social resources" (Fredrickson, 2004, p. 147). In her theory, she explains how grateful behaviour helps expand an individual's universe of thoughts by broadening awareness and encouraging novel actions (Garg *et al.*, 2021). This broadened behaviour helps individuals develop social and psychological resources like self-resilience, optimism, hope and confidence (Fredrickson, 2004). Social resources comprise trust, loyalty, respect and love for family, peers and others (Fredrickson, 2004; Bansal *et al.*, 2022). On the other hand, psychological resources like hope and optimism may help the perpetrators and victims deal with shame and guilt arising out of weight-based teasing. Even studies have suggested a positive correlation between good mental health and psychological resources (Jans-Beken *et al.*, 2019; Susman, 2021; Sarkar and Garg, 2020). Psychological resources like self-resilience help an individual reduce the risk of developing depressive symptoms (Wu *et al.*, 2020), whereas social resources like love and trust safeguard mental health (Fasihi Harandi *et al.*, 2017; Li *et al.*, 2021). Also, family and peer support help reduce the risk of developing depressive symptoms (Wise *et al.*, 2019). A study by Sundar *et al.* (2016) proved the importance of positive psychology interventions to promote mental well-being and reduce the ill effects of depression. The study further suggested that positive psychology can be a low-cost tool to develop values and strengths that may cushion adolescents against depression. Further, studies have shown a positive linkage between grateful behaviour and an individual's psychological capital (PsyCap) (Chen *et al.*, 2019; Fox *et al.*, 2018; Sarkar and Garg, 2020). Increased PsyCap leads to better work-related results because an individual can tackle any work or academics-related misbehaviour like bullying at the organization (Chen *et al.*, 2019; Garg, 2020). Even Fredrickson suggests that "gratitude—like joy, interest, contentment, love, pride and elevation—broadens people's modes of thinking, which in turn builds their enduring personal and social resources" (2012, p. 159). Xiang and Yuan (2020) further tested gratitude based on Fredrickson's broaden and build theory and suggested that gratitude helped increase life satisfaction in the subjects. Tachon *et al.* (2021) also suggested that gratitude moderates the relationship between daily hassles and satisfaction with life. Their study demonstrated that respondents with high gratitude scores had built resilience towards daily hassles and were not affected by them. Also, this behaviour led to higher life satisfaction scores in the respondents compared to those with lower gratitude scores (Tachon *et al.*, 2021). Similarly, Guan and Jepsen (2020) also suggested that gratitude can be used as a moderator to buffer the exhaustion employees face when they regulate others' behaviour while not doing so with them. They further suggested that gratitude provides social and cognitive resources to buffer these negative emotions. Therefore, based on the theoretical premise, the authors propose that an individual who experiences gratitude may be less likely to develop depressive symptoms arising from teasing, especially college students. In other words, the authors suggest that people with high gratitude would develop psychological resources that will help them be resilient towards negative comments on their weight. Also, a person with high gratitude will find it easy to talk and share their feelings with their near and dear ones on their experiences of weight-based teasing. These resources will help foster positive emotions in victims and hinder the development of depressive symptoms arising from weight-based teasing. Although the literature talks about gratitude as an intervention, it does not adequately consider gratitude's moderation effect on depressive symptoms, especially in collegiates. The authors propose that gratitude may moderate the depressive symptoms arising from weight-based teasing. We propose the following model and hypotheses based on the said theoretical premise.

H2. Gratitude negatively impacts the relationship between weight-based teasing and depressive symptoms

H3. Gratitude moderates the relationship between weight-based teasing and depressive symptoms

Methodological framework

Sample and data collection

The study was conducted on Indian college students with a sample of 835 students. Google Form, divided into five sections, was randomly shared via social media and educational communities like groups on Telegram and Facebook to collect data. The form was shared with 1,100 students. As per our practice, the form consists of ethical and privacy statements. For instance, the participants were ensured that the data would not be shared with any third party and would be strictly used for academic purposes. Also, informed consent from the participants was taken. The consent was taken in the form of choice. The choice of whether the participants wanted to proceed further with the survey or not (after reading the ethical and privacy statements) was provided. The form was designed so that if the participants agreed with the statements, they could proceed with the survey. If not, the survey automatically ended at that point. The second section consists of the demographic details of the respondents. There are 390 females (46.7%) and 445 males (53.3%) participants. Students enrolled in undergraduate (UG) programs are 440 (52.7%), in postgraduate (PG) 255 (30.5%) and in Doctor of Philosophy (PhD) 140 (16.8%). Additionally, students from science, commerce and humanities are 361 (43.2%), 408 (48.8%) and 66 (8.0%), respectively. The age of participants ranged between 18 and 34 years Table 1 describes the demographic details of the participants.

Variables	Category	Details
Gender	Female	390
	Male	445
Age	Below 20 years	400
	21–25 years	207
	26–30 years	167
	Above 30 years	61
Course	Under-graduate	440
	Postgraduate	225
	PhD	140
Stream	Science	361
	Commerce	408
	Humanities	66

Source(s): Primary Data

Table 1.
Demographic details

Measures

Teasing. The third section consists of a weight-based teasing scale. The scale was developed by Eisenberg *et al.* (2003) and is a five-item, five-point rating ranging from 0 (never) to 4 (at least once a week) and a unidimensional scale. It was used to capture the victim's teasing-based feelings. A few statements are “You are teased about your weight” and “People act as if they're better than you”.

Gratitude. In the fourth section, McCullough *et al.*'s (2002) gratitude questionnaire-6 (GQ-6) was used to assess the gratitude of the respondents. It is a six-item scale with seven-point ranging from 1 (strongly disagree) to 7 (strongly agree) and unidimensional (Bansal *et al.*, 2022; Garg *et al.*, 2021). A few statements of GQ-6 are “I have so much in life to be thankful for” and “when I look at the world, I don't see much to be grateful for.” As directed by McCullough *et al.* (2002), statements “When I look at the world, I don't see much to be grateful for” and “Long amounts of time can go by before I feel grateful to something or someone” were reverse coded.

Depression. The last section explored depressive symptoms in Indian college students. For this purpose, we used Kroenke *et al.*'s (2001) Patient Health Questionnaire (PHQ-9) scale. The scale is a nine-item, four-point rating ranging from 1 (not at all) to 4 (nearly every day) and a unidimensional scale. A few scale items are “little interest or pleasure in doing things” and “poor appetite or overeating”.

Reliability and validity and common method variance

Internal consistency reliability was assessed using Cronbach's alpha (CA) and composite reliability (CR) values. According to George and Mallery (2003), the accepted values of Cronbach's alpha and composite reliability are 0.70. Cronbach's alpha values of weight-based teasing, depression and gratitude are 0.82, 0.89 and 0.84, respectively. The composite reliability values of teasing, depression and gratitude are 0.88, 0.91 and 0.94, respectively. These values indicated and confirmed the internal consistency reliability of every variable used in the study. Further, average variance explained (AVE) was used to assess and verify the convergent validity. According to Fornell and Larcker (1981), the acceptable value of AVE should be equal to and above 0.50. The AVE values of teasing, depression and gratitude are 0.59, 0.53 and 0.72, respectively. Hence, it confirms the convergent validity of the variables.

Additionally, the data related to independent (weight-based teasing), moderating (gratitude) and dependent (depression) variables were collected simultaneously, so the problem of “common method variance (CMV)” was explored. We resorted to statistical and procedural remedies to address such a problem. Procedural remedies included anonymity and confidentiality of the participants, which reduced CMV risk (Podsakoff *et al.*, 2012). Therefore, personal information like the name was not collected. Furthermore, participants were assured of the confidentiality and non-shareability of the data with any third party. Also, participants were told that teasing and depression are two separate studies as a cover story. After the survey, the truth was revealed to them and the reason behind the cover stories, as Podsakoff *et al.* (2012) suggested. Lastly, as a statistical measure, Harman's (1960) one-factor method was used to assess CMV among the variable. The factor analysis of all the variables did not result in a single factor, according to Chang *et al.* (2010). Hence, the possibility of CMV was refuted. The most significant factor explained the 29.86% variance. The descriptive statistics, validity and reliability are provided in Table 2.

Analysis of data

Correlation and hierarchical regression analysis was used to explore the relationship between weight-based teasing and depressive symptoms. While exploring weight-based teasing's influence on depressive symptoms, gender and age were controlled. Finally, the moderation effect of gratitude on the relationship between teasing and depression was assessed using PROCESS Macro by Hayes (2017) in IBM SPSS v21. Mean values of the measures were considered for this research.

Results

The correlation between weight-based teasing and depressive symptoms is 0.714 ($p < 0.05$). Hence, these results suggest that weight-based teasing correlates with depressive symptoms positively, confirming that individuals who are teased based on their weight are also likely to develop depressive symptoms. Also, the studies conducted by Madowitz *et al.* (2012) and Greenleaf *et al.* (2013) confirm that teasing (in various forms) leads to psychological disorders like stress and depression. Our study is in line with these findings. Three staged hierarchical regression was used to conduct a regression analysis, whose results are presented in Table 3, and summaries of the three models in Table 4. The only control variables, age and gender, were considered in the first stage. It was found that those control variables were insignificant depressive symptom predictors. In the second model, the independent variable, i.e. weight-based teasing, was considered. It proved that although age and gender were insignificant predictors as control variables, weight-based teasing was still a significant predictor with a p-value less than 0.05. The results also proved that in Indian college students, age and gender are insignificant variables that can control the relationship between weight-based teasing and depressive symptoms. The correlation between the different variables of the study is

	Descriptive statistics					Correlation matrix		
Variables	Mean	SD	CA	CR	AVE	1	2	3
Teasing	12.97	2.71	0.82	0.88	0.59	1		
Gratitude	27.21	4.02	0.84	0.94	0.72	−0.612*	1	
Dep. Symp	18.02	3.13	0.89	0.91	0.53	0.714*	−0.550*	1

Source(s): Primary data, Teasing- Weight-Based Teasing, Dep. Symp- Depressive Symptoms, CA- Cronbach'alpha, CR- Composite reliability and AVE- Average Variance Explained, Correlation Matrix, * Sig at 0.05

Table 2. Descriptive statistics and correlation matrix

		Unstandardized coefficient		Standardized coefficient		
Model	Variable	B	Std. error	Beta	*t*-value	Sig
1	Constant	1.478	0.079		23.303	0.000*
	Gender	0.540	0.433	0.043	1.248	0.213
	Age	0.598	0.348	0.060	1.720	0.086
2	Constant	1.079	0.086		12.073	0.000*
	Gender	0.429	0.382	0.034	1.123	0.262
	Age	0.085	0.309	0.009	0.276	0.783
	Teasing	0.626	0.041	0.472	15.358	0.001*
3	Constant	0.601	0.130		4.623	0.000*
	Gender	0.253	0.042	0.038	1.270	0.205
	Age	0.194	0.034	0.010	0.323	0.747
	Teasing	0.340	0.022	0.461	15.143	0.000*
	Gratitude	−0.491	0.020	−0.511	−4.506	0.000*

Source(s): Primary data, Teasing- Weight-Based Teasing * Sig. at 0.05

Table 3. Results of hierarchical multiple regression (DV- Depression)

reported in Table 2. Therefore, the study concludes: weight-based teasing can be associated as one of the reasons behind the development of depressive symptoms in Indian college students.

Further, gratitude as a control variable was introduced in the third model. The study explored that gratitude negatively correlates with weight-based teasing (r = −0.612) and depressive symptoms (r = −0.550), which means individuals with highly grateful behaviour or those who experience or practice gratitude in some way or another are resilient to the ill effects of teasing and depression. The third model suggested a negative impact of gratitude on the relationship between weight-based teasing and depressive symptoms. The standardized regression coefficient for teasing is 0.461 ($p < 0.05$), and for gratitude is −0.511 ($p < 0.05$), which accepts hypotheses 1 and 2, respectively. The comparative assessment of the three models confirmed the improvement in the coefficient of determination from 0.110 to 0.475 and 0.110 to 0.494, implying that control variables explain only 11% of the variation in the dependent variable. Factors like workplace bullying, peer teasing, medicine like isotretinoin, or the loss of some significant other explain the rest, 36.5 and 38.3% (see Figure 1).

Moderation analysis using the PROCESS macro

Moderation analysis was further conducted to confirm the results of Model 3 (gratitude's intervention) of hierarchical regression analysis. Table 5 presents the moderation analysis. The explored moderation effect of gratitude on the relationship between weight-based teasing and depressive symptoms observes a reduction in the regression weights of teasing from 0.741 to −0.100 when interacting with gratitude. The results were in line with those of Model 3. The results indicate that gratitude helps in reducing the effect of weight-based teasing on depressive symptoms in Indian collegiates. In layperson's terms, a collegiate with high gratitude score is less likely to be impacted by weight-based teasing. Also, the F-value indicate a significant moderation effect of gratitude. Further, Figure 2 shows the gratitude moderation effect diagram. It shows that an individual with a high gratitude score

Table 4.
Model summary

				Change statistics		ANOVA	
Model	R	R Square	Adjusted R square	F changes	Sig F changes	F	Sig
1	0.110	0.006	0.004	2.465	0.043*	2.465	0.086
2	0.475	0.226	0.223	235.861	0.000*	80.728	0.000*
3	0.494	0.244	0.240	20.301	0.000*	66.997	0.000*

Source(s): Primary data, * Sig. at 0.05

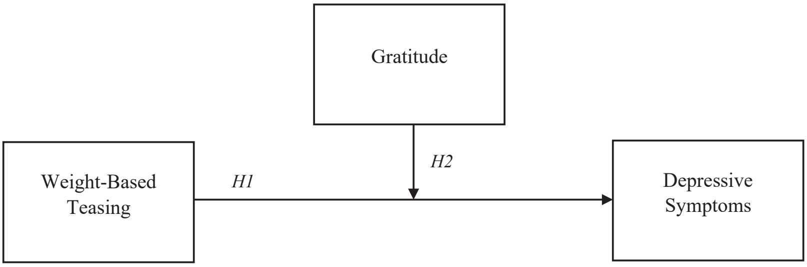


Figure 1.
Proposed moderation model

Source(s): Authors' creation based on Model 1 given by Prof. Hayes (represented by a line with red triangles with gratitude at 6 in Figure 2) can cope with and resist weight-based teasing-induced depressive symptoms. Hence, these findings and those explored using correlation and hierarchical regression empirically support the third hypothesis.

Discussion

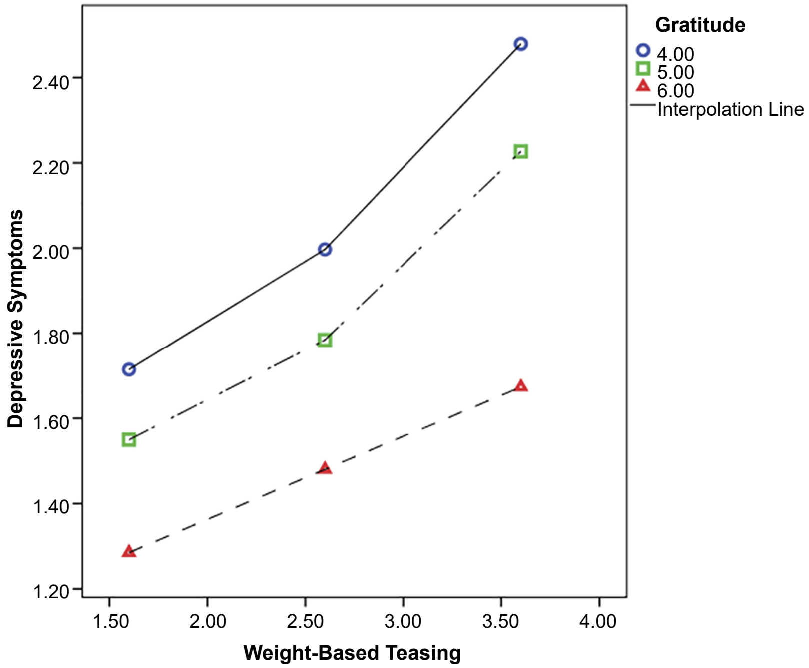
In the preceding sections, it has been clearly shown with appropriate references that teasing (specifically weight-based teasing) has played a crucial role in communicating emotions, albeit with positive and negative consequences (Myers *et al.*, 2013). Keltner *et al.* (2001) observe that teasing has various meanings, depending on the scenario and the

Model	Coefficient	S. E	*t*-value	*p*	LLCI	ULCI
Constant	0.675	0.284	0.288	0.041	0.423	0.682
Weight-Based Teasing	0.741	0.035	5.120	0.000	0.361	0.756
Gratitude	−0.301	0.084	−3.610	0.003	−0.711	−0.103
Interaction (T*G)	−0.100	0.006	−2.220	0.026	−0.248	−0.005

Model	R^2-Change	F-value	*p*
Interaction (T*G)	0.214	4.903	0.026*

Source(s): Primary Data, G-Gratitude, T- Weight-Based Teasing, * Sig at 0.05

Table 5. Results of moderation analysis (teasing - depression)



Source(s): Primary Data

Figure 2. Moderation effect

receiver's interpretations (Winfrey, 1993). Barnett *et al.* (2004) propose that when teasing is done in a friendly manner or for fun, it promotes pro-social behaviours without the intent to hurt the receiver's emotions. These pro-social behaviours can lead to stronger peer friendships or promote positive intimate relations between partners (Meston and O'Sullivan, 2007; Burnett Heyes, 2020). However, when an individual is teased for a long or when the receiver's emotions get hurt, then teasing is said to be dangerous teasing (Payne *et al.*, 1999). Blanco *et al.* (2019) have suggested that teasing can gravely impact mental well-being based on weight or body type when a child is teased. This impact on mental health can make teenagers susceptible to drug abuse (News Bureau, 2020) and other psychological problems like social anxiety (Webb *et al.*, 2020) and depression (Szwimer *et al.*, 2020; Merlo *et al.*, 2018). Also, the present literature has suggested various interventions like campaigns and workshops to create awareness on reducing weight-based teasing to deal with depression due to weight-based teasing (Eisenberg *et al.*, 2003; Gredler and Olweus, 2003). It does not adequately study the impact of weight-related teasing and tools of positive psychology 2.0 to reduce the ills of such teasing at the individual level, especially among college students. This inadequacy prompted the current study, which adequately explores the effect of weight-based teasing on Indian college students and suggests that gratitude moderates the relationship between weight-based teasing and depressive symptoms.

The study proposes that collegiates may experience depressive symptoms due to weight-based teasing. The study found a significant negative correlation between gratitude and weight-based teasing and gratitude and depressive symptoms. Initially, the study deployed correlation and hierarchical regression analysis to substantiate a relationship between weight-based teasing, depressive symptoms and gratitude. The correlation results presented in Table 1 proved that weight-based teasing significantly positively correlates with depressive symptoms. The hierarchical regression results of Model 1 and Model 2 (Table 3) also suggest that irrespective of victims' age and gender, weight-based teasing can lead to the development of depressive symptoms. In other words, weight-based teasing is one of the reasons behind increased depression-linked cases in Indian college students. Other factors can be cyberbullying (Bansal *et al.*, 2022), techno-ethical stress (Verma *et al.*, 2022), (Anju *et al.*, 2021) or even Brain Emotional Systems (Giacolini *et al.*, 2021). Hence, the present study further expanded upon Beck's (1967) Cognitive Theory on depression.

The results of hierarchical regression (Model 3) and moderation analysis presented in Tables 3 and 5 and Figure 2 suggest that gratitude moderates the relationship between weight-based teasing and depressive symptoms. When weight-based teasing interacted with gratitude, the total effect of weight-based teasing on depressive symptoms reduced from 0.741 to −0.100. In other words, it can be said that students with higher grateful behaviour are less likely to get affected by weight-based teasing and are at a lower risk of developing depressive symptoms. All the results follow Fredrickson's (2004) analysis that gratitude, like any other positive emotion, builds an individual's social resources and follows Fredrickson's broaden and build theory (2004). Fredrickson further suggested that gratitude broadens individuals' modes of thinking and helps in developing their "enduring personal and social resources" (2012, p. 159). In our case, the authors also suggest that gratitude will help students broaden their minds, making them resilient towards negative emotions like weight-based teasing. Further, when we talk about Lord Shree Krishna's message in Shree Bhagavad Gita when he suggests praising nature as one of the ways to show gratitude towards the almighty, in many ways, the Lord meant to broaden the perspective (Upadhyaya, 2018; Prabhupada, 1997). The broadened perspective will help people see negative emotions or events as part and parcel of life, which will make people resilient towards the negativity (like developing depressive symptoms) arising out of those bad events. Also, students are considered the torch bearer of the future. Hence, the study suggests that students with a positive outlook toward their surroundings, including people, will positively impact the future (Yoshida, 2022; Garg and Sarkar, 2020). Such individuals will be less impacted by teasing or other incidents. Hence, the world may see reduced revenge crime and mental issues arising from crimes.

Conclusion

To summarise, the study explored weight-based teasing and its impact on college students' psychological well-being. The psychological impact was studied by assessing depressive symptoms. The study suggested a significant positive correlation between the two. Finally, the study recommended practising gratitude to reduce the impact of weight-based teasing-related depressive symptoms by assessing gratitude's moderation effect on the relationship between weight-based teasing and depressive symptoms.

Implications

Theoretical

The present study has contributed to the literature's theoretical development in the following ways.

One of our study's contributions is improving the present understanding of the relationship between weight-based teasing and its effects on the psyche of individuals (Schvey *et al.*, 2017). Literature has reported non-satisfactory results based on the age and gender of victims. In contrast, this study has proven that the gender and age of victims are insignificant to relating episodes of weight-based with the development of depressive symptoms, especially in the Indian context. The study also expands Beck's (1967) Cognitive Theory on Depression. The study contributes an empirically tested relationship between weight-based teasing and depressive symptoms. Though the literature on the said theory contains many studies, they are inadequate, especially from weight-based teasing's point of view. This study expands the literature by empirically testing the relationship between weight-based teasing and depressive symptoms in the Indian context. The results of this study suggest that weight-based teasing acts as one of the sources of developing depressive symptoms in college students. In other words, the negative experiences and emotions arising out of weight-based teasing create a systematic bias in the thinking process that leads to the development of mental issues like depressive symptoms in college students. As the study assesses this relationship among college students, it fulfils a long pending literature gap. Earlier studies have assessed the said relationship in children and its associated impact in adulthood (Costello *et al.*, 2003).

Additionally, the study has contributed to improving the present understanding of the moderating effect of gratitude as an individual level of intervention to tackle the rising cases of depressive symptoms in the younger generation, especially collegiates. Previous studies have suggested a community-wide approach, like reducing stigmatization of weight-related teasing, which has played a role in tackling depressive symptoms (Meghani and Harvey, 2016). In contrast, this study expands the literature on Fredrickson's (2004) broaden and build theory. The study suggests that positive emotions like gratitude can help an individual build hope, optimism and resilience that can cushion the impact of negative emotions like teasing and depressive symptoms. The literature also lacks adequate gratitude-related studies on the organizational setup. The authors explore the significance of practising and experiencing gratitude in the educational setup to address this gap through this study. Finally, the literature does not adequately explore the moderation effect of gratitude. Hence, this pioneering study assesses the moderating role of gratitude in the relationship between weight-based teasing and depressive symptoms.

Practical

The results of this study will help society, academicians and policymakers in the following ways.

The results from this study will help create awareness about the negative impact of weight-based teasing on the victims. They further help policymakers and academicians create effective policies and programs to sensitize people on such issues. As weight-based teasing leads to depressive symptoms among college students, the present study will also help academicians conduct seminars and workshops on sensitizing students about their body type and the ill effects of weight-based teasing on them. Once the students are sensitized about the body types, teasers can introspect their deeds, and their thoughts about body types may change. From this, we can see a lower rate of weight-based teasings. Also, the victims of weight-based teasing may feel a shift in their perception of self (Porter *et al.*, 2012), and they might start to take negative comments as small jokes.

Also, in the Indian context, where people find it taboo to talk about mental health, awareness program about depressive symptoms and tips to combat it can be introduced to create awareness in the citizens by the Health Ministry. These programs will not only help society to understand depression as a mental disease having serious repercussions but also broaden the perspective of society not to find it taboo when individual talks about their mental health. For instance, these awareness programs can help people understand that depressive symptoms can develop due to many factors, and in some instances, they are curable too. Also, it will encourage victims of depressive symptoms to seek help and improve their lives. People will be free from the fear of getting stigmatized and talk more freely about their mental health (Schvey *et al.*, 2017).

Further practising and experiencing gratitude can help to combat the situation, as suggested by the results. A curriculum on the same can be introduced at the university level. Like Delhi Technological University, India, a few other universities have also introduced courses related to gratitude at the undergraduate level and strive to promote grateful behaviour in the institution. Researchers like Garg (2020), and Watkins (2014), recommend numerous interventions related to gratitude, like “count your blessings and ‘self-introspection’”. Garg (2019) further explains that the positive attitude of a teacher is contagious, and such an attitude inspires students during their life. He further suggests that a teacher should practice grateful behaviour. Thus, this study suggests cultivating an environment that promotes grateful behaviour in the organization. Hence, gratitude will broaden the thinking perspective of individuals and foster positive emotions. For instance, positive emotions like joy, happiness, hope and confidence will cushion negative comments on weight and see the bigger picture. The victims with broadened thinking will be able to judge and filter out weight-related comments and instead put those negative comments or emotions in a broader prospect. Positive emotions will help them give a better meaning to life and see their current situation, not as their destiny. A person with broadened thinking will take negative comments as a source of motivation to lose weight (if they want to).

Limitations and future recommendations

The study presents exciting findings, but it too has certain limitations. Firstly, the sample is based on college students. Hence, the validation of results across generations needs to be conducted. Secondly, the study uses cross-sectional data. Though the problem of CMV was addressed adequately, we suggest researchers use longitudinal research design in the future. Thirdly, the scales used were unidimensional; in the future, researchers may study other dimensions of the variables, like hope or appreciation, to explore gratitude and its effects. Future research can further probe the areas like eve-teasing to add to the literature on teasing and depression. Fourthly, the use of self-administered psychometric tools may lead to some biases, like an individual with pre-existing depressive tendency, even subclinical or as a personality characteristic, may be more likely to interpret a simple joke as teasing. Though the definition of teasing, as mentioned in the literature review, suggests the interpretation of an act or verbal communication as teasing a personal perception solely, future researchers are suggested to consider this limitation of self-administered psychometric tools and perform the research adequately. Lastly, although India is a diverse country, the study did not explore the impact of culture and its practices on gratitude, teasing and depression. Hence, future studies can explore these aspects.

References

Abu Sumaqa, Y. and Hayajneh, F.A. (2022), "Depression and anxiety in patients with heart failure: contributing factors, consequences and coping mechanisms: a review of the literature", *Working With Older People*, Vol. 26 No. 3, pp. 174-186, doi: 10.1108/wwop-12-2021-0061.

Anju, A., Amandeep, A., Punia, B., Punia, V. and Garg, N. (2021), "Life dissatisfaction among students: exploring the role of intrapersonal conflict, insufficient efforts and academic stress", *Rajagiri Management Journal*, Vol. 15 No. 2, pp. 113-128, doi: 10.1108/ramj-09-2020-0058.

Asanov, I., Flores, F., McKenzie, D., Mensmann, M. and Schulte, M. (2021), "Remote-learning, time-use, and mental health of Ecuadorian high-school students during the COVID-19 quarantine", *World Development*, Vol. 138, 105225, doi: 10.1016/j.worlddev.2020.105225.

Banerjee, P. (2022), "Anxiety, depression, panic attacks: how covid took mental toll on Indians", The Times of India, 13 June 2022, available at: https://timesofindia.indiatimes.com/india/anxiety-depression-panic-attacks-how-covid-took-mental-toll-on-indians/articleshow/90492430.cms

Bansal, S., Garg, N. and Singh, J. (2022), "Exploring the psychometric properties of the Cyberbullying Attitude Scale (CBAS) and its relation with teasing and gratitude in Indian collegiates", *International Journal of Educational Management*, Vol. 37 No. 1, pp. 225-239, doi: 10.1108/ijem-05-2022-0198.

Barnett, M.A., Burns, S.R., Sanborn, F.W., Bartel, J.S. and Wilds, S.J. (2004), "Antisocial and pro-social teasing among children: perceptions and individual differences", *Social Development*, Vol. 13 No. 2, pp. 292-310, doi: 10.1111/j.1467-9507.2004.000268.x.

Bartlett, M.Y. and DeSteno, D. (2006), "Gratitude and pro-social behavior", *Psychological Science*, Vol. 17 No. 4, pp. 319-325, doi: 10.1111/j.1467-9280.2006.01705.x.

Beck, A.T. (1967), *Depression: Clinical, Experimental, and Theoretical Aspects*, Harper & Row, PA.

Blanco, M., Solano, S., Alcántara, A.I., Parks, M., Román, F.J. and Sepúlveda, A.R. (2019), "Psychological well-being and weight-related teasing in childhood obesity: a case–control study", *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity*, Vol. 25 No. 3, pp. 751-759, doi: 10.1007/s40519-019-00683-y.

Burnett Heyes, S. (2020), "Just banter? Friendship, teasing and experimental aggression in adolescent peer networks", *Developmental Science*, Vol. 23 No. 3, doi: 10.1111/desc.12926.

Chang, S.J., van Witteloostuijn, A. and Eden, L. (2010), "From the editors: common method variance in international business research", *Journal of International Business Studies*, Vol. 41 No. 2, pp. 178-184, doi: 10.1057/jibs.2009.88.

Charlson, F.J., Stapelberg, N.J., Baxter, A.J. and Whiteford, H.A. (2011), "Should global burden of disease estimates include depression as a risk factor for coronary heart disease?", *BMC Medicine*, Vol. 9 No. 1, doi: 10.1186/1741-7015-9-47.

Chaudhary, A.P., Sonar, N.S., Jamuna, T.R., Banerjee, M. and Yadav, S. (2021), "Impact of the COVID-19 pandemic on the mental health of college students in India: cross-sectional Web-Based study", *JMIRx Med*, Vol. 2 No. 3, e28158, doi: 10.2196/28158.

Chen, X., Zeng, G., Chang, E.C. and Cheung, H.Y. (2019), "What are the potential predictors of psychological capital for Chinese primary school teachers?", *Frontiers in Education*, Vol. 4, doi: 10.3389/feduc.2019.00050.

Conejero, I., Olié, E., Calati, R., Ducasse, D. and Courtet, P. (2018), "Psychological pain, depression, and suicide: recent evidences and future directions", *Current Psychiatry Reports*, Vol. 20 No. 5, doi: 10.1007/s11920-018-0893-z.

Costello, E.J., Mustillo, S., Erkanli, A., Keeler, G. and Angold, A. (2003), "Prevalence and development of psychiatric disorders in childhood and adolescence", *Archives of General Psychiatry*, Vol. 60 No. 8, p. 837, doi: 10.1001/archpsyc.60.8.837.

Dahill, L., Mitchison, D., Morrison, N.M.V., Touyz, S., Bussey, K., Trompeter, N., Lonergan, A. and Hay, P. (2021), "Prevalence of parental comments on weight/shape/eating amongst sons and daughters in an adolescent sample", *Nutrients*, Vol. 13 No. 1, p. 158, doi: 10.3390/nu13010158.

Day, S., Bussey, K., Trompeter, N. and Mitchison, D. (2021), "The impact of teasing and bullying victimization on disordered eating and body image disturbance among adolescents: a systematic review", *Trauma, Violence, and Abuse*, Vol. 23 No. 3, pp. 985-1006, 152483802098553, doi: 10.1177/1524838020985534.

Demirbağ, B.C., Çiçek, Z., Yiğitbaş, A., Özkan, I.G. and Dinçer, A. (2017), "The relationship between types of bullying experienced by primary school students and their anxiety, State-Trait, Self-Esteem and certain Socio-Demographic characteristics", *Procedia - Social and Behavioral Sciences*, Vol. 237, pp. 398-404, doi: 10.1016/j.sbspro.2017.02.077.

Eckert, J., Winkler, S.L. and Cartmill, E.A. (2020), "Just kidding: the evolutionary roots of playful teasing", *Biology Letters*, Vol. 16 No. 9, 20200370, doi: 10.1098/rsbl.2020.0370.

Eisenberg, M.E., Neumark-Sztainer, D. and Story, M. (2003), "Associations of Weight-Based teasing and emotional well-being among adolescents", *Archives of Pediatrics and Adolescent Medicine*, Vol. 157 No. 8, doi: 10.1001/archpedi.157.8.733.

Eisenberg, M., Neumarksztainer, D., Haines, J. and Wall, M. (2006), "Weight-teasing and emotional well-being in adolescents: longitudinal findings from project EAT", *Journal of Adolescent Health*, Vol. 38 No. 6, pp. 675-683, doi: 10.1016/j.jadohealth.2005.07.002.

Eisenberg, M.E., Berge, J.M., Fulkerson, J.A. and Neumark-Sztainer, D. (2011), "Associations between hurtful weight-related comments by family and significant other and the development of disordered eating behaviors in young adults", *Journal of Behavioral Medicine*, Vol. 35 No. 5, pp. 500-508, doi: 10.1007/s10865-011-9378-9.

Ettman, C.K., Abdalla, S.M., Cohen, G.H., Sampson, L., Vivier, P.M. and Galea, S. (2020), "Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic", *JAMA Network Open*, Vol. 3 No. 9, e2019686, doi: 10.1001/jamanetworkopen.2020.19686.

Fasihi Harandi, T., Mohammad Taghinasab, M. and Dehghan Nayeri, T. (2017), "The correlation of social support with mental health: a meta-analysis", *Electronic Physician*, Vol. 9 No. 9, pp. 5212-5222, doi: 10.19082/5212.

Fornell, C. and Larcker, D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18 No. 1, p. 39, doi: 10.2307/3151312.

Fox, C., Webster, B.C. and Casper, W.C. (2018), "Spirituality, psychological capital and employee performance: an empirical examination", *Journal of Managerial Issues*, Vol. 30 No. 2, p. 194, available at: https://www.questia.com/library/journal/1G1-545892001/spirituality-psychological-capital-and-employee-performance

Fredrickson, B.L. (2001), "The role of positive emotions in positive psychology: the broaden-and-build theory of positive emotions", *American Psychologist*, Vol. 56 No. 3, pp. 218-226, doi: 10.1037/0003-066x.56.3.218.

Fredrickson, B.L. (2004), "The broaden–and–build theory of positive emotions", *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, Vol. 359 No. 1449, pp. 1367-1377, doi: 10.1098/rstb.2004.1512.

Fredrickson, B.L. (2012), "Gratitude, like other positive emotions, broadens and builds", in Emmons, R.A. and McCullough, M. (Eds.), *The Psychology of Gratitude*, online, pp. 145-166, Oxford Academic, doi: 10.1093/acprof:oso/9780195150100.001.0001, Original work published 2004.

Gandhi, P.A. and Kishore, J. (2020), "Prevalence of depression and the associated factors among the software professionals in Delhi: a cross-sectional study", *Indian Journal of Public Health*, Vol. 64 No. 4, doi: 10.4103/ijph.IJPH_568_19.

Garg, N. (2019), "High performance work practices and organizational performance-mediation analysis of explanatory theories", *International Journal of Productivity and Performance Management*, Vol. 68 No. 4, pp. 797-816, doi: 10.1108/ijppm-03-2018-0092.

Garg, N. (2020), "Exploring role of gratitude in developing teacher leadership in Indian universities", *International Journal of Educational Management*, Vol. 34 No. 5, pp. 881-901, doi: 10.1108/ijem-07-2019-0253.

Garg, N. and Sarkar, A. (2020), "Vitality among university students: exploring the role of gratitude and resilience", *Journal of Organizational Effectiveness: People and Performance*, Vol. 7 No. 3, pp. 321-337, doi: 10.1108/joepp-03-2020-0041.

Garg, N., Katiyar, N. and Mehak (2021), "Gratitude questionnaire (GQ-6)—exploring psychometric properties in India", *Journal of Religion and Health*, Vol. 60 No. 5, pp. 3716-3731, doi: 10.1007/s10943-021-01419-y.

Garg, N., Kumari, S. and Punia, B.K. (2022), "Interrelation between gratitude and vitality among students: exploring mediating role of resilience", *Psychological Studies*, Vol. 67 No. 2, pp. 200-207, doi: 10.1007/s12646-022-00654-x.

George, D. and Mallery, P. (2003), *SPSS for Windows Step by Step: A Simple Guide and Reference, 11.0 Update*, 4th ed., Allyn & Bacon, MI.

Giacolini, T., Conversi, D. and Alcaro, A. (2021), "The Brain emotional systems in addictions: from attachment to dominance/submission systems", *Frontiers in Human Neuroscience*, Vol. 14, doi: 10.3389/fnhum.2020.609467.

Goldfield, G., Moore, C., Henderson, K., Buchholz, A., Obeid, N. and Flament, M. (2010), "The relation between weight-based teasing and psychological adjustment in adolescents", *Paediatrics and Child Health*, Vol. 15 No. 5, pp. 283-288, doi: 10.1093/pch/15.5.283.

Goldschmidt, A.B., Wall, M., Choo, T.H.J., Becker, C. and Neumark-Sztainer, D. (2016), "Shared risk factors for mood-, eating-, and weight-related health outcomes", *Health Psychology*, Vol. 35 No. 3, pp. 245-252, doi: 10.1037/hea0000283.

Gredler, G.R. and Olweus, D. (2003), "Bullying at school: what we know and what we can do", Malden, MA: Blackwell publishing, 140 pp., $25.00, *Psychology in the Schools*, Vol. 40 No. 6, pp. 699-700, doi: 10.1002/pits.10114.

Greenleaf, C., Petrie, T.A. and Martin, S.B. (2013), "Relationship of Weight-Based teasing and adolescents' psychological Well-Being and physical health", *Journal of School Health*, Vol. 84 No. 1, pp. 49-55, doi: 10.1111/josh.12118.

Greenleaf, C., Petrie, T. and Martin, S. (2017), "Exploring weight-related teasing and depression among overweight and obese adolescents", *European Review of Applied Psychology*, Vol. 67 No. 3, pp. 147-153, doi: 10.1016/j.erap.2017.01.004.

Guan, B. and Jepsen, D.M. (2020), "Burnout from emotion regulation at work: the moderating role of gratitude", *Personality and Individual Differences*, Vol. 156, 109703, doi: 10.1016/j.paid.2019.109703.

Harman, H. (1960), *Modern Factor Analysis*, 1st ed., 3rd Printing, UNIV OF CHICAGO + PRESS, Chicago.

Hayes, A.F. (2017), *Introduction to Mediation, Moderation, and Conditional Process Analysis, second edition: A Regression-Based approach (methodology in the social sciences)*, 2nd ed., The Guilford Press, New York.

Hosen, I., Al Mamun, F., Sikder, M.T., Abbasi, A.Z., Zou, L., Guo, T. and Mamun, M.A. (2021), "Prevalence and associated factors of problematic smartphone use during the COVID-19 pandemic: a Bangladeshi study", *Risk Management and Healthcare Policy*, Vol. 14, pp. 3797-3805, doi: 10.2147/rmhp.s325126.

Ievers-Landis, C.E., Dykstra, C., Uli, N. and O'Riordan, M.A. (2019), "Weight-related teasing of adolescents who are primarily obese: roles of sociocultural attitudes towards appearance and physical activity self-efficacy", *International Journal of Environmental Research and Public Health*, Vol. 16 No. 9, p. 1540, doi: 10.3390/ijerph16091540.

Jans-Beken, L., Jacobs, N., Janssens, M., Peeters, S., Reijnders, J., Lechner, L. and Lataster, J. (2019), "Gratitude and health: an updated review", *The Journal of Positive Psychology*, Vol. 15 No. 6, pp. 743-782, doi: 10.1080/17439760.2019.1651888.

Keltner, D., Capps, L., Kring, A.M., Young, R.C. and Heerey, E.A. (2001), "Just teasing: a conceptual analysis and empirical review", *Psychological Bulletin*, Vol. 127 No. 2, pp. 229-248, doi: 10.1037/0033-2909.127.2.229.

Kerr, D.C., Gini, G. and Capaldi, D.M. (2017), "Young men's suicidal behavior, depression, crime, and substance use risks linked to childhood teasing", *Child Abuse and Neglect*, Vol. 67, pp. 32-43, doi:10.1016/j.chiabu.2017.02.026.

Kroenke, K., Spitzer, R.L. and Williams, J.B.W. (2001), "The PHQ-9", *Journal of General Internal Medicine*, Vol. 16 No. 9, pp. 606-613, doi: 10.1046/j.1525-1497.2001.016009606.x.

Krukowski, R.A., Smith West, D., Philyaw Perez, A., Bursac, Z., Phillips, M.M. and Raczynski, J.M. (2009), "Overweight children, weight-based teasing and academic performance", *International Journal of Pediatric Obesity*, Vol. 4 No. 4, pp. 274-280, doi: 10.3109/17477160902846203.

Lattie, E.G. Lipson, S.K. and Eisenberg, D. (2019), "Technology and college student mental health: challenges and opportunities", *Frontiers in Psychiatry*, Vol. 10, doi: 10.3389/fpsyt.2019.00246.

Li, Luo, S., Mu, W., Li, Y., Ye, L., Zheng, X., Xu, B., Ding, Y., Ling, P., Zhou, M. and Chen, X. (2021), "Effects of sources of social support and resilience on the mental health of different age groups during the COVID-19 pandemic", *BMC Psychiatry*, Vol. 21 No. 1, doi: 10.1186/s12888-020-03012-1.

Liu, Y., Yang, Y., Jia, X., Wang, C., Mo, J., Jiang, Z. and Lin, L. (2020), "The role of hopelessness and gratitude in the association between brooding and Chinese college students' suicidal ideation: a moderated mediation model", *Psychology, Health and Medicine*, Vol. 26 No. 6, pp. 764-770, doi:10.1080/13548506.2020.1800053.

Madowitz, J., Knatz, S., Maginot, T., Crow, S.J. and Boutelle, K.N. (2012), "Teasing, depression and unhealthy weight control behaviour in obese children", *Pediatric Obesity*, Vol. 7 No. 6, pp. 446-452, doi: 10.1111/j.2047-6310.2012.00078.x.

McCullough, M.E., Emmons, R.A. and Tsang, J.A. (2002), "The grateful disposition: a conceptual and empirical topography", *Journal of Personality and Social Psychology*, Vol. 82 No. 1, pp. 112-127, doi: 10.1037/0022-3514.82.1.112.

Mcleod, S. (2015), "Depression theories - simply psychology", available at: https://www.simplypsychology.org/depression.html

Meghani, D.T. and Harvey, E.A. (2016), "Asian Indian international students' trajectories of depression, acculturation, and enculturation", *Asian American Journal of Psychology*, Vol. 7 No. 1, pp. 1-14, doi: 10.1037/aap0000034.

Merlo, E.M., Frisone, F., Settineri, S. and Mento, C. (2018), "Depression signs, teasing and low self-esteem in female obese adolescents: a clinical evaluation", *Mediterranean Journal of Clinical Psychology*, Vol. 6 No. 1, doi: 10.6092/2282-1619/2018.6.1819.

Meston, C.M. and O'Sullivan, L.F. (2007), "Such a tease: intentional sexual provocation within heterosexual interactions", *Archives of Sexual Behavior*, Vol. 36 No. 4, pp. 531-542, doi: 10.1007/s10508-006-9167-7.

Myers, B.J., Mackintosh, V.H., Kuznetsova, M.I., Lotze, G.M., Best, A.M. and Ravindran, N. (2013), "III. Teasing, bullying, and emotion regulation in children of incarcerated mothers", *Monographs of the Society for Research in Child Development*, Vol. 78 No. 3, pp. 26-40, doi: 10.1111/mono.12019.

News Bureau (2020), “Teasing about weight and shape can lead teens into drug use”, Indian Flash News, 25 May 2022, available at: https://indianf.com/teasing-about-weight-and-shape-can-lead-teens-into-drug-use/

Obesity (2020), “WHO”, 4 April 2022, available at: https://www.who.int/health-topics/obesity#tab=tab_1

Obesity and overweight (2021), “WHO”, 4 April 2022, available at: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

Payne, D.L., Lonsway, K.A. and Fitzgerald, L.F. (1999), “Rape myth acceptance: exploration of its structure and its measurement using the Illinois rape myth acceptance scale”, *Journal of Research in Personality*, Vol. 33 No. 1, pp. 27-68, doi: 10.1006/jrpe.1998.2238.

Pössel, P. and Smith, E. (2019), “Integrating Beck’s cognitive theory of depression and the hopelessness model in an adolescent sample”, *Journal of Abnormal Child Psychology*, Vol. 48 No. 3, pp. 435-451, doi: 10.1007/s10802-019-00604-8.

Podsakoff, P.M., MacKenzie, S.B. and Podsakoff, N.P. (2012), “Sources of method bias in social science research and recommendations on how to control it”, *Annual Review of Psychology*, Vol. 63 No. 1, pp. 539-569, doi: 10.1146/annurev-psych-120710-100452.

Porter, J.S., Stern, M., Mazzeo, S.E., Evans, R.K. and Laver, J. (2012), “Relations among teasing, body satisfaction, self-esteem, and depression in treatment-seeking obese African American adolescents”, *Journal of Black Psychology*, Vol. 39 No. 4, pp. 375-395, doi: 10.1177/0095798412454680.

Prabhupada, S.B.A.C. (1997), *Bhagavad-Gita as it is (Revised)*, Bhaktivedanta Book Trust.

Puhl, R.M. and Latner, J.D. (2007), “Stigma, obesity, and the health of the nation’s children”, *Psychological Bulletin*, Vol. 133 No. 4, pp. 557-580, doi: 10.1037/0033-2909.133.4.557.

Racine, N., McArthur, B.A., Cooke, J.E., Eirich, R., Zhu, J. and Madigan, S. (2021), “Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19”, *JAMA Pediatrics*, Vol. 175 No. 11, p. 1142, doi: 10.1001/jamapediatrics.2021.2482.

Rosenberg, E.L. (1998), “Levels of analysis and the organization of affect”, *Review of General Psychology*, Vol. 2 No. 3, pp. 247-270, doi: 10.1037/1089-2680.2.3.247.

Santomauro, D.F., Mantilla Herrera, A.M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D.M., Abbafati, C., Adolph, C., Amlag, J.O., Aravkin, A.Y., Bang-Jensen, B.L., Bertolacci, G.J., Bloom, S.S., Castellano, R., Castro, E., Chakrabarti, S., Chattopadhyay, J., Cogen, R.M., Collins, J.K., Dai, X., Dangel, W.J., Dapper, C., Deen, A., Erickson, M., Ewald, S.B., Flaxman, A.D., Frostad, J.J., Fullman, N., Giles, J.R., Gierf, A.Z., Guo, G., He, J., Helak, M., Hulland, E.N., Idrisov, B., Lindstrom, A., Linebarger, E., Lotufo, P.A., Lozano, R., Magistro, B., Malta, D.C., Månsson, J.C., Marinho, F., Mokdad, A.H., Monasta, L., Naik, P., Nomura, S., O’Halloran, J.K., Ostroff, S.M., Pasovic, M., Penberthy, L., Reiner, R.C., Jr., Reinke, G., Ribeiro, A.L.P., Sholokhov, A., Sorensen, R.J.D., Varavikova, E., Vo, A.T., Walcott, R., Watson, S., Wiysonge, C.S., Zigler, B., Hay, S.I., Vos, T., Murray, C.J.L., Whiteford, H.A. and Ferrari, A.J. (2021), “Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic”, *The Lancet*, Vol. 398 No. 10312, pp. 1700-1712, doi: 10.1016/s0140-6736(21)02143-7.

Sarkar, A. and Garg, N. (2020), “‘Peaceful workplace’ only a myth?”, *International Journal of Conflict Management*, Vol. 31 No. 5, pp. 709-728, doi: 10.1108/ijcma-11-2019-0217.

Schmidt, J. and Martin, A. (2019), “Appearance teasing and mental health: gender differences and mediation effects of Appearance-Based rejection sensitivity and dysmorphic concerns”, *Frontiers in Psychology*, Vol. 10, doi: 10.3389/fpsyg.2019.00579.

Schmidt, U., Adan, R., Böhm, I., Campbell, I.C., Dingemans, A., Ehrlich, S., Elzakkers, I., Favaro, A., Giel, K., Harrison, A., Himmerich, H., Hoek, H.W., Herpertz-Dahlmann, B., Kas, M.J., Seitz, J., Smeets, P., Sternheim, L., Tenconi, E., van Elburg, A., van Furth, E. and Zipfel, S. (2016), “Eating disorders: the big issue”, *The Lancet Psychiatry*, Vol. 3 No. 4, pp. 313-315, doi: 10.1016/s2215-0366(16)00081-x.

Schvey, N.A., Barmine, M., Bates, D., Oldham, K., Bakalar, J.L., Spieker, E., Maurer, D., Stice, E., Stephens, M., Tanofsky-Kraff, M. and Sbrocco, T. (2017), “Weight stigma among active duty

U.S. military personnel with overweight and obesity", *Stigma and Health*, Vol. 2 No. 4, pp. 281-291, doi: 10.1037/sah0000057.

Sethi, H. (2021), "Overcoming the taboo of mental health in India amidst the pandemic", Financial Express, 13 June 2022, available at: https://www.financialexpress.com/lifestyle/health/overcoming-the-taboo-of-mental-health-in-india-amidst-the-pandemic/2294235/

Shapiro, J.P., Baumeister, R.F. and Kessler, J.W. (1991), "A Three-Component model of children's teasing: aggression, humor, and ambiguity", *Journal of Social and Clinical Psychology*, Vol. 10 No. 4, pp. 459-472, doi: 10.1521/jscp.1991.10.4.459.

Smith, M., Robinson, L. and Segal, J. (2022), "Depression symptoms and warning signs", HelpGuide.org, 4 April 2022, available at: https://www.helpguide.org/articles/depression/depression-symptoms-and-warning-signs.htm

Southam-Gerow, M., McLeod, B., Brown, R., Quinoy, A. and Avny, S. (2011), "Cognitive-behavioral therapy for adolescents", *Encyclopedia of Adolescence*, Vol. 3 No. 1, pp. 100-108, doi: 10.1016/b978-0-12-373951-3.00106-x.

Sundar, S., Qureshi, A. and Galiatsatos, P. (2016), "A positive psychology intervention in a hindu community: the pilot study of the hero lab curriculum", *Journal of Religion and Health*, Vol. 55 No. 6, pp. 2189-2198, doi: 10.1007/s10943-016-0289-5.

Susman, D. (2021), "How to use a theory of positive emotions to feel better", Verywell Mind, 7 April 2022, available at: https://www.verywellmind.com/broaden-and-build-theory-4845903

Szwimer, E., Mougharbel, F., Goldfield, G.S. and Alberga, A.S. (2020), "The association between Weight-Based teasing from peers and family in childhood and depressive symptoms in childhood and adulthood: a systematic review", *Current Obesity Reports*, Vol. 9 No. 1, pp. 15-29, doi: 10.1007/s13679-020-00367-0.

Tachon, G., Shankland, R., Marteau-Chasserieau, F., Morgan, B., Leys, C. and Kotsou, I. (2021), "Gratitude moderates the relation between daily hassles and satisfaction with life in university students", *International Journal of Environmental Research and Public Health*, Vol. 18 No. 24, 13005, doi: 10.3390/ijerph182413005.

Upadhyaya, K.N. (2018), *The Bhagavad Gita: Message Divine*, 1st ed., RSSB, Punjab.

Velly, A.M. and Mohit, S. (2018), "Epidemiology of pain and relation to psychiatric disorders", *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, Vol. 87, pp. 159-167, doi: 10.1016/j.pnpbp.2017.05.012.

Verma, S., Garg, N. and Arumugam, T. (2022), "Being ethically resilient during COVID-19: a cross-sectional study of Indian supply chain companies", *The International Journal of Logistics Management*, Vol. ahead-of-print No. ahead-of-print, in press, doi: 10.1108/ijlm-05-2022-0203.

Wang, H., Tian, X., Wang, X. and Wang, Y. (2021), "Evolution and emerging trends in depression research from 2004 to 2019: a literature visualization analysis", *Frontiers in Psychiatry*, Vol. 12, doi: 10.3389/fpsyt.2021.705749.

Watkins, P.C. (2014), "Gratitude and the good life: toward a psychology of appreciation", *Choice Reviews Online*, Vol. 51 No. 07, pp. 51-4145, doi: 10.5860/choice.51-4145.

Webb, H.J., Kerin, J.L. and Zimmer-Gembeck, M.J. (2020), "Increases in emotional eating during early adolescence and associations with appearance teasing by parents and peers, rejection, victimization, depression, and social anxiety", *The Journal of Early Adolescence*, Vol. 41 No. 5, pp. 754-777, doi: 10.1177/0272431620950469.

WHO (2021), "Depression", 13 June 2022, available at: https://www.who.int/news-room/fact-sheets/detail/depression

Winfrey, C. (1993), "Just teasing", *American Health*, Vol. 18, pp. 66-68.

Wise, A.L., Manolio, T.A., Mensah, G.A., Peterson, J.F., Roden, D.M., Tamburro, C., Williams, M.S. and Green, E.D. (2019), "Genomic medicine for undiagnosed diseases", *The Lancet*, Vol. 394 No. 10197, pp. 533-540, doi: 10.1016/s0140-6736(19)31274-7.

Wong, P. (2016), "Mapping the contour of second wave positive psychology (PP 2.0) | Dr. Paul wong", Drpaulwong, 7 April 2022, available at: http://www.drpaulwong.com/inpm-presidents-report-september-2016/

Wood, A.M., Maltby, J., Stewart, N. and Joseph, S. (2008), "Conceptualizing gratitude and appreciation as a unitary personality trait", *Personality and Individual Differences*, Vol. 44 No. 3, pp. 621-632, doi: 10.1016/j.paid.2007.09.028.

Wood, A.M., Froh, J.J. and Geraghty, A.W. (2010), "Gratitude and well-being: a review and theoretical integration", *Clinical Psychology Review*, Vol. 30 No. 7, pp. 890-905, doi: 10.1016/j.cpr.2010.03.005.

Wright (2018), "Cyberstalking victimization, depression, and academic performance: the role of perceived social support from parents", *Cyberpsychology, Behavior, and Social Networking*, Vol. 21 No. 2, pp. 110-116, doi: 10.1089/cyber.2016.0742.

Wu, P.E., Styra, R. and Gold, W.L. (2020), "Mitigating the psychological effects of COVID-19 on health care workers", *Canadian Medical Association Journal*, Vol. 192 No. 17, pp. E459-E460, doi: 10.1503/cmaj.200519.

Xiang, Y. and Yuan, R. (2020), "Why do people with high dispositional gratitude tend to experience high life satisfaction? A broaden-and-build theory perspective", *Journal of Happiness Studies*, Vol. 22 No. 6, pp. 2485-2498, doi: 10.1007/s10902-020-00310-z.

Yoshida, M. (2022), "Network analysis of gratitude messages in the learning community", *International Journal of Educational Technology in Higher Education*, Vol. 19 No. 1, doi: 10.1186/s41239-022-00352-8.

Zhang, X., Gao, F., Kang, Z., Zhou, H., Zhang, J., Li, J., Yan, J., Wang, J., Liu, H., Wu, Q. and Liu, B. (2022), "Perceived academic stress and depression: the mediation role of mobile phone addiction and sleep quality", *Frontiers in Public Health*, Vol. 10, doi: 10.3389/fpubh.2022.760387.

Further reading

Eisenberg, N., Cumberland, A., Spinrad, T.L., Fabes, R.A., Shepard, S.A., Reiser, M., Murphy, B.C., Losoya, S.H. and Guthrie, I.K. (2001), "The relations of regulation and emotionality to children's externalizing and internalizing problem behavior", *Child Development*, Vol. 72 No. 4, pp. 1112-1134, doi: 10.1111/1467-8624.00337.

Greenleaf, C., Pecinosvky, C., Petrie, T. and Martin, S. (2014), "Media use and body satisfaction, social comparisons, physical activity, and dietary restraint in adolescent females: 828 board# 243 May 28, 3: 30 PM-5: 00 PM", *Medicine and Science in Sports and Exercise*, Vol. 46 No. 5S, pp. 220-221.

Corresponding author

Naval Garg can be contacted at: naval.garg@dtu.ac.in